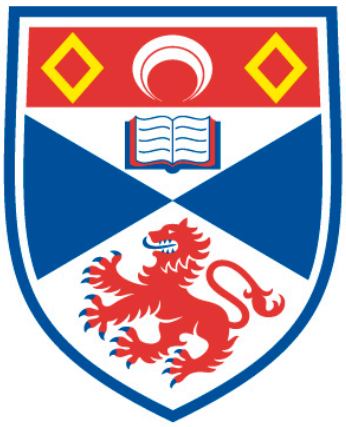
Brain death: what we are and when we die

Lukas Jost Meier

A thesis submitted for the degree of PhD
at the
University of St Andrews



2020

Identifier to use to cite or link to this thesis:
DOI: https://doi.org/10.17630/sta/17

Brain Death

What We Are and When We Die

Lukas Jost Meier

University of
St Andrews

This thesis is submitted in partial fulfilment for the degree of
Doctor of Philosophy (PhD)
at the University of St Andrews.

July 2020

Abstract

When does a human being cease to exist? For millennia, the answer to this question had remained largely unchanged: death had been diagnosed when heartbeat and breathing were permanently absent. Only comparatively recently, in the 1950s, rapid developments in intensive-care medicine called into question this widely accepted criterion. What had previously been deemed a permanent cessation of vital functions suddenly became reversible.

A new criterion of death was needed. It was suggested that the destruction of the brain could indicate the death of the organism in the presence of external life support. Soon the so-called *brain death* became the new worldwide standard. In recent years, however, doubts about this neurological criterion have been growing. Is brain death really our death?

This is the question that this thesis seeks to answer. To this end, we shall connect the medical debate about the definition of death to the philosophical debate about personal identity. While we will find that the destruction of its brain does in fact not correspond to an organism's death, we shall also ask whether the assumption that we are essentially organisms is correct. May brain death be the ceasing to exist of a different entity?

Substituting clinical case reports and considerations about human physiology for the use of thought experiments, the thesis takes a novel and philosophically unconventional approach to the problem of what we essentially are. We shall analyse various pathological conditions and their respective effects on the bodily and mental characteristics of our existence. We will conclude that brain death is indeed our death – but for reasons entirely different from those cited in the original justification of this criterion.

General Acknowledgements

The success of a PhD project is largely dependent on the support that one receives from other people. I was very fortunate to have had three excellent supervisors – Prof Katherine Hawley (University of St Andrews), Prof Michael Wheeler (University of Stirling), and Prof Thomas Fuchs (University of Heidelberg) – who supported the writing of this thesis with great commitment that extended far beyond their official duties. Their valuable comments on earlier drafts and the numerous discussions that we had during the supervisions have improved my thesis substantially. I could always turn to my supervisors for any kind of advice, for which I am most grateful.

I have also greatly benefitted from conversations with many other people, including Anita Avramides, Natalie Bergmann, Ralf Conrad, Colin Johnston, Ralf Jox, John Lizza, Simon Prosser, Alan Shewmon, Peter Sullivan, and Christian Tewes, whom I would like to thank for sharing their insights with me. For creating the diagrams included in this thesis, I thank Adrian Krohn.

Although he was not directly involved with this PhD project, I am also very much indebted to Prof Jeff McMahan, my former BPhil supervisor at the University of Oxford. Jeff has been a major influence on my philosophical thinking and I have profitted enormously from his advice and kind mentorship.

The German National Academic Foundation very generously funded not only this PhD project but also my undergraduate and graduate studies. I am extremely grateful for this extensive financial support, which also included trips to international conferences, as well as for numerous memorable experiences and encounters during the foundation's summer academies and PhD symposia.

Finally, I am deeply grateful to my parents, Tima and Dietrich Meier, for supporting me throughout my studies – emotionally, financially, and in every other imaginable way.

St Andrews, July 2020
Lukas J. Meier

Funding

This work was supported by
the German National Academic Foundation (Studienstiftung des deutschen Volkes).

For my parents

Contents

The time has been, that, when the brains were out, the man would die,
and there an end; but now they rise again.

(William Shakespeare)

1 Introduction

When does a human being cease to exist? For millennia, the answer to this question had remained largely unchanged: death had been diagnosed when heartbeat and breathing were permanently absent.[1] Only comparatively recently, three developments in intensive care medicine called into question this widely accepted criterion: the advent of positive pressure ventilation and the promotion of cardiopulmonary resuscitation, both in the early 1950s, and the first successful heart transplantation in 1967. What had previously been deemed a permanent cessation of vital functions suddenly became reversible. Not only could doctors keep patients oxygenated in the total absence of spontaneous diaphragmatic function; they could also now replace a defective heart with a functioning transplant. The traditional boundaries between life and death became blurred. Eventually it was realised that cardiopulmonary criteria were

> no longer valid when modern resuscitative and supportive measures are used. These improved activities can now restore 'life' as judged by the ancient standards of persistent respiration and continuing heart beat. This can be the case even when there is not the remotest possibility of an individual recovering to consciousness following massive brain damage.[2]

A different criterion of death was needed. It would have to be based on an organ that is indispensable to the survival of the organism as a whole, but whose function, unlike that of heart and lungs, is not replaced in intensive-care settings. Only this would permit to make diagnoses of death despite the provision of external life support. French neurologists Descotes, Jouvet, and Wertheimer were the first to publish the insight that such a criterion could only be a neurological one in their seminal 1959 paper:

> Le perfectionnement des appareils modernes de respiration artificielle permet, actuellement, d'obtenir des survies de plusieurs jours, voire de plusieurs semaines, dans le cas d'un arrêt respiratoire prolongé. Ces tentatives de réanimation semblent devoir être abandonnés, quand la mort du système nerveux central peut être établie.[3]

The brain bears a special relation to the rest of the body. While it, for example, regulates the heart rate and induces the movements of the diaphragm, it does not itself circulate blood or inhale air. Its task is not the execution but the direction of bodily functions. When the brain is destroyed, this central control mechanism is irreversibly lost,

1 The inaccuracy of the available tests resulted in frequent misdiagnoses (Winslow 1748; Garrigues 1889).

2 Ad Hoc Committee of the Harvard Medical School to Examine the Definition of Brain Death 1968, 87.

3 Wertheimer / Jouvet / Descotes 1959, 88. Translation (by L. J. M.): 'The improvement of modern devices for artificial ventilation now enables patients to survive for additional days, or even additional weeks, in cases of prolonged respiratory arrest. These attempts at reanimation must, it seems, be abandoned when the death of the central nervous system can be diagnosed.'

and the organism ceases to function in an integrated way. Any further treatment attempts, it was believed, would then be futile.

It took nearly another decade until an ad hoc committee of the Harvard Medical School finally proposed the first comprehensive guidelines for what is nowadays known as *brain death*. According to this novel criterion, the destruction of the brain is 'both a necessary and a sufficient definition of the death of a human being, in that it provides a physiological substratum for the irreversible loss of function of the organism as a whole'.[4] The committee suggested that the total loss of brain function has occurred when the patient's body shows no response to external stimuli and inner need and when spontaneous breathing as well as all other brainstem-mediated reflexes are irreversibly absent. An electroencephalography (EEG) may be performed for additional confirmation that electrical activity in the cortex has ceased. Normally, two separate sets of tests ought to be conducted. If they do not elecit any reaction, the patient should be pronounced dead.[5]

This proposal proved extremely influential, and neurological definitions of death became soon enshrined into law all over the world.[6] Most countries decided on the whole-brain criterion, which requires that all parts of the brain have lost their function before a patient can be declared dead. Canada, India, and the United Kingdom adopted a brainstem criterion, according to which only tests of brainstem function are mandatory.[7]

Recently, doubts have been growing as to whether the destruction of the brain is indeed tantamount to the ceasing to exist of the organism as a whole.[8] In the fifty years that have passed since the neurological criterion was first implemented, the world saw a huge increase in medical knowledge accompanied by rapid advancements in intensive-care technology. The number of artificially replaceable functions has been growing steadily, and brain-dead bodies can be maintained on life support for increasingly longer periods – including even the successful continuation of pregnancies for up to several months.[9] That these warm bodies whose hearts are beating, that digest food and excrete wastes, are indeed corpses as the neurological criterion advises is met with growing scepticism.

Is brain death still a suitable criterion of the death of an organism? And is it *our* death? These are the questions that this thesis seeks to answer. We will begin with an in-depth analysis of the physiological profile of brain-dead bodies in chapter 3. We

4 Lamb 1985, 7.

5 Ad Hoc Committee of the Harvard Medical School to Examine the Definition of Brain Death 1968, 85 f.

6 See, for example, the following national guidelines: Bundesärztekammer 2015; Schweizerische Akademie der Medizinischen Wissenschaften 2011; Law Reform Commission of Canada 1981; see also Wikler 1993, 239.

7 Academy of Medical Royal Colleges 2008, 18.

8 See, for example, Meier 2020c; Lizza 2018; Shewmon 2018; Miller / Truog 2016; McMahan 2009; McMahan 2006, Kurthen / Linke 1995; McMahan 1995; Veatch 1993; Zaner 1988; Gervais 1986; Youngner / Bartlett 1983; Green / Wikler 1980.

9 Miller / Truog 2009, 186.

shall compare this profile to that of other pathological conditions in which integrated organismic functioning is also reduced, but in which there can be no doubt that living organisms are present. We will identify the major differences between these conditions and brain death, and present evidence why these dissimilarities do not warrant the conclusion that brain-dead bodies on life support are dead.

We shall then take these physiological considerations to a more abstract level. I will introduce a typology of the different ways in which vital functions can cease to be performed and show why it is highly problematical to base one's judgment as to whether a biological entity is dead or alive on the status of the neurological control mechanism of a function, rather than on the execution of the task itself. We will conclude that given the great sophistication of intensive-care medicine, brain death does no longer signify that the organism has died – just as the cardiopulmonary criterion became obsolete with the advent of ventilators.

With the heart, the lungs, and the brain excluded, there seems to be no other organ whose status is suited to indicate the death of the organism as a whole. We are therefore left without any criterion that is applicable in intensive-care settings, which has serious consequences. The most worrying of these would be a sharp decline in the availability of donor organs since, according the dead-donor rule that forms the basis of organ donation programmes in nearly all countries, explantations may only be performed following a diagnosis of death.[10] How could one escape this dilemma?

In the debate about determining when we die, authors generally distinguish between definitions, criteria, and tests.[11] The *definition* of death specifies the kind of entity whose death is to be established. Most people believe that this entity is the organism. The *criteria* stipulate which particular physiological discontinuity evidences that the entity named in the definition has ceased to exist. Brain death and the loss of cardiopulmonary function are therefore two rivalling criteria that are both supposed to indicate the death of the organism. Finally, *tests* are diagnostic procedures that demonstrate the occurrence of the physiological discontinuity that the criteria specify. In the case of the neurological criterion, these are, for example, brainstem-reflex tests, EEG readings, and cerebral blood-flow studies.

When one cannot make any progress at the level of the criteria of death because there is no other organ whose destruction would be indicative of an organism's death, a good strategy may be to concentrate one's efforts at a different *level*, namely, at the level of the definition. Rather than asking what part of an organism must cease to function for the latter to die, one questions instead whether the organism is the appropriate subject of death to begin with. Is our ceasing to exist really the death of our organisms or might it be the case that we have different persistence conditions – and that our deaths therefore also require a different criterion?

10 Deutscher Ethikrat 2015, 96–113.

11 Bernat / Culver / Gert 1981.

To answer this question, the aim of chapter 4 will be to establish what we essentially are. Only briefly will we consider the possibility of us being immaterial souls, before moving on to the two main rivalling accounts of modern times, namely, biological and psychological views of personal identity: are we essentially organisms or persons? Our strategy will be to determine whether one can isolate mental characteristics from the organism. We shall do this by examining human anatomy and analysing clinical cases in which the communication between the brain and the rest of the body is severely impaired. Combined, this evidence suggests that the brain can give rise to mental phenomena even in isolation from its original body, and that, consequently, our persistence conditions may not be congruent with those of our organisms.

If this is correct, which mental continuities could be responsible for our diachronic persistence? This shall be the topic of chapter 5. According to the most influential psychological account of personal identity, the traditional Lockean view, long-term memories are the basis of our transtemporal existence. To test this view, we will apply the Lockean criterion to two conditions in which memory is at least partly extinguished – retrograde amnesia and Alzheimer's disease – and try to discern whether memory-based accounts of personal identity can provide convincing descriptions of these disorders. We shall then carry out a crosscheck by considering a rare condition in which only the neural correlates of memories are retained while all other mental capacities are irreversibly absent. Memory-based views of personal identity, we will conclude, have trouble accommodating facts about human neurophysiology.

If it is not the retention of memories that underlies our persistence, may it be the capacity for being conscious per se, that is, for being awake and aware, devoid of any specific mental content? We shall explore how such an account of personal identity could be constructed and what problems focusing on the capacity for consciousness would incur. These difficulties, we will find, are surmountable. Outcomes of neurosurgical operations will then help us to assess the relation that consciousness bears to its material substrate, the brain.

Having thus arrived at a new *definition* of what we essentially are, we can proceed to specify the necessary and sufficient conditions for this kind of entity to persist. With these findings in hand, we are finally in a position to return to the level of the *criteria* of death to determine which physiological discontinuity corresponds to the ceasing to exist of this entity. Our conclusion will be that brain death is indeed our death – but for reasons entirely different from those given in its original justification.

The task of defining death has been approached from moral as well as from metaphysical angles. The former strategy seeks to ascertain when it is *morally justifiable* to declare someone dead.[12] Its proponents believe that since death represents a change in moral status, it should also be moral considerations that indicate when this change has occurred: 'Saying people are alive is simply shorthand for saying that they are bearers of (…) rights. (…) The determination of who is alive – who has full moral

12 Lizza 2011, 745; DeGrazia 2005, 134–141; Quante 1995, 169–172; Rachels 1986, 42.

standing as a member of the human community – is fundamentally a moral, philosophical, or religious determination, not a scientific one', submits, for example, Robert Veatch.[13] To determine whether or not life support may be discontinued or organs be explanted, one must therefore establish 'what minimal quality of life in a human body possesses sufficient intrinsic value to obligate us to regard it as a living person'.[14]

Addressing the matter in this way, it seems to me, is asking a secondary question first. From the fact that deciding whether or not a patient's ventilator should be withdrawn can sometimes be a moral dilemma, it does not follow that the reasoning central to its resolution must be moral, too.[15] It is true that what rights an individual has and how she ought to be treated can be inherently moral questions. Whether or not brain-dead patients or patients in a persistent vegetative state *are* still the kinds of entities who are bearers of rights, however, is not a moral but a metaphysical issue. For not only is passing judgment on the value of life on behalf of the – usually unconscious – patients a very problematical undertaking; the loss of what makes life worth living is also not identical with the loss of life itself.[16] Reasoning on the basis of the quality of life can therefore at best show that terminating life support for brain-dead patients would be morally tenable, but such considerations do not demonstrate that these patients are indeed dead.[17] This, however, is what we are trying to establish. We shall therefore pursue a strictly metaphysical strategy. Our interest will not lie with the patient's estimated life quality, but solely with whether there is still a *subject* to whom this quality, or any other property, could be attributed. Whether or not this is the case in brain death, we will shortly explore.

Before we do so, we shall in the following chapter reflect on the method that I will be employing. Philosophical works in general, and those concerned with questions of personal identity in particular, usually rely on the use of thought experiments to develop concepts, to test hypotheses, and to decide between rivalling theories. In the domain of personal identity, these imagined situations are often fantastical and located in rather distant possible worlds. I will be arguing that this practice results in these thought experiments not conforming to the standards of good scientific experimental design; and that sometimes they even actively mislead by making unwarranted background assumptions about human physiology. It is not surprising, then, that the method of thought experimentation has not led to the resolution of many conflicts in the debate about personal identity but often rather cemented the disagreement. We shall therefore avoid the use of fantastical imaginary situations and – although rather uncommon in philosophy – take a predominantly empirical approach that relies on human neuroanatomy and physiology, on clinical case reports, and on the results of certain surgical procedures.

13 Veatch 1993, 21.
14 Hoffman 1979, 445.
15 Gervais 1986, 76.
16 Lamb 1978, 145; Schwager 1978, 44.
17 Green / Wikler 1980, 117.

The method of science fiction has its uses in philosophy, but (…) I wonder whether the limits of the method are properly heeded. To seek what is 'logically required' for sameness of person under unprecedented circumstances is to suggest that words have some logical force beyond what our past needs have invested them with.

(Willard V. O. Quine)

2 Deficiencies of Thought Experiments in Solving Problems of Personal Identity

2.1 Introduction

Thought experiments are mental test scenarios that purport to deliver scientifically acceptable results in the absence of actual physical execution. Scientists use imaginary situations as a method to test a hypothesis or to explore the scope of a concept when the respective domain is inaccessible to ordinary experimentation or because conducting a physical experiment would be too costly, ethically impermissible, or even deemed unnecessary. Hypothetical reasoning is employed in a variety of disciplines, including physics[18] and economics, and it has a particularly long and important tradition in philosophical discourse, beginning as early as in pre-Socratic time.[19] Also engaging in a meta-discourse on the thought-experimental technique, however, is a comparatively recent phenomenon.

The term *thought experiment* was introduced in 1812 by Danish physicist and philosopher Hans Christian Ørsted and later became popular through the writings of Ernst Mach, who was the first systematically to debate thought experimentation as a scientific method.[20] The past thirty years finally saw heightened interest in the topic, culminating in the publication of several monographs.[21]

The thought-experimental method has had many prominent advocates, including such major figures as Descartes and Leibniz. Proponents of thought experimentation usually maintain that

> philosophy is the analysis or articulation of the conditions of application of our concepts. As masters of these concepts (…) we have at least an implicit grasp of their application conditions; this tacit knowledge of when they apply and when they should be withheld can be manifested equally well in real and imaginary cases.[22]

18 Famous examples are Galileo's Leaning Tower of Pisa experiment and Schrödinger's cat-paradox.

19 See Rescher (2005, 61–72) for some examples from that period. Probably the most famous ancient thought experiment is Plato's Tale of the Ring of Gyges. Plato inquired whether one would remain moral if all sanctions were removed, which he tried to establish by imagining that there existed two rings that made their owners invisible, one worn by a just, the other by an unjust person (Plato 1997, 1000; *Republic*, 359d-360a).

20 Mach 1926.

21 These include Bertram 2012; Cohnitz 2006; Kühne 2005; Rescher 2005; Gendler 2000; Häggqvist 1996; Sorensen 1992.

22 Johnston 2016, 91; see also Noonan 2003, 199 f.

Appealing to intuitions about imaginary cases has also seemed dubious to some, however. At the beginning of the 20th century, French physicist Pierre Duhem made the following observation.

> Invoquer une telle expérience fictive, c'est donner une expérience à faire pour une expérience faite; c'est justifier un principe non pas au moyen de faits observés, mais de faits dont on prédit la réalisation; et cette prédiction n'a d'autre fondement que la croyance au principe à l'appui duquel on l'invoque.[23]

Some contemporary authors echo this position. A thought experiment is, as Kühne remarks, an experiment of which the main part is missing.[24] And Bernard Williams worried that it is often the way in which an author *describes* a certain situation that determines whether or not it appears to support a particular theory, while a slightly different account of the same setting could yield entirely different results.[25] Is this criticism well-founded?

Especially in debates about personal identity, philosophers have always relied heavily on thought experiments, and the intuitions that they elicit serve as weighty evidence in favour or against the proposed concepts and accounts. One of the main reasons for the great dependence on thought experiments in this field is that deciding between biological and psychological approaches to the question of what we essentially are requires situations in which these two characteristics come apart. In real-life settings, an individual's bodily continuity – however understood – and her mental continuity – however understood – either occur conjoined or else bodily continuity occurs in isolation, as in a persistent vegetative state. While we cannot learn anything from the former case, we do not know what we should say in the latter. The interesting permutation, it appears, would be the third one: the presence of psychological features in the absence of bodily continuity. Especially proponents of psychological accounts of personal identity therefore often introduce hypothetical situations that are designed to provide us with this configuration, for the study of which, it appears, we cannot resort to empirical evidence.

Pioneered by Locke's case of the prince whose soul enters a cobbler's body and his thought experiment featuring the rational parrot, authors have made frequent use of a great variety of hypothetical situations to prove or disprove their respective views about personal identity.[26] We are invited to envisage being teletransported to Mars,[27] existing as mere brains in a vats,[28] or even being on a mission to retrieve a 'Supersonic

23 Duhem 1906, 331; see also Mach 1926, 188 f. Translation (by L. J. M.): 'Employing such a hypothetical experiment is passing off an experiment yet to be executed as one already performed; it is justifying a principle not on the basis of observed facts, but on the basis of facts whose realisation one predicts; and this prediction has no other foundation than the belief in the principle on whose support it is based.'

24 Kühne 2005, 10; see also DeGrazia 2005, 26 f.

25 Williams 1970, 179 f.

26 Locke 2008, II.XXVII, § 15 and § 8.

27 Parfit 1987, 199.

28 Putnam 1999.

Tunneling Underground Device' whose special type of radiation makes it necessary to remove the brain and connect it to the decerebrated body by means of 'microminiaturized radio transceivers'.[29] Such thought experiments are certainly ingenious and creative; the question is whether they are also suited to decide between biological and psychological accounts of personal identity and to act as testing grounds for other hypotheses concerning our synchronic and diachronic persistence.

What place thought-experimental techniques should have in science in general and in philosophy in particular, as well as what epistemic status one can grant the results that this method delivers, is a very important but difficult issue. I can in this chapter only give a very brief account of what I take to be the two most severe weaknesses of using thought experiments to solve problems of personal identity, and I shall strictly limit the focus to this very area of philosophy. I will be arguing that since questions of personal identity often require hypothetical scenarios that are very distant from the actual world, many of the latter do not comply with the standards of experimental design, which is why they are ineffective at resolving conflicting intuitions; and that these scenarios can be misleading as their authors often make empirically unwarranted background assumptions about human physiology. In the final section, I shall explain how I intend to bypass some of these difficulties by following an empirical approach.

2.2 Nonconformity to the Standards of Scientific Experimental Design

Advocates of biological accounts and proponents of psychological views of personal identity differ in what they claim their respective intuitions are about situations in which bodily and mental features come apart.[30] While the former find it perfectly natural to suppose that we are wherever our bodies are, the latter insist that we must be wherever our mental features are located. Which of these mutually exclusive positions is the right one is often to be established by constructing hypothetical test scenarios in which physical and mental characteristics come apart. Usually, the opponents reach differing conclusions even when considering one and the same thought experiment; improved imagined scenarios are devised, but instead of settling the matter, they often only cement the disagreement.[31] Seldom does any newly suggested thought experiment manage to put the respective issue to rest.

How can the two camps disagree about the conclusions to be derived from the same thought experiment? In physical experiments, as they are conducted in the natural sciences, the hypothesis that is being tested can normally be regarded as either confirmed or refuted when the experiment was carefully designed and carried out according to appropriate standards. Not so in philosophy. More than three centuries have

29 Dennett 1998, 310 f.
30 See section 4.2.
31 Johnston 2016, 96; Cohnitz 2006, 165.

passed since Locke first introduced his thought experiments into the debate, yet many of the hypothetical situations that are currently discussed are simply modernised versions of his imaginary setups. If one does not doubt that there *is* something to discover in questions of personal identity, that is, that there is a fact to the matter as to what we are and in what our synchronic and diachronic persistence consists, one may start to question the aptness of employing thought experiments as a scientific method in this area of philosophy. What is it about hypothetical situations that makes them evoke so radically differing reactions?

The answer may lie in the fact that solving problems of personal identity often requires thought experiments so fantastical that they transgress the standards to which physical experiments are commonly held. One demands of physical experiments that they be objective, reliable, and valid. Put simply, an experiment is *objective* if it manages to exclude all unwanted outside influences on its result; it is *reliable* if, whenever repeated, it always yields the same outcome; and it is *valid* if it measures what it claims to be measuring. These are widely agreed standards among the scientific community.[32] Let us illustrate each of them with an example and try to establish whether classical thought experiments in the debate on personal identity adhere to these criteria.

2.2.1 Objectivity

When a person wants to establish her weight, she steps onto her bathroom scales. The outcome of this very simple experiment is objective if the value that the scales display is the result only of the person's body mass and the magnitude of the local gravitational acceleration. Other factors, like the room temperature or the person's political views, must not be taken into account.

Most experiments are much more complicated than this simple model, and the possible sources of error that threaten their objectivity are abundant. An important prerequisite of ensuring that the result of an experiment is objective is therefore strictly to differentiate between *causality* and *correlation*, that is, to distinguish a change in outcome that results from a modification of the factor under consideration from one that simply occurs contemporaneously. Hence, in experiments in the natural sciences only very few variables are actively being manipulated while all the remaining ones are held as constant as possible. This is what distinguishes a proper experiment from mere observation, where the surrounding conditions are not under the scientist's control.

The parameter that is being influenced is called the *independent variable*, and the one that is monitored for change is the *dependent variable*. If only one variable is being manipulated at a time, any difference in outcome is attributable to this very change alone. When an uncontrolled factor emerges, however, the experiment's ob-

32 See, for example, Lienert / Raatz 1994, 7–14.

jectivity is threatened as it is now unclear whether the change in the dependent variable is indeed caused by the manipulated independent variable (causation) or whether it results from a source of unwanted influence (mere correlation). Consequently, if a team of scientists wants to determine which of the two drugs that they have developed relieves pain, they will either form two groups of patients, of which one receives drug *x* while the other group is treated with drug *y*, or they will give both drugs to the same group at different times. What they must not do, however, is to administer both substances to the same patients at the same time as this simultaneous modification of two crucial variables would render the result unattributable to the variable that was actually causally responsible. The experiment's designers will also need to control for the placebo effect, for the test persons' varying bodily dispositions, and for other factors that are known to influence pharmaceutical studies. Once they have correctly factored in all potential influences, the experiment should be objective. Is the same true of hypothetical experiments in personal identity?

In his defense of the thought-experimental method, Daniel Kolak writes that '[i]n thought experiments about persons, a properly imagined set-up allows us to leave out all factors but the one under examination'.[33] If this were indeed the case, it would be great news for the objectivity of thought experiments, elevating it even above that achieved in most physical experiments, where, despite meticulous planning, it is often impossible to control each and every variable. If thought experimentation was superior in this regard, one should expect great unanimity among philosophers when it comes to questions of personal identity: unless there was a problem with the experiment's reliability or validity, everyone should happily accept the outcome. This is not as things stand. Let us examine why this is so on the basis of what is probably the most famous thought experiment in the literature: Derek Parfit's teletransportation case.

> I enter the Teletransporter. (…) When I press the button, I shall lose consciousness, and then wake up at what seems a moment later. In fact I shall have been unconscious for about an hour. The Scanner here on Earth will destroy my brain and body, while recording the exact states of all of my cells. It will then transmit this information by radio. Travelling at the speed of light, the message will take three minutes to reach the Replicator on Mars. This will then create, out of new matter, a brain and body exactly like mine. It will be in this body that I shall wake up.[34]

The imagined situation is intriguing and it has sparked off a long and lively debate. Will it really be *I* who wakes up in the newly created body? Parfit confronts us with a world that is radically different from our actual one – even more so than it would initially seem. It is a world in which technology is so advanced that there exists a machine that can translate the characteristics of living matter into information to manufacture a qualitatively identical duplicate according to the blueprint. Such procedures may not be possible in the actual world, but that is not the main worry. In thought experiments

33 Kolak 1993, 46.

34 Parfit 1987, 199. For a slightly different setup, see Carruthers 1995, 198.

we are precisely invited to speculate, and the stage of technological advancement, one may argue, is irrelevant to the question of personal identity. The problem is that with a world so disposed other factors creep in, further independent variables, that are not explicitly mentioned in the description but have the potential to influence the conclusion.[35]

One should, for example, expect that the inhabitants of a world in which teletransportation devices are available would have very different attitudes towards life and death. An person's life would be something that could be suspended and stored on any data medium to become later manifested in different matter. As there is no requirement that the blueprint, once created, is immediately transmitted to the replicator on Mars, this imagined world must also include the possibility of time travel into the future. Before entering the teletransporter, one could instruct the machine to delay the transmission for hundred years and finally leave the replicating booth without having grown older even a single day.[36] It would be a world in which parents could meet their children at an older age than they have reached themselves if, after their child is born, the parents enter the teletransporter and delay replication for a sufficient amount of time. In what would a person's death in such an environment even consist? Would it be the deletion of the final remaining copy of the blueprint? Or would it be the ceasing to exist of the last living clone, on the condition that another replication will never be attempted?

Phrased in the terminology of physical experiments, these are independent variables, that is, manipulable integral parts of the setting in which the experiment takes place that potentially exert an influence on its outcome. Instead of, as it initially appears, making alterations only to the one specifically named parameter that distinguishes the experimental world from our actual one – 'humans can be teletransported' – a high number of other variables were also tacitly amended. Among these are so elementary ones as 'adult human beings can be created from non-living matter' and 'people can travel to the future'.

Such a large number of independent variables changed at the same time makes it impossible to determine whether the observed change in the dependent variable and the conclusion derived from the thought experiment ('the individual leaving the teletransporter would / would not be numerically identical with the one who entered it', and thus 'psychological continuity is / is not a sufficient condition of a person's diachronic existence') is indeed the result only of the one condition of the actual world that the thought experiment was originally supposed to waive or just *correlates* with it, while resulting from any of the other independent variables that were also modified. Causality and correlation cannot be distinguished.

35 Wilkes 1999, 45.

36 Whether this applies to numerically or qualitatively identical individuals depends, of course, on the conclusion that one derives from the thought experiment.

When scientists cannot banish all undesired influences in physical experiments, certain countermeasures are available. They include the use of positive and negative controls, blinding, randomisation, and many other statistical devices. Which of these are employed varies between disciplines. If implemented correctly, these measures can be very powerful. Unwanted influences in thought experiments, however, cannot be balanced out in these ways. It would not be helpful to form control groups of thought experimenters, to blind the experimenter to some facts of the hypothetical scenario, or somehow to apply statistical methods to the imagined situation. Consequently, in thought experiments that are as distant from reality as most setups in the debate about personal identity, the tacit introduction of spurious variables is not only very common; the countermeasures traditionally used in physical experiments are also impotent in neutralising them.

2.2.2 Reliability

The second principle of proper experimental design is reliability. Going back to our initial example of establishing one's weight, the experiment is reliable when the scales always display the same result unless there is an actual change in body mass. When the person steps on and off the scales a hundred times in a row and the established value remains constant, the experiment is most likely reliable.[37] This overall constancy of a measurement under unchanged conditions is crucial to ensure that experiments are comparable and that their results are reproducible.

How reliable is experimenting with hypothetical situations? Intrapersonally, that is, when one and the same philosopher executes a thought experiment, the results that imagined situations deliver are indeed rather consistent. While there have been cases in which philosophers have changed their minds about certain thought experiments, they usually adhere to their favourite interpretation.[38] Thought experiments are reliable devices in this regard.

Obviously, scientific experiments count as producing reliable results only if the latter are *inter*personally consistent, that is, if *different* experimenters come to the same conclusion. Whether, for instance, a drug to be tested is administered by a doctor who appears trustworthy or by a colleague who gives off an unskilled impression may influence how patients rate the effectiveness of the medication that they receive. Well-devised experimental designs consequently aim to minimise these effects to the largest possible degree to achieve intersubjective reliability. Regardless of whether the

37 In this context it is indecisive whether the experiment indeed measures what it purports to be measuring – this is a question of *validity* (see below). If, for example, the result displayed is consistently 1 kg lower than the individual's true weight, the test is still reliable; it is just not valid.

38 One might argue that in these cases the thought experimenters have arrived at their new position in response to persuasive philosophical debate, not through re-evaluating the imaginary scenario. If true, however, this only shows that hypothetical experiments are weak decision factors that can be overruled by other considerations if deemed appropriate.

same person conducts the experiment twice or whether different individuals execute the respective rounds, the result should not change.

Over the years, not only Parfit's imagined teletransportation has evoked highly contradictory interpretations from different philosophers, but many other famous thought experiments, like Searle's *Chinese Room* or Jackson's *Mary the super-scientist*, share this fate.[39] Let us stick to the field of personal identity, however. Whether or not we would survive a procedure like teletransportation is not just a trivial disagreement but a huge difference. Consequently, there must be something about this and other hypothetical situations that allows the philosophical community to be as divided as it is.

As in the case of objectivity, certain procedures are in place to identify and correct poor reliability in physical experiments. The easiest strategy is replicating and repeating the test, if necessary several times, with a different set of experimenters while other factors are held constant. When experimenting with thoughts, however, exchanging the experimenter is not possible. While philosophers can collectively develop and amend a certain thought experiment, its actual execution remains confined to each individual alone. One needs to imagine the respective situation and then go through the sequence of steps on one's own, and while one may communicate with other people during this process, every conclusion reached is the very own conclusion of the respective thinker. In a sense, thought experiments can only be carried out in private. This is in sharp contrast to physical experiments, which not only can, but often also must, be jointly executed.[40]

Consequently, an important factor which has the potential to affect reliability cannot be controlled for since one cannot simply exchange the experimenter *and* conduct the experiment under otherwise identical conditions. When imagined situations are relatively close to the actual world, this does not present a problem because such situations should normally elicit similar reactions among philosophers. The difficulty appears when the setup is very far-fetched, as is often the case in situations meant to solve problems of personal identity.[41] As detailed in the foregoing section, some parameters of imagined scenarios are fixed in the description that their respective authors gave, but many are usually left unarticulated. The subjects who carry out the experiments must then deal, consciously or subconsciously, with a multitude of indeterminate variables. The more distant the hypothetical situation is from reality, the more of these variables demand specification and the more varied will be the assumptions that the thought experimenters must make.[42] How an experimenter fills in these gaps will

39 See Searle 1980 and Jackson 1982. An overview of the doubts about these thought experiments can be found in Cole 2020, section 4, and Nida-Rümelin / O Conaill 2019, section 4.

40 Sorensen 1992, 241.

41 See also DeGrazia 2005, 26 f.

42 Their number will in most cases be higher than in physical experiments since most of the parameters of the latter are automatically set by the situation that obtains in the very place and at the very time at which the experiment is executed. This is not true of most hypothetical situations, whose description ordinarily does not exceed a couple of sentences despite demanding enormous modifications vis-à-vis the familiar conditions of the actual world (see the following section).

depend on her general philosophical beliefs, on her cultural background, and on many other influences, of which not all are necessarily available to introspective evaluation in an explicit form and hence to interpersonal discussion. Exchanging the experimenter would therefore inevitably mean also to replace most of these background assumptions. Isolating the one from the other, as would be required for proper intersubjective control and thus for achieving reliability, is impossible.

2.2.3 Validity

The three principles of good experimental design form a certain hierarchy. Objectivity is a necessary, but not a sufficient, condition of reliability; and reliability is a necessary, but not a sufficient, condition of validity. Hence, an experiment that is not objective can be neither reliable nor valid. And an experiment that is not reliable may well be objective but it cannot be valid either. We have seen that many thought experiments employed in the debate about personal identity do not fulfil the requirements of objectivity and reliability. If this is correct, their validity is equally threatened.

An experiment is valid when it measures the very parameter that it was designed to measure. If the scales in our example determined, for instance, the room temperature, this could well be an objective measurement (if it was free from other influencing factors) and it could also be a reliable one (if the established value always corresponded to the true temperature), but the experiment would not deliver a verdict on the parameter demanded of it, namely, the person's weight. Hence, it would not be valid.

Whether thought experiments devised to decide questions of personal identity yield valid results can obviously not be answered in general. Many of them of them certainly do. Others do not. Consider, for instance, the following one due to Peter Unger.

> One of my half-brains may be gradually both bisected and fitted with radio transceivers at the opening interface. To get a very gradual spectrum, we may use this plodding procedure: At each stage, we always bisect just one largest brain-part of those then available in the situation. So, after we have half-brains, we will have one half-brain communicating with two quarter-brains. Then there will be four quarter-brains communicating; then two eighth-brains and three quarter-brains all communicating; then four eighth-brains and two quarter-brains, and so on, and so forth. This bisecting procedure can be repeated time and again, with arbitrary assignment to one side or the other in cases where the starting number of cells is odd, not cooperatively even. Eventually each neuron of my present brain will be in a supportive dish of its own, in splendid isolation from the others, while hooked up to an enormously complex device that is, among other things, a radio communicator. At any stage in this spectrum of radio communications, each resulting brain-part may be moved so that it is a few miles from

> the others then maximally intact. (…) Will I exist in such a tremendously scattered condition as that?[43]

To be valid, the thought experiment should demonstrate how our stream of consciousness would behave in a possible world that is mostly like the actual world, but contains a minimal technological alteration that is needed to make the experiment work. What it does in fact establish, however, is whether we would presumably continue to exist when a machine replaced the brain's synaptic circuitry while the neurons themselves remain organic, when cerebral tissues can be sliced into infinitely small units that nonetheless retain all their functions, when data can be extracted from such minute portions of brain tissue and be transmitted to other units in real time, and so forth. What can we learn about ourselves from hypothetical situations of this kind?

With so many conditions of the actual world manipulated, it becomes totally unclear whether the experiment really measures what it is supposed to measure, and its validity is severely endangered. After all, if one put the scales from our example on the moon, the displayed weight would be different, too, although the change would just be one of location rather than a modification of the inner workings of our brains. Thus, while we can imagine scenarios like Unger's and formulate questions about these possible worlds, it is, I take it, very doubtful that the answers one might find in them should tell us what we had originally sought to establish about *our* world and *our* persistence. Thompson and Cosmelli remark that

> if all that matters is conceivability, then we can avail ourselves of whatever conceivable technical resources we need, regardless of whether such resources are remotely feasible or even possible in our world or in worlds with our laws of nature. But such conceivability or possibility in principle tells us virtually nothing of interest with regard to what concerns us here, namely, the explanatory framework of the neuroscience of consciousness.[44]

However, many hypothetical situations on which investigators of personal identity rely in order to probe their intuitions are exactly that: testing grounds so distant that the results they yield can hardly be pertinent to our world as it is, and thus be of any relevance to the hypothesis that the thought experiment is meant to evaluate or to the concept that it seeks to explore.[45] Hence, they are not valid.

2.3 Unwarranted Background Assumptions about Human Physiology

I have been arguing that many thought experiments in the debate about personal identity fail to conform to the basic methodological standards of scientific experimentation, which makes them ineffective at resolving conflicting intuitions. The main reason that we identified for this was that the imagined *possible* worlds are often bizarre. An-

43 Unger 1990, 205.

44 Thompson / Cosmelli 2011, 174; see also Wilkes 1999, 46. Beck (2014, 193) does not find this problematic.

45 Wittgenstein 1967, 64 (§ 350).

other worry is this: thought experiments employed in questions of personal identity sometimes actively mislead by making unwarranted background assumptions about physiological facts obtaining in the *actual* world. Let us consider another prominent example.

Probably the most frequently used hypothetical scenario in the debate is that of brain transplantation. Shoemaker was the first to present such a case, but several authors have been following his lead, suggesting many different variations.[46] It is a modern version of the Lockean scenario in which the prince and the cobbler switch souls.

> So let us imagine the following. First, suppose that medical science has developed a technique whereby a surgeon can completely remove a person's brain from his head, examine or operate on it, and then put it back in his skull (regrafting the nerves, blood vessels, and so forth) without causing death or permanent injury; (…) One day, to begin our story, a surgeon discovers that an assistant has made a horrible mistake. Two men, a Mr. Brown and a Mr. Robinson, had been operated on for brain tumors, and brain extractions had been performed on both of them. At the end of the operations, however, the assistant inadvertently put Brown's brain in Robinson's head, and Robinson's brain in Brown's head. One of these men immediately dies, but the other, the one with Robinson's body and Brown's brain, eventually regains consciousness. (…) Over a period of time he is observed to display all of the personality traits, mannerisms, interests, likes and dislikes, and so on that had previously characterized Brown, and to act and talk in ways completely alien to the old Robinson.[47]

We are now asked whether this individual is Brown or Robinson and thereby also to reach a verdict on whether our diachronic persistence consists in mental or in bodily continuity. The variable that Shoemaker officially changed in this case may be paraphrased as 'medicine is so advanced that brain transplantations are technologically possible'. While this would, of course, also result in some other factors being modified (people will, for example, have higher life expectancies), it may be reasonable to assume that, unlike in the case of teletransportation, which was located in a much more distant possible world, these alterations would not be far-reaching enough significantly to influence the conclusion. Consequently, the thought experiment's objectivity does not appear to be threatened. Of the many hypothetical situations that philosophers have devised to test their hypotheses about personal identity, whole-brain transplantations are certainly among the least demanding ones. A possible world in which this operation can be carried out is reasonably close.

The problem here is a different one. The hypothetical situation presupposes that a specific relation obtains, in the actual world, between the brain and the rest of body: it is assumed that a particular body does not exert a significant influence on mental features – and vice versa.[48] We are invited to decide between physical proper-

46 See, for example, Perry 1972, 463 and Williams 1970, 162 f.

47 Shoemaker 1964, 23 f.

48 Bernard Williams (1970, 161) raised the worry that the new body might not be able to *express* the personality traits that it now houses. I do not think that this in itself presents a problem as there are

ties (remaining with the body) and psychological properties (relocated with the brain), without considering the possibility that the characteristics of the latter may depend on those of the former. The standard reply would be that this should not matter precisely because we are dealing with a *hypothetical* situation, not with reality. One must look closely to see why this is not so: the counterfactual assumptions that this thought experiment makes are not supposed to extend to human physiology but to remain within the realm of technological advancements. If it is simply taken for granted that a brain would behave sufficiently alike in a different bodily environment, the reasoning becomes circular: in tacitly conjecturing that the body would not exert an influence on the brain that would be relevant to the person's psychological identity, one is begging the question against the advocate of biological views; for whether a person would persist when her brain was separated from the original organism and transferred to a different body is exactly the question at issue. Whether or not, or to which extent, bodies influence their brains and the mental processes to which the latter give rise is therefore highly relevant to the verdict that the thought experiment delivers if a *petitio principii* is to be avoided.

Brain and body are intimately connected via the nervous-, the endocrine-, and the vascular systems, through which they constantly exchange high amounts of electrical impulses and chemical substances. The question of how much influence a certain body exerts on the brain or on the mind is a genuinely empirical one, and engaging in purely conceptual speculation about this point is futile. One may well have an intuition as to what would happen if a brain was transplanted, but this intuition could easily be false. It could, for instance, turn out that the interaction between the brain and the rest of the body was so peculiar to a certain organism that in the new environment the brain could not give rise to mental properties at all.[49] That one can *imagine* existing in Napoleon's body or that one can *coherently entertain* the possibility of being a brain in a vat does not mean that thought experiments based on such conjectures yield meaningful conclusions. The problem of being unfamiliar with the respective empirical findings is, as Wilkes remarks,

> particularly pertinent to thought experiments concerning personal identity, precisely because most of the thought experimenters know little (and

other situations in which we accept that a particular individual is present despite a total absence of motor output as, for example, locked-in syndrome evidences (see section 3.3). But what if the target of the cerebral transplantation was an organism of the opposite sex (Steinhart 2001, 21 n. 12)? In this case, the receiving body would exert a radically different hormonal influence on the implanted brain. See the section on hormones (4.4.2) for a general overview.

49 In section 4.4, I will present evidence which shows that this is in fact not so. The point is, however, that one is not here dealing with an a priori proposition at which one can arrive without taking into account neurophysiological data. One may, of course, try to minimise the effects of such confounding variables by conducting one's thought experiment with the bodies of identical twins, as McMahan (2002, 20) does. This is a clever move, which, however, is not entirely free from empirical assumptions either: even identical twins are, strictly speaking, not phenotypically identical. And the question to which extent phenotypical differences affect mental characteristics is, once again, an empirical one.

> unfortunately care less) about biology and physiology (…), and relevant obstacles to the derivation of the conclusion (…) will be ignored.[50]

To ensure that this is not just a feature of the specific hypothetical situation that we have selected for discussion and does not extend to other thought experiments, let us analyse another prominent example from the literature. Originally suggested by Wiggins and later developed by Parfit, authors usually employ it to evaluate psychological criteria of personal identity under the condition of fission.[51]

> My body is fatally injured, as are the brains of my two brothers. My brain is divided, and each half is successfully transplanted into the body of one of my brothers. Each of the resulting people believes that he is me, seems to remember living my life, has my character, and is in every other way psychologically continuous with me.[52]

The brainstem houses the ascending reticular activating system (ARAS), which is an indispensable contributor to the generation of consciousness.[53] For this thought experiment to work, one must therefore make the empirical assumption that brainstems can be divided without rendering them functionless, so that each half can be transplanted together with the respective cerebral hemisphere. Parfit acknowledges that 'it seems likely that it would never be possible to divide the lower brain, in a way that did not impair its functioning', but he contends that this would not matter as this impossibility would be 'merely technical'.[54] In this case, the physiological background assumption is thus made explicit, which is very helpful. What, however, does it mean for an impossibility to be merely technical and how does this bear on the strength of the conclusion derived?

One may distinguish several notions of impossibility. A case of *logical* impossibility would, for example, obtain if one and the same thing was both a person and not a person. By *metaphysical* (or conceptual) impossibility we mean what 'could not happen given our backing scientific knowledge: what our theories [do not] allow to be possible'.[55] Things that are *physically* (or nomologically) impossible are not in accordance with the laws of nature. *Technical* impossibility, finally, is the weakest of these notions, or, in other words, the one located in the closest possible world. It denotes, I take it, something that is logically sound, physically possible as well as in accordance with scientific concepts, but that has not been realised due to contingent reasons like a temporary lack of scientific progress. It is possible in principle.

Removing this obstacle in a thought experiment would consequently only necessitate the modification of the level of technological advancement as compared to the one prevalent in the actual world. It would not require that one tinker with any

50 Wilkes 1999, 19 f.
51 Wiggins 1967, 53.
52 Parfit 1987, 254 f.
53 As the role of this system will become important later, I shall postpone a detailed discussion until section 5.2.2.
54 Parfit 1987, 255.
55 Wilkes 1999, 18.

laws of nature. Prima facie, adjustments of this type should not pose a problem when the thought experiments in which they feature are designed carefully.[56] But does the procedure on which Parfit's thought experiment relies – the division of the brainstem – really fall under this category? Is it really only some contingent technical difficulty that prevents its realisation in the actual world, such as a lack of available surgical instruments?

Unlike the cerebrum, the brainstem is not a paired organ. The nuclei of the ascending reticular activating system, which serves both cerebral hemispheres simultaneously, are interrelated in a manner that precludes any attempt at slicing them in half without destroying this delicate structure. It is therefore not the case that if only we had at our disposal more sophisticated medical equipment, we could create two separately functioning ascending reticular activation systems out of one brainstem.[57] While it may be logically possible to divide a whole brain and obtain two independently functioning slices, it is, for all we know, a physical and consequently a biological impossibility.[58] These are not merely technical hurdles as Parfit maintained.

Whether we are in this imagined procedure considering a philosophically useful situation or just something we can somehow *conceive* depends on neuroanatomical and physiological properties of the brainstem – and thus on empirical facts that lie beyond what thought experiments can reliably establish or presuppose. As Adina Roskies remarks: 'Insofar as philosophy aims to tell us about the world we live in, it is (or should be) as bound by fact as other disciplines'.[59] Without taking into account biological facts, one is doing philosophy in the realm of science fiction, and it is highly doubtful that conclusions arrived at in this way are reliable guides to *our* persistence conditions. Not all questions can be answered from the armchair.

2.4 The Empirical Approach

'Philosophers who wish to ban thought experiment need to articulate a feasible alternative', warns Roy Sorensen.[60] The obvious alternative to relying on hypothetical test scenarios is the use of physical experiments, that is, the incorporation of empirical data. If done properly, the collection of this data can be expected to fulfil the aforementioned standards of objectivity, reliability, and validity. Ideally, therefore, causality

56 In the previous sections, we have seen that even seemingly small alterations can impact on more variables than it initially appears or was originally intended, thereby rendering unintelligible the concepts against which the hypothesis is to be tested. But this need not be so in all cases.

57 Wilkes 1999, 38 f.

58 For a surgical procedure in which doctors sever the connections between the cerebral hemispheres but leave the latter *in situ*, see section 5.3.3. These so-called *commissurotomies* are sometimes cited as real-life counterparts of Parfit's thought experiment. *If* one regards the outcome of this procedure as proof that consciousness can be divided, one should work directly with the available clinical data instead of engaging in thought experimentation. What the procedure unquestionably does not show, however, is that the brainstem can be divided, since only the pathways between the cerebral hemispheres are severed whereas the structures of the lower brain remain untouched.

59 Roskies 2016, 592.

60 Sorensen 1992, 19.

and mere correlation are carefully distinguished, the experiments yield the same results when they are repeated by different individuals, and they measure the exact parameters that they were designed to measure. Empirical evidence obtained according to these standards does usually not evoke intuitions as contradictory as those that the discussed fantastical thought experiments elicited.

Moreover, while thought experiments are often crude simplifications of highly complex situations, physical experiments naturally provide an infinitely high level of detail. It is mostly unclear what exactly setups like teletransportation would actually involve as the short description commonly provided for such thought experiments only specifies the bare minimum of the scenario. In consequence, one easily underestimates how different the imagined conditions are from those obtaining in the actual world. Conversely, frequently occurring medical disorders or the results of certain operations that have been repeated hundreds of times can be studied in full detail.

Finally, although as a merely practical consideration not philosophically decisive, working on an empirical basis secures greater applicability of the results so derived in settings in which the use of thought experimentation is uncommon or looked upon with suspicion. Our ultimate goal is to devise a definition of death that has potential relevance in practice. We will use considerations about personal identity as vehicles to arrive at this definition, but the aim is to produce results that are acceptable not only within the philosophical community but also in clinical settings. This goes for both their content and the method by with they are obtained.

Just like the thought-experimental method, empirical approaches also have their limits. While one can usually control the conditions of physical experiments better than those of bizarre hypothetical setups, the acquired data is often still open to different interpretations. Furthermore, when applied to the field of personal identity, a predominantly empirical modus operandi is restricted to what can be established about *human* persons in the *actual* world, and is thus unsuited to construct ontological accounts that specify the persistence conditions for other logically possible forms of personhood in other logically possible settings. Those who are interested in these latter scenarios will inevitably have to resort to the use of rather distant hypothetical situations.[61] Since, however, the question that this thesis seeks to answer is that of *our* death under the conditions that obtain in *this* world, I deem this constraint acceptable for the present project.

What kind of empirical evidence is it on which an investigation into the conditions of our persistence could rely? Establishing what we essentially are requires the incorporation of findings from biology, medicine, and psychology, with a focus on anatomical and physiological data, the results of neurosurgical operations like hemispherectomy or vagotomy, and the descriptions of disorders like locked-in syndrome or amnesia. Simultaneously, however, the conditions of our persistence are not solely empirically discoverable facts either, and establishing them is not like uncovering natural

61 See note 142.

laws – if this were so, disciplines other than philosophy would be better suited to deal with this question.[62] Philosophical reflections without empirical grounding may be speculative; but empirical data without philosophical conceptualisation is blind. Data does not speak for itself. It needs to be interpreted and understood through the lens of our philosophical concepts and views.

The main challenge will therefore be to process and combine the fairly concrete pieces of empirical evidence in a way that enables us to reach the highly abstract level that the discourse about personal identity requires. When our philosophical views are applied to, and tested against, this gathered data, we can hopefully reach conclusions about our persistence that are both conceptually sound and more relevant to *this* world, the world in which we are living, than those derived from distant hypothetical examples.

2.5 Conclusion

For most philosophers, thought experimentation is the method of choice for illuminating the metaphysics of personal identity. In this chapter, we were investigating the strengths and weaknesses of this method. We began with the question as to how it can be that one and the same hypothetical situation elicits contradictory intuitions and found that especially the more fantastical test scenarios that are often employed in an attempt to establish our persistence conditions fail to meet the standards of ordinary physical experimentation.

Many of these thought experiments are not *objective* because in imagined worlds that are very different from the actual one, a multitude of uncontrolled variables – rather than only the purposefully manipulated independent variable – exert an influence on the outcome. It then becomes unclear from the modification of which parameters the observed change in the dependent variable originates.

Most fantastical thought experiments have yielded results that are interpersonally inconsistent. They are therefore not *reliable*. One reason for this that we have identified is that the more distant a possible world is from the actual one, the more non-*ceteris paribus* conditions demand specification. Descriptions of the imagined setups usually do not exceed a few paragraphs, however, and some situations are so bizarre that full accounts of the envisaged world's features would assume the length of books. It is therefore the subject carrying out the respective thought experiment who must fill in these gaps, which makes the obtained result dependent on the experimenter's particularities – an influence that is to be avoided in science at all cost.

Finally, the more distant a hypothetical scenario is, the less likely does it become that the conclusions drawn on the basis of the laws and concepts that obtain in the imagined setup are applicable to our world as it is. When asking questions about personal identity, we are normally inquiring about *our* persistence conditions. Conclu-

62 Birnbacher 2017, 10; Steigleder 2015, 105.

sions derived from fantastical possible worlds therefore run the risk of not being *valid* when re-applied to the actual world.

From difficulties with *possible* worlds we moved to unwarranted background assumptions about the *actual* world. Many popular thought experiments in the personal-identity literature rely on unjustified presumptions about physiological details of the human body. While this is not a feature that is necessarily inherent to the method of thought experimentation since one could always take into account the available empirical facts, it appears that there has been little enthusiasm in the philosophical community for having the sheer endless options of imaginary setups constrained by anatomical or physiological limitations.

In this thesis, I shall try to mitigate these difficulties by relying on empirical data wherever possible. This data fulfils the same function as thought experiments would otherwise do: it acts as a testing ground for philosophical concepts. Mark Johnston wonders whether it would not

> be odd to restrict our evidence base to the adventitious experiments of nature, when we could also avail ourselves of the full range of ingeniously designed thought experiments. Wouldn't that be like only considering the moves that have been made in actual chess games, rather than the full range of moves that *could have been* made?[63]

Chess is only meaningful when it is played by the rules, and this is exactly what we should require of good thought experiments. Without rules, anything is possible, but the moves that one makes become meaningless. In areas where thought-experimental setups would have to be far-fetched and where they *can* be replaced by empirical data, this real-life testing ground is more reliable than any one located in a very distant possible world.

As we haven seen, however, authors widely disagree about which epistemic status one can grant conclusions derived from fantastical thought experiments. The arguments presented in this thesis are therefore entirely independent of the verdict that one reaches on this methodological question. Those who do not deem the thought-experimental method problematical can simply regard the empirical approach that we will pursue as complementary. Nearly everyone will hopefully concur that one ought to take into account all available evidence when issues as serious as the conditions of our persistence and the definition of our death are at stake – with implications ranging from the withdrawal of life support to saving patients through organ donation.[64]

63 Johnston 2016, 92.

64 Mark Johnston (2016, 96), for example, insists that in the debate about personal identity 'we must resort to empirical means, we must use all we collectively know and all of our capacities for argumentative ingenuity to settle the question'. Philosophical works that make use of medical data to decide questions of personal identity include Reid 2016; Savulescu / Persson 2016; Campbell / McMahan 2010; Damasio 2010; McMahan 2002; Steinhart 2001; Nagel 1985; Puccetti 1969. The most systematic approach is Kathleen Wilkes's (1999) monograph *Real People: Personal Identity without Thought Experiments*. The great majority of authors, however, rely exclusively on hypothetical situations.

Thought experimentation remains a great tool for making scientific progress. In the natural sciences, this method has been pivotal to devising new theories and models. And in philosophy, some of the most intriguing exchanges of arguments have sprung from the use of cleverly designed hypothetical situations. It is not the hypothetical method as such that is questionable; nor even is it the hypothetical method applied to questions of personal identity. It is this method combined with *overly fantastical* scenarios. This is where the conclusions drawn become unreliable or even entirely inapplicable to the actual world. It seems to be a feature of problems of personal identity that these are often the kinds of scenarios that their solution necessitates.[65]

In other domains of philosophy, thought experimentation is often much less problematic. Dilemma situations constructed to uncover moral axioms, for instance, can be set up in ways that neither necessitate distant possible worlds nor require any specialist knowledge like physiological facts. A classical trolley dilemma consists nearly entirely of *ceteris paribus* conditions and does not demand a greater understanding of the functioning of railways than any average person possesses. The stark contrast to setups in which a person's brain is cut in half and transplanted into different bodies should be obvious.

Conducting thought experiments is one of the most important trademarks of philosophical thinking. No other discipline muses about individuals who can disconnect and reunite their cerebral hemispheres at will,[66] exchange their brain states,[67] or have a dead tree in a swamp turned 'entirely by coincidence' into a physical replica of Donald Davidson.[68] We need not abandon this unconventional method of reasoning; nor, however, should we underestimate its shortcomings and limitations.

When Locke and others began to employ thought experiments to test their views on personal identity, they did not have much empirical data at their disposal, nor could they conduct systematic clinical studies. Only comparatively recently have we begun to acquire the abundance of empirical facts that are available to us today. This may explain why philosophers have not incorporated many of these pieces of information into their thinking. But it is not an excuse for continuing to refrain from putting them to use.

65 Hence, the methodological worries expressed here lie with the use of bizarre thought experiments in questions of personal identity *per se* as opposed to being targeted only at those thought experiments that conflict with my own position in the debate. In fact, authors who endorse a psychological approach to personal identity – the type of view to which I will ultimately be sympathetical – have been relying much more heavily on a 'seemingly endless litany of fantastical thought experiments' (Blatti 2019) than their opponents.

66 Parfit 1971, 6.

67 Shoemaker / Swinburne 1984, 108 f.

68 Davidson 1987, 443.

The case for whole-brain death has not been successfully made.
The tolerance of logical disorder by the medical community
in this matter is quite remarkable.

(Daniel I. Wikler)

3 Is Brain Death Organismic Death?

3.1 Introduction

A widely shared conviction is that an organism is alive when its organs function in an integrated way; and that, consequently, death occurs when this somatic unity is lost.[69] In most cases, this point is reached after heartbeat and breathing have stopped and have failed to resume spontaneously. For hundreds of years, this so-called *cardiopulmonary criterion* had been the standard for determining death until, in the middle of the 20th century, intensive-care medicine became advanced enough to replace these vital functions. What used to be permanent became reversible. In 1968, an ad hoc committee of the Harvard Medical School ultimately argued that the cardiopulmonary criterion was no longer applicable under these circumstances and suggested that neurological criteria be used instead.[70] The total loss of brain function – nowadays known as *brain death* – soon became the new international standard.

Identifying the destruction of a single organ with the death of the organism as a whole requires an exceptionally well-founded justification. In this chapter, we shall explore whether such a justification can be given. We will begin by comparing brain death to conditions that are universally accepted as constituting living organisms – the persistent vegetative state and anencephaly – to demonstrate that cognitive capacities are not essential to organismic functioning and that, hence, only brainstem-mediated functions can be relevant to biological definitions of death. With cognitive capacities excluded, five major differences remain between brain-dead bodies and bodies in a persistent vegetative state, whose respective significance to integrated functioning we shall evaluate in section 3.3 by contrasting them with conditions like high cervical spine transection, locked-in syndrome, and panhypopituitarism. We will conclude that the dissimilarities between bodies in these conditions and brain-dead bodies on life support do not warrant considering the former alive but the latter dead.

In the final section, we shall take these physiological considerations to a more abstract level. I will introduce a classification of the different ways in which vital functions can cease to be performed and show why it is highly problematical to base one's

69 Bernat 2001, 175 f.; President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research 1981, 33. I will speak of the *body* when I want to withhold judgment as to whether the entity in question is still a living organism. When I use the term *organism*, this is meant to imply that the whole entity, rather than solely some of its organs, is alive in a biological sense and constitutes a unified whole.

70 Ad Hoc Committee of the Harvard Medical School to Examine the Definition of Brain Death 1968; see also Mollaret / Goulon 1959 and Wertheimer / Jouvet / Descotes 1959.

judgment as to whether a biological entity is dead or alive on the status of the neurological control mechanism of a function rather than on the execution of the task itself.[71]

3.2 Brain Death or Brainstem Death?

An organism comprises many interrelated subsystems that work in concert. By way of this integrated functioning, the organism is able to perform higher-order functions that can only be brought about through the collaborative, internally coordinated work of mutually dependent organs or organ systems, but not by one organ or tissue in isolation. The detoxification and recycling of cellular wastes throughout the body, for instance, does not only require the participation of those organs that carry out the actual purification processes; it also presupposes an intact circulation that transports the toxins from wherever they accumulate to the target organs, supplies the latter with oxygen, and removes carbon dioxide that is produced in the process.

Determining the criterion of an organism's death therefore means identifying a change in status of an organ or of any other physiological subsystem that is indicative of the irreversible cessation of these higher-order functions and thereby marks the transition from the organism's constituting an integrated whole to being a mere collection of independent organs. According to the neurological criterion of death, it is the destruction of the brain that is to be identified with this transition. The brain executes, or enables the execution of, certain functions that are indispensable to an organism's persistence. When it stops carrying out these tasks, so the assumption goes, somatic unity dissolves.[72]

Roughly speaking, the functions to which the brain gives rise fall in two categories: cognitive and vegetative. Should both play a role in biological definitions of death? At least at present, we cannot artificially replace cognitive brain functions. If the justification given for equating brain death with our death is *psychological*, that is, if it is meant to indicate a *person*'s ceasing to exist, the status of cognitive capacities must be of primary importance. This alternative justification has lately attracted a growing number of supporters, and we shall explore it in more detail in the subsequent chapters.[73] However, if – as things currently stand – the underlying rationale is a *biological* one, that is, if it is concerned with the death of the *organism*, then the absence of cognitive capacities must be irrelevant.[74]

---

71 To avoid repetition, I shall postpone defining some of the medical terms that occur in this chapter until the section *Anatomy and Terminology* (4.4.1.1).

72 See, for example, Bernat 1998, 19 f., Lamb 1985, 37, President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research 1981, 32 f., Korein 1978, 26, and most other authors who endorse a biological justification of brain death.

73 They include Lizza 2018, McMahan 2006, Kurthen / Linke 1995, McMahan 1995, Veatch 1993, Zaner 1988, Gervais 1986, Youngner / Bartlett 1983, Green / Wikler 1980. Recently, Bernat (2018) has also expressed sympathies for the psychological rationale.

74 McMahan 2006, 45 f. Requiring that cognitive capacities be absent is intuitively very appealing. Lizza (2004, 52) speculates that the relatively high acceptance of brain death in society stems from this very fact rather than from its official justification as marking the cessation of integrated functioning and organismic unity.

One can infer this from two medical conditions. The first is anencephaly. Infants suffering from this disorder are born without a cerebrum but usually possess an intact brainstem. An anencephalic infant 'can breathe spontaneously, swallow, and grimace in response to painful stimuli. Its eyes are open. The heart can beat normally for many weeks'.[75] Thus, this infant is undeniably a functioning organism. But it is never going to be conscious or develop any cognitive capacities.[76] The other medical condition is the persistent vegetative state. The difference between anencephalic infants and patients in a persistent vegetative state is that the former, lacking a cerebrum, can never acquire cognitive capacities, whereas the latter have lost them owing to the destruction of the upper brain. We will discuss the persistent vegetative state in greater depth in the following section.

There is another argument. To the best of our knowledge, the vast majority of creatures do not develop the capacities for complex cognitive processes, yet we consider these animals to be fully functioning organisms. One could, of course, legitimately hold that the persistence conditions of human and non-human organisms need not be congruent. The universal applicability of their account, however, is what proponents of biological definitions of death generally see as one of its major advantages over the psychological rival.[77]

If anencephalic infants and patients in a persistent vegetative state are functioning organisms despite their lack of an intact cerebrum, and if many creatures that we consider living organisms do not possess any cognitive capacities to begin with, then these capacities cannot be necessary conditions for organisms to retain their integrative unity. On a definition of death that is rooted in a biological framework, the only consistent position is therefore that, in Becker's words, the 'loss of consciousness is not death any more than is the loss of a limb'.[78] Thus, if one chooses to endorse a neurological criterion of death based on a biological (rather than psychological) understanding of human life, then its anatomical locus must not include the upper brain but focus solely on the brainstem as the supposed apex of integrated functioning.

From this it follows that the so-called *whole-brain criterion*, which, as we have seen, is the standard in nearly all countries that subscribe to neurological criteria, is inappropriate. The conceptual error can, of course, be avoided if the status of the cerebrum is merely utilised as a confirmatory criterion, as the Havard Committee ini-

75 Pallis / Harley 1996, 5.

76 Merker (2007) and Shewmon et al. (1999) provide evidence for the possibility of consciousness without an intact cerebral cortex in hydranencephalic infants, arguing that the latter are awake and display emotional and orienting reactions in response to their environment (see also Fuchs 2018, 113 f. and Miller / Truog 2016, 88–95). The question is to which degree this 'primary consciousness' (Merker 2007, 80) is equivalent to the ordinary clinical notion of awareness. Several authors (cf. e.g. Coenen 2007; Collerton / Perry 2007; Doesburg / Lawrence 2007; Morin 2007) maintain that it is anatomically impossible that the structures remaining in hydranencephaly can give rise to any kind of consciousness. We shall return to the task of defining consciousness in section 4.3.

77 See, for example, Bernat 2001, 177.

78 Becker 1975, 353.

tially suggested.[79] In this case, no claim is made regarding the participation of the upper brain in integrated functioning.

If the retention of cognitive capacities is not a prerequisite for organisms to function in an integrated way, our evaluation of the neurological criterion of death must not compare, as has often been suggested, brain-dead bodies with healthy ones, but with bodies in a persistent vegetative state or in similar conditions in which cognitive capacities are irreversibly absent while vegetative functions are preserved. We must consequently ask: is the discrepancy in integrated functioning between brain-dead bodies on life support and bodies in a persistent vegetative state large enough to license the conclusion that the latter are living organisms while the former are not? This is the focus of the following section.

3.3 Functional Deficits in Brain Death Compared with Related Conditions

The persistent vegetative state, also known as *unresponsive wakefulness syndrome*, is a long-term condition in which most higher-brain functions are absent. There is no awareness of self or the environment nor any purposeful behaviour. Sleep-wake cycles, however, persist as do other brainstem-mediated autonomic functions. Spontaneous breathing is retained and cardiovascular and gastrointestinal functions continue nearly unimpaired. The body is in a state of homeostasis and homeothermia.[80]

When one compares this functional profile to that of brain-dead bodies, five major differences become apparent: (1) brain-dead bodies are irreversibly comatose, while those in a persistent vegetative state exhibit sleep-wake cycles; (2) somatomotor function is abolished in brain-dead bodies, while it is intact in the persistent vegetative state (albeit not under voluntary control); (3) all functions mediated by cranial nerves are absent in brain death but usually present in the persistent vegetative state; (4) autonomic nervous system function is heavily depressed in brain death but unaffected in the persistent vegetative state; (5) endocrine system function is often altered in brain death but normal in the persistent vegetative state. How are we to interpret these differences?

(1) As concerns the preservation of *sleep-wake cycles*, the case is clear: wakefulness is brought about by the ascending reticular activating system (ARAS) in the brain-

79 Ad Hoc Committee of the Harvard Medical School to Examine the Definition of Brain Death 1968, 338. As already mentioned, the United Kingdom is one of the few countries that do not subscribe to the whole-brain criterion but focus exclusively on the brainstem (Academy of Medical Royal Colleges 2008, 11).

80 Multi-Society Task Force on PVS 1994, 1500 f. Functional magnetic resonance imaging (f-MRI) revealed conscious awareness in some patients who met the behavioural criteria for the vegetative state in clinical assessment (Monti et al. 2010; Owen et al. 2006). While interpreting these findings is difficult (Miller / Truog 2016, 91–95), they seem to show that there are cases in which a profound dissociation between observable motor output and the actual level of residual cognitive function can occur, which may necessitate a revision of the standardised test procedures for the vegetative state (Shewmon 1997, 58–60). Since this is predominantly a *diagnostic* problem, albeit an important one, we shall not pursue it further.

stem. But devoid of a functional cerebrum, which is responsible for the awareness component of consciousness, episodes of arousal have no experiential content.[81] The persistent vegetative state is therefore very appropriately described as *wakeful unawareness*. Lacking a target organ that the ARAS could activate, its functioning does not contribute to organismic unity. Thus, the absence of sleep-wake cycles cannot be what makes the difference between life and death of an organism.

(2) Without consciousness *voluntary movements* are not possible, yet patients in a persistent vegetative state are not immobile. They may move their limbs or trunk in meaningless ways.[82] Are these non-purposeful movements, which are absent in brain-dead bodies, essential to organismic unity? One can best establish the significance of these subcortically coordinated motions by considering C1-quadriplegia. In this condition, a lesion at the level of the neck has damaged the spinal cord and prevents motor signals that originate from the brain from reaching their target muscles. Limbs and torso are completely paralysed. Hence, the spontaneous movements characteristic of the vegetative state cannot occur. In spite of this fact, we regard quadriplegic bodies as living organisms. Given that the lesion is located at the very same level at which the functional spinal cord ends in brain death, namely, at the foramen magnum, quadriplegia and brain death are exactly on a par as far as the extracephalic somatomotor deficit is concerned. And, correspondingly, spinal reflexes are preserved in both conditions since they operate independently of cerebral input.[83] If quadriplegic bodies are functioning organisms despite their lack of voluntary movements below the neck, the same criterion should apply to brain-dead bodies.[84]

Quadriplegic patients often lead fulfilling lives, and there can be no question regarding their status as persons and full members of society. In contrast, brain-dead bodies and those in a persistent vegetative state do not possess any cognitive capacities and thus they are devoid of what is probably the most defining feature of our existence. As we have seen, however, cognitive capacities are irrelevant to the question of organismic integration, and therefore to those justifications of the neurological criterion that are rooted in a *biological* framework. When a quadriplegic patient lapses into a coma from which he or she is never going to emerge, the organism does not thereby disintegrate and die. Rather, the death of the organism is a separate event that may take place months or even years later.[85]

(3) For conscious subjects, all functions mediated by *cranial nerves* are of utmost importance, as they enable them to see, hear, or smell, to move their eyes and to speak. In a persistent vegetative state, some of these functions are usually retained, so

81 Meier 2020a, 101; see section 5.2.2.1.
82 Multi-Society Task Force on PVS 1994, 1500.
83 Gordon / McKinlay 2012, 228; Han et al. 2006, 588; Pallis / Harley 1996, 9; Walker et al. 1977, 985.
84 A critic might point out that the fact that in the former case, but not in the latter, motor instructions are generated by the brain presents a relevant disanalogy – despite their ineffectiveness. This objection is blocked by the discussion in the following section.
85 See also Wikler 1993, 243 and section 5.4.

that eye-opening, grimacing, shedding tears, or occasional vocalisations may occur.[86] Conversely, in brain-dead bodies, all functions mediated by cranial nerves – sensory, motor, and parasympathetic – are absent. Prima facie, this difference appears extensive. When cognitive capacities are extinguished, however, the status of nerves I to VIII, XI, and XII may well be of great diagnostic importance as it permits doctors to test the integrity of the brainstem and thus to distinguish the persistent vegetative state from brain death;[87] but none of the functions that these nerves mediate have any bearing on somatic unity.[88]

That we can regard an organism as alive despite the paralysis of cranial nerve function is also exemplified by another condition: locked-in syndrome. Resulting from damage to the base of the pons, the transmission of practically all motor signals from the brain to the target organs, both outside of and within the head, is blocked in affected patients. Only voluntary blinking is usually possible. Consciousness is preserved.[89] Since most reflex tests will be negative and thus mimic brain death, diagnostic difficulties may occur. However, the fact that consciousness is maintained in locked-in patients is best evidence that brainstem function can only be partly absent since an entirely defective brainstem would entail a destroyed ARAS, which, in turn, would inevitably preclude any cognitive activity regardless of the status of the cerebrum.[90] Locked-in syndrome can therefore not serve as a counter-argument to the neurological criterion of death. What it does show, however, is that cranial (motor) nerve function is inessential to the basic level of integrated functioning that we require for regarding a body as a living organism. Locked-in patients sometimes survive for many years.[91]

(4) The task of the *autonomic nervous system* is to control many of the automatic functions that an organism has to perform and to adapt the activity of its organs to the requirements of different situations. The autonomic nervous system can be subdivided into the *sympathetic*, the *parasympathetic*, and the *enteric* system. Roughly speaking, the sympathetic branch increases the activity of organs, while the parasympathetic branch decreases it. The enteric nervous system governs digestion. As the autonomic nervous system operates without conscious direction, the absence of higher-brain function in the persistent vegetative state does not terminate its activities. In brain death, however, where not only the cerebrum but also the brainstem is destroyed, this system is deprived of its primary controller.[92]

86 Multi-Society Task Force on PVS 1994, 1500.

87 Cranial nerves IX and X participate in extracephalic autonomic nervous system function, are therefore potentially relevant to integrated functioning, and are consequently covered under (4).

88 As Veatch (1993, 21) notes, considering brainstem reflexes as constitutive of bodily integration while denying spinal reflexes – which persist after brain death – the same status, would be arbitrary.

89 Bauer / Gerstenbrand / Rumpl 1979, 78.

90 Schlake / Roosen 2001, 70 f.

91 Bernat 2001, 131.

92 There are two exceptions: the enteric nervous system can operate largely independently of central nervous system input (Silbernagl / Despopoulos 2009, 236); and the heart depolarises without external influence as its rhythm is generated by the sinoatrial node (Shewmon 2012, 444 f.).

Of the five major functional differences between brain-dead bodies and those in a persistent vegetative state that we have identified, the disabling of large parts of the autonomic nervous system in the former is, with regard to somatic unity, undoubtedly the most crucial one. Does a body whose organs have ceased to be connected by this overarching network turn into a mere collection of organs, that is, into a corpse? We can assess this by returning to a condition that resembles brain death in its impairment of autonomic nervous system function and with which we have already dealt: quadriplegia.

As quadriplegia results from a transection of the spinal cord, not only are somatosensory and somatomotor pathways severed, as detailed in (2), but so, too, are all fibres of the autonomic nervous system that travel through the spinal cord – namely, the entirety of the sympathetic pathways as well as the sacral branch of the parasympathetic pathways. When these fibres are disconnected from cerebral input, a multitude of visceral functions is affected. Quadriplegic patients with lesions above the third cervical segment are dependent on a ventilator for breathing, exhibit imbalances in cardiovascular- and thermoregulation, and suffer from a loss of bladder and bowel control.[93] Brain-dead bodies display precisely the same symptoms.[94] This is to be expected given that from the perspective of all body parts below the neck, there is, neurologically speaking, no difference between a transected spinal cord above which the brainstem is functional (quadriplegia) and a destroyed brainstem (brain death) since in both cases no impulses between body and brain can be exchanged via the spinal route. We will return to this peculiarity of the neurological criterion of death in the following section. Brain death is '*from the cord's perspective*, a transection at the cervicomedullary junction'.[95] In summary, as concerns the extent of neurological integration mediated by somatosensory-, somatomotor-, and sympathetic autonomic function relayed by the spinal cord, quadriplegic and brain-dead bodies are exactly on a par.

There are, however, also differences in the preservation of neurological unity between quadriplegic and brain-dead bodies. Since the non-sacral parasympathetic fibres travel through the extraspinal vagus (X) and glossopharyngeal (IX) nerves, they are unaffected by transections of the cord and continue to transmit impulses between brain and viscera.[96] This parasympathetic influence is now unopposed, which is why quadriplegic patients often suffer from severe bradycardia, hyperthermia, and bladder

93 Karlsson 2006, 2–5. It is worth noting that even devoid of rostral input, cardiac activity remains under the influence of the sympathetic centres of the spinal cord (Ouaknine 1978, 254).

94 Gordon / McKinlay 2012, 225; Wijdicks / Atkinson 2001, 35.

95 Shewmon 1999, 316. The non-endocrinologic characteristics of spinal shock after high cord transection are so similar to those of brain death that Alan Shewmon (1999) devoted a whole article to comparing the two conditions.

96 The nervus vagus (X) is the longest nerve of the autonomic nervous system. It is responsible for the parasympathetic control of several organs, in most of which its influence leads to a decrease in activity. The glossopharyngeal nerve (IX) is involved in detecting changes in blood pressure (baroreception) and in the composition of arterial blood (chemoreception). It also carries out other functions, which are not relevant to the present investigation. Technically, cranial nerves III and VII also belong to the parasympathetic system, but they do not contribute to somatic unity. See also section 4.4.1.3.

flaccidity.[97] The symptoms are especially pronounced in the acute phase of spinal shock, that is, in the months directly following the injury, but may improve later.[98]

This imbalance in autonomic nervous system activity in quadriplegic individuals leads Alan Shewmon to conclude that brain-dead bodies could in fact be viewed as being even *better* somatically integrated than the former.[99] One may reply that whether a certain function is physiologically advantageous and whether it contributes to unifying a collection of organs into an organism are two related but inherently distinct issues. Although unbalanced parasympathetic influence is usually a physiological disadvantage, its presence still means that a higher degree of communication and control exists among the body's constituents.[100] It is therefore safe to assume that quadriplegic bodies indeed manifest a greater level of neurological integration than brain-dead ones. However, as the foregoing considerations show, this difference is not profound enough to mark the line between the presence and the absence of somatic unity. To reiterate, as regards cognitive capacities, the contrast could not be starker; but we are here concerned only with those physiological characteristics that form the basis of the current justification of the neurological criterion.

(5) There is another potentially unity-conferring network with the brain at its apex that is preserved in all conditions reviewed so far, but affected in brain death: the *endocrine system*. This chemical messenger system is complex, and we shall here focus only on what is absolutely necessary for answering the question at issue. In section 4.4.2, we shall deal with the endocrine system in more detail. Just like the nervous system, the endocrine system integrates signals from different parts of the body. While the nervous system elicits immediate responses, endocrine activity is mostly geared towards long-term effects. To this end, certain glands secrete hormones into the blood stream, which then regulate physiological processes at their respective target organs.

97 Grundy / Swain 1993, 13.

98 Gordon / McKinlay 2012, 227; Karlsson 2006, 3 and 7 f.

99 Shewmon 1999, 321.

100 Insisting that the preservation of the vagal parasympathetic branch is essential to organismic life would not help the advocate of the neurological criterion, however. The side effect of an operation helps to see this: bilateral truncal vagotomy is the surgical transection of the two main trunks of the abdominal vagus nerve, which is sometimes performed to treat peptic ulcer disease. The procedure causes a decrease in peristalsis, and patients who underwent it usually report minor digestive inconveniences (Martin 2015, 3088; Clark et al. 1964, 902 f.). Governed by the enteric nervous system, digestion continues even in the absence of parasympathetic input. One might object that in a truncal vagotomy, the nervus vagus is not transected at the neck but where it enters the abdomen, so that the operation, while denervating the stomach, intestines, pancreas, and the liver, leaves intact the connections to those structures that lie above the point of separation. To be equivalent to the non-sacral parasympathetic visceral denervation that occurs in brain death, the vagus would instead have to be severed at the skull base. In this case, sensation in the supraglottis is also lost and the pharyngeal musculature as well as the vocal cord become paralysed. Due to dysphagia, tube feeding may be necessary (Montgomery / Evans / Gullane 2009, 515). These are configurations that also occur in many other intensive-care patients, however (Bernat 2001, 126). Finally, a complete vagotomy would additionally denervate the heart, thereby impairing heart rate adjustment. All recipients of cardiac transplants live with denervated hearts, and while the loss of vagal input profoundly disturbs the smooth functioning of many organs, none of the listed consequences is life-threatening given the provision of adequate support. In addition to surgical transection, vagus nerve function can also be suppressed pharmacologically, for example, by administering atropine.

The primary controller of large parts of this system is the hypothalamus – a brain structure that synthesises releasing hormones which, in turn, prompt the secretion of hormones from the pituitary gland.[101] Most of the hormones that the pituitary gland secretes subsequently act on effector hormone glands in the body. How does brain death impact on this delicate system?

Hypothalamic-pituitary function can be altered in brain death, but it is not normally completely abolished. Presumably, this is due to the fact that the inferior hypophysial artery, which perfuses parts of the pituitary gland and of the hypothalamus, arises from extradural branches of the internal carotid arteries and thus remains unaffected by the stoppage of intracranial circulation.[102] Although the hypothalamus and the pituitary gland are located within the neurocranium, their blood supply is therefore less vulnerable to increases in intracranial pressure.

Hence, brain death does not necessarily lead to endocrine failure. Anterior pituitary hormone release seems to be preserved 'on a functional level sufficient to maintain circulating hormones at least in the lower reference range even for prolonged periods'.[103] Posterior pituitary function is usually more seriously affected, which is why the majority of brain-dead bodies develop diabetes insipidus.[104] Diabetes can easily be treated by introducing antidiuretic agents, however.[105] Even if the pituitary gland *were* entirely dysfunctional, so that all types of hormones that it normally produces stopped being secreted, doctors could still substitute the products of the effector glands, as they do in the case of neurologically unimpaired patients who suffer from panhypopituitarism.[106]

Since the hypothalamus is part of the brain, preserved hypothalamic activity is, strictly speaking, inconsistent with the notion of whole-brain death.[107] Some authors tried to circumvent this problem by labelling hypothalamic neurosecretion a non-critical function.[108] In the light of other functions that they do class as critical, however, this categorisation seems ad hoc.[109]

In summary, the functional profile of brain death resembles in several important aspects that of other conditions in which the brain is functional or partly functional but its exchange of information with the rest of the body is hindered. In high-level quadriplegia and locked-in syndrome, brain and body cannot communicate via the spinal cord; after a vagotomy, in which the major parasympathetic nerve is severed, a whole branch of the autonomic nervous system ceases to function; and in panhypo-

101 There are also non-brain-regulated types of hormones, which are secreted independently of the hypothalamic-pituitary axis (Silbernagl / Despopoulos 2009, 270).
102 Schlake / Roosen 2001, 23; Wijdicks / Atkinson 2001, 30.
103 Gramm et al. 1992, 856.
104 Shewmon 2012, 459 f.; Emery / Robertson 2001, 204.
105 Wijdicks / Atkinson 2001, 32.
106 Moskopp 2015, 33.
107 Shewmon 2007, 376; Potts 2001, 482.
108 Bernat 1998, 17.
109 Shewmon 2018, S23; Nair-Collins / Miller 2017, 751; Miller / Truog 2016, 61.

pituitarism, endocrine signalling from brain to body is entirely disrupted. All of these maladies are survivable given appropriate treatment.

It should be emphasised, however, that in each of the conditions that we contrasted with brain death, a multitude of other bodily functions are retained that would be absent if the brain was destroyed. It was therefore not the aim of this section to argue that the *number* of functions lost after brain death is surpassed by any other disorder, but to demonstrate that there is no single *type* of vital function for which the absence of neurological control in brain death is incompatible with the survival of the organism. However, although the extent of functional loss in brain-dead bodies is greater than in those that are afflicted with these other conditions, the discrepancy is still not large enough to warrant regarding the latter as living organisms and the former as corpses. We shall return to this issue in section 3.5.

3.4 Decoupling of the Performance of Functions from Retention of Neurological Control

Let us now take these physiological considerations to a more abstract level. Being based on a neurological criterion, diagnoses of brain death establish the status of the brain's capacity to *direct* a certain task instead of determining whether the task *itself* is being executed.[110] 'When an individual's breathing and circulation lack *neurologic integration*, he or she is dead', submits, for instance, the 1981 President's Commission in its report on the determination of death.[111] This approach makes sense since the introduction of the neurological criterion was motivated by the desire to be able to make a diagnosis in the presence of a ventilator. The traditional cardiopulmonary criterion, which focuses directly on the *performance* of heartbeat and breathing, does not permit a diagnosis in intensive-care settings. Hence, when doctors carry out an apnoea test as part of the brain-death diagnosis routine, they do not test whether the body is being oxygenated. One would establish this by checking the oxygen saturation monitor. Rather, what they determine is whether the organism retains the *neurological capacity* to breathe, that is, whether it would in principle – *in principle* because a positive result does not necessarily entail that the respective target organ, in this case the diaphragm, is working effectively, too – be capable of breathing.

Prima facie, this difference seems trivial. But it is not. The more comprehensive life-supporting machinery becomes, the less closely does the status of the brain correspond to the functions that are in fact being carried out in the body. The diagnosis of the absence or the retention of a mere neurological capacity then overrides the much more important question of whether the respective functions are actually being performed. To see this more clearly, consider the following classification of the different

110 This would obviously not apply to the capacity for consciousness as it is a function that the brain itself performs. We have already determined, however, that the status of cognitive capacities must not figure in definitions of death that are rooted in a biological framework.

111 President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research 1981, 33; emphasis added.

ways in which a certain function, understood as the interplay between the neurological controller and its target organ, can break down. There are four possible types of malfunctioning, and I shall illustrate each of them with an actual medical condition.

Type 1	Damage to the *target organ*, while the brainstem retains the capacity for directing its function (example: Duchenne muscular dystrophy).
Type 2	Disruption of the *pathway* between brainstem and target organ, while both the brainstem as well as the organ are intact (example: spinal transection).
Type 3	Destruction of the *brainstem*, while the target organ remains undamaged (example: respiratory centre failure).
Type 4	Loss of function in *both* the brainstem and the target organ.[112]

An example of the first type is Duchenne muscular dystrophy, a condition that leads to progressive skeletal muscle degeneration. When it finally affects the diaphragm, the patient requires external ventilation.[113] Although the function itself – breathing – cannot be executed any longer, the brainstem retains the capacity to direct this task. Respiratory arrest, if irreversible, would have constituted one of the two clinical signs of death on the traditional criterion (the other one being asystole), since criteria other than whether or not a function was actually being carried out were not taken into account. On the neurological criterion, the converse is true: since it only takes into consideration whether a function *could* be controlled by the brainstem, that is, whether the stem retains the respective capacity, but not whether it is in practice being executed, absent function of type 1 does not constitute death on this definition. Hence, the organism counts as alive although not its brainstem but a ventilator is responsible for controlling the oxygenation of the body.

The same goes for type 2, where the communication between controller and target is disrupted. Despite both organs being intact, the respective function must be provided externally. When cervical spine transection occurs at cord segment C1, the patient is not only quadriplegic but the pathways connecting the brainstem to the phrenic nerves, which supply the diaphragm and exit the cord at C3 to C5, are also severed. Hence, the brainstem cannot communicate with this main muscle of respiration, and the patient is unable to breathe. Since the brainstem is intact, however, the body is regarded as alive according to the neurological criterion.

An example of what I have labelled a type-3 condition is respiratory centre failure. When this area of the brainstem is damaged, for example due to hemorrhage or trauma, breathing stops even though the diaphragm and the intercostal muscles are not affected. This is a case of a defective controller with an intact target organ.

Let us assume that the three disorders are irreversible. As concerns the demand for the provision of life support, the conditions are exactly on a par: in all three cases,

112 Typology taken from my BPhil thesis (Oxford, 2016, 28).
113 Lo Mauro / Aliverti 2016, 324 f.

the brainstem *de facto* does not control the function, which means that a ventilator is required. That external ventilation is sufficient to provide the lost function in all scenarios shows that the presence of an intact brainstem is redundant in this regard. The sole dissimilarity between the situations is that while in the first two the brainstem could *in principle* act as the controller, although, in fact, it does not, it *cannot* – not even in principle – do so in the third scenario. Is this a crucial difference?

According to the neurological criterion, those patients who suffer from conditions of the first and second types would be classified as alive, while the third patient would be regarded as dead. However, there are no physiological dissimilarities between these cases that warrant these attributions. A patient who suffers from respiratory centre failure does not differ in her ability to breathe from another patient whose diaphragm is paralysed as a result of Duchenne muscular dystrophy since the respective function is absent in both cases. The former must be a living organism if the latter is.

This also goes for type 4. Whether, in addition to a dysfunctional diaphragm, the respiratory centre is defective or not has no bearing on whether the function in question is in fact being carried out. All four scenarios yield exactly the same result: the organism is unable to oxygenate itself. That on the neurological criterion types 3 and 4 would constitute death, while types 1 and 2 would not, is arbitrary – especially in the light of the fact that the mechanical stand-in for the lost function in the first two cases is not under the control of the brainstem either. Besides the *de facto* controller of ventilation (which does not need to be inorganic – it could also be a nurse operating a bag valve mask), there exists an additional control centre, the brainstem, which is idle. Its presence is not physiologically required, as a comparison between types 1 and 3 shows.

The reason why patients survive the described conditions is not that the brain is still intact and only unable to communicate, but rather that adequate external support substitutes for a vital function that the brainstem would otherwise direct. If this were not the case, these patients would die of anoxia within a few minutes, irrespective of the status of the brain. To base the decision as to whether or not the patient is dead and life support should be terminated on a diagnosis of the brainstem's status, then, is ad hoc.

Obviously, the brainstem directs many other functions in addition to ventilation, and a patient is, of course, only declared dead when *all* its controlling capacities have irreversibly ceased. As long as the potential for neurological control over other bodily functions is retained, the brainstem is taken to integrate the different organs into a unified whole, and the organism is regarded as alive. We have, as an example, confined our considerations to ventilation, yet the point of the classification that we introduced is a more general one. When there are no grounds to require an intact respiratory centre in the case of external ventilation, then, by parity of reasoning, nor can one insist on an unimpaired neurological control mechanism of other replaceable functions.

In the previous section, we compared brain death to other conditions. Quadriplegia, for instance, is a type-2 case since it results from high spinal-cord transection: the brain and all target organs are intact, but they cannot communicate, except via the vagus nerve. If the neck injury that severed the spinal cord also damaged the vagus or if a quadriplegic patient underwent bilateral vagotomy, controller and target organs would be entirely neurologically separated.[114] The neurological criterion would nonetheless demand that the status of organismic unity be assessed by tests conducted on the brain. However, whether this neurologically isolated organ is intact or not would not make any difference whatsoever to the functioning of the organism as a whole, while the cognitive dissimilarities, which are enormous, must not be taken into account.

In every condition that we have analysed in the foregoing section, there always remain *some* vital functions that the brainstem still controls, which means that considered in isolation none of them present a problem for the neurological criterion. *Combined*, however, they show that there is, in fact, no single vital brainstem-mediated function that one cannot, at least temporarily, artificially replace; for taken together they preclude all means by which the brainstem could control integrated functioning in the organism: neural – most notably via the spinal pathways and the vagus nerve – and endocrine, via the hypothalamic-pituitary axis.

As soon as all vital functions that the brainstem normally directs can be maintained with the help of external means, the status of this organ loses its justifiability as the sole indicator of an organism's death. Fifty years after the introduction of the neurological criterion, this point has already been reached. Intensive-care medicine can provide tailored *oxygenation* by constantly adjusting various ventilatory settings to live blood gas measurements; maintain *haemodynamic stability* through the automatic administering of vasoactive drugs; maintain *normothermia* using fluid warmers and heated ventilator circuits; manage *diabetes insipidus* by administering antidiuretic agents; and regulate *glucose homeostasis* and *electrolyte balance* via targeted infusions of insulin and various other substances. These are just the more important of the technologically feasible interventions in a brain-dead body.[115]

To be clear, none of these interventions achieve the regulatory perfection that an intact brainstem would have provided. They are only relatively crude attempts at substituting fine-grained physiological processes, which is why in many cases all measures that doctors take are unsuccessful and fail to prevent asystole. Often, however, they permit them to keep a brain-dead body functioning for a considerable amount of time – sometimes even for years.[116]

It may be objected that even if a brain-dead body can be maintained for a while, eventual asystole is inevitable. In the vast majority of cases, brain death is indeed a

114 Both scenarios are relatively unlikely to occur, but represent sound theoretical possibilities.
115 For a more detailed description, see Emery / Robertson 2001, 202–206.
116 Shewmon 2018, S23; Nair-Collins / Miller 2017, 749; Shewmon 2012, 456; Shewmon 1997, 68 f.

precursor to total organ failure.[117] Arguing this way, however, is confusing a diagnosis with a prognosis.[118] The neurological criterion of death purports to pinpoint the one event whose occurrence is *identical* with the ceasing to be of a living organism. Even if brain death did inevitably herald an organism's destruction, these would still be separate events, regardless of how long the interval between them is. The new situation of the neurological criterion in current intensive-care settings is very similar to the one that the cardiopulmonary criterion was facing when heart and diaphragmatic function became replaceable. When a body's circulation is maintained by extracorporeal membrane oxygenation, one may assess the status of heart and lungs, but this is not an indication of whether or not the body is being perfused with oxygenated blood. As long as these organs are bypassed, no conclusions regarding the life or death of the organism as a whole can be drawn from their functional status. The same has now become true of the brainstem: while it is a necessary condition of organismic integration that tasks like ventilation or circulation are being carried out, it is not essential that they be neurologically directed by the brain. The relevant distinction is between the presence and the absence of a function, not between its being controlled internally or externally – provided that there remains a certain degree of coordination and regulation among the organs themselves to account for somatic unity. It is therefore no longer the case that 'when an individual's breathing and circulation lack neurologic integration, he or she is dead'.[119]

When the brain-death criterion was introduced in the 1960s, the diagnostic decoupling of the neurological control mechanism from the actual performance of vital functions was a great advancement as it made possible diagnoses in the presence of ventilators. But the legitimacy of the criterion began to shrink as the number of brainstem-directed functions that could be replaced was increasing. When the brainstem is found to be destroyed while none or only very little external assistance can be provided, it is obvious that the organism is dead already or at least in the process of dying since indispensible brainstem-mediated functions are bound to be absent. When, however, extensive life support successfully stands in lieu of all vital functions that the brainstem would otherwise direct, the status of the brainstem becomes immaterial to the organism's continued existence. A test carried out on a dysfunctional brainstem in a successfully maintained body will therefore yield a false positive, that is, the organism will be regarded as dead when it is in fact alive.[120]

---

117 Emery / Robertson 2001, 204; Wijdicks / Atkinson 2001, 35 and 39; Pallis / Harley 1996, 30; President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research 1981, 17; Korein 1978, 26 f.

118 Miller / Truog 2016, 63; Steigleder 2015, 108; Green / Wikler 1980, 110. Lamb (1985, 37) submits that the continuing functioning of the various bodily subsystems after brain death only *mimics* integrated life.

119 President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research 1981, 33.

120 This does not imply that we *should* continue life support for brain-dead bodies – whether we should do so is a *moral* question, whereas we are here concerned with a strictly metaphysical one. One may, for instance, come to the conclusion that we are justified in letting die brain-dead organisms due to

3.5 Conclusion

Organisms are characterised by internal cooperation and regulation of reciprocally dependent processes among their various parts. Advocates of the biological justification for the neurological criterion of death hold that this integrated functioning ceases irreversibly with the destruction of the brain. We put forward two objections against this assertion.

In the first part of the chapter, we drew parallels between brain death and other pathological conditions that are comparable in relevant aspects. Whenever we regard as compatible with organismic unity the absence, or the artificial replacement, of function *x* in disorder *y*, then we also ought to tolerate the loss, or the artificial replacement, of function *x* in brain death. Combining the characteristics of high cervical spine transection, locked-in syndrome, bilateral vagotomy, and panhypopituitarism enables one to show that all vital functions can continue in a body on adequate life support even if all means by which the brain could exercise control, neural and endocrine, are lost. From this it follows that the destruction of the brain does not disband somatic unity.

What this comparative method fails to establish, however, is the *exact threshold* below which too many functions are either absent or executed by external means for the body to be a living organism rather than a mere collection of isolated organs and machines. The cardiopulmonary criterion specified the irreversible cessation of breathing and circulation as clinical signs of death. Brain death, which was supposed to replace the former criterion in intensive-care settings, even narrows the possible loci down to a single organ whose status is deemed to be indicative of the state of the whole organism, thereby purporting to deliver a yet more precise cut-off point. While this degree of precision may seem attractive for diagnostic purposes and facilitates the timely procurement of organs, matters may in reality not be so straightforward. Where exactly the threshold lies between life and death, or, put differently, at which point the organism vanishes and the corpse (or the machine) begins, is a question to which we cannot give a determinate answer. It seems that with what we are here confronted is a Sorites Paradox.[121] While a body in a persistent vegetative state is clearly an organism, and a body in rigor mortis is clearly a corpse, we cannot point to a *single* event in the transition process between the two states that would mark the transformation from life to death – just as it is unclear at which point a heap of sand disappears when grains are taken away from it. Eventually, the heap will cease to exist, but it is impossible to attribute this change to the removal of a particular grain. Likewise, an organism persists as long as there exists a certain degree of internal coordination and regulation among its different organs, and when the functions that underlie these integrative processes are gradually terminating, or are being replaced externally, ultimately a

the fact that they have lost all cognitive capacities.

121 The original formulation of the paradox is attributed to Eubulides. See Hyde / Raffman 2018.

point is reached at which there ceases to be a biological life present over and above organ level.[122]

Is this vagueness linguistic or ontic, that is, is our concept of organism imprecise or are matters of biological life and death indeterminate in and of themselves?[123] The answer to this question is contingent on one's deeper attitudes towards realism, which is a topic far beyond the scope of this thesis. Suffice it, therefore, to say that if the linguistic view is correct, one could in principle eliminate the indeterminacy by making the concept of an organism more precise, for example, when additional physiological details come to light at a future time. There is no such possibility if facts about the persistence of organisms are ontically vague.

Hence, for assessing whether the destruction of the brain is a proper indicator of organismic death, the best we can presently do is locating brain-dead bodies on life support on a *spectrum* of organismic unity. We have been trying to achieve this by comparing the characteristic functional profile of brain-dead bodies to that of conditions that exemplify a smaller degree of organismic unity than the persistent vegetative state and are thereby closer to organismic death, while unanimously being regarded as belonging to the realm of life. We found the difference in integrated functioning between these conditions and brain death to be relatively small – not great enough, in any case, to warrant classifying one group as organisms and the other as corpses.

Conversely, on the other side of the spectrum, the difference in integrated functioning between brain-dead bodies on life support and bodies that begin to exhibit the classical signs of death is extensive. According to Shewmon's canonical list, the former, with the help of only limited external support, maintain homeostasis of mutually interacting chemicals, macromolecules, and physiological parameters; eliminate, detoxicate, and recycle cellular wastes; maintain energy balance and temperature regulation (to a certain degree); heal wounds; fight infections; display cardiovascular and hormonal stress responses; are able to gestate fetuses; and show sexual maturation and proportional growth (in children).[124]

The sheer number of items on this list is impressive. Even more relevant to the question at issue, however, is the fact that all of them, without exception, presuppose the coordinated participation of several organs or tissues. They involve interactions between systems as complex as the cardiovascular, the endocrine, the immune, or the lymphatic as well as of smaller components of the body, like blood cells or bone marrow. All listed functions are realised without the brainstem (or any external mechanism) exerting centralised control, yet they still achieve a high level of somatic integration through mutual interdependence. We can therefore conclude that if one leaves

122 Some might argue that an organism can comprise, or acquire, a large or even an indefinite number of inorganic parts, so that it would not cease to exist when it consisted predominantly or entirely of artificial devices. See section 4.4.3.2.

123 For an explanation of the distinction between linguistic and ontic vagueness, see Hawley 2001, 100–116.

124 Shewmon 2001, 467 f.; see also Nair-Collins / Miller 2017, 750; Jox 2014, 37; Sadovnikoff / Wikler 2014, 39 f.; Miller / Truog 2009, 186; Emery / Robertson 2001, 206; Wikler 1993, 241; Wikler 1984, 101.

aside cognitive capacities, the functional profile of brain death is reasonably close to that of the pathological conditions that we have analysed, but very distant from that of a cold, stiff corpse.

In the second part of the chapter, we argued that the growing sophistication of life-support systems gave rise to a dangerous decoupling of a function's performance from the retention of neurological control over it. We introduced a classification of ways in which a bodily function can be lost, and demonstrated that two out of four permutations constitute death on the neurological criterion, despite the number of vital functions that the body actually performs, as well as the amount of external assistance that it requires, being identical in all cases.

Provided that the level of internal coordination between the different organs is still high enough to account for a sufficient degree of somatic unity, the continued existence of an organism is not conditional on the *means* by which a certain vital function is directed, but rather on its being performed or having ceased. In intensive-care settings, the status of the brain does therefore not reliably indicate whether an organism is dead or alive since the former need not correspond to the functions that are being carried out in the body – a discrepancy that can yield false positives. For these reasons, the brain is not a suitable locus for determining the death of an organism in the presence of extensive life support. Fifty years after its introduction, the neurological criterion is facing the same fate as its cardiopulmonary predecessor.[125]

125 A modified version of this chapter appeared in *The British Journal for the Philosophy of Science* under the title *The Demise of Brain Death* (Meier 2020c).

From the brain, and from the brain only, arise our pleasures, joys, laughter and jests, as well as our sorrows, pains, griefs and tears. Through it, in particular, we think, see, hear, and distinguish the ugly from the beautiful, the bad from the good, the pleasant from the unpleasant.

(Hippocrates)

4 What Are We?

4.1 Introduction

The outcome of the previous chapter is highly problematic. If brain death is indeed not the death of an organism, we lose the sole criterion of death applicable in intensive-care settings. While the permanent cessation of heartbeat and breathing is still indicative of a dying organism when no external assistance is being provided, one cannot simply return to this criterion in the presence of ventilators, cardiopulmonary bypass, extracorporeal membrane oxygenation, and other life-preserving technological interventions.

Abandoning brain death would have many undesirable consequences. First, there would be great uncertainty as to the status of the brain dead among medical staff, relatives, and in the general public. Secondly, a multitude of legal questions about the end of life would arise. Thirdly, and most importantly, we would see a sharp drop in the number of available donor organs. According to the dead-donor rule, the widely accepted principle that forms the basis of practically all transplantation programmes worldwide, vital organs may only be procured from deceased patients.[126] As we have just established, however, brain-dead bodies are biologically alive. If this is correct, explantations must not be performed.[127] The organs from thousands of potential donors could then not be used to save other patients' lives. How could one avoid this unacceptable ramification?

An alternative way to procure organs that is not affected by the difficulties that plague the neurological criterion is the so-called *donation after circulatory death* (DCD). In this procedure, which is also referred to as *non-heart beating donation* (NHBD), life support is withdrawn from critically ill patients who do not fulfil the neurological criterion *before* they are declared dead. Doctors then wait for pulselessness to manifest, after which they observe the patient for a short interval and, if cardiopulmonary function does not return spontaneously, begin the explantation.

Unfortunately, this practice raises a host of serious conceptual and ethical problems itself.[128] To name just a few of these: how long a period of pulselessness is required following the withdrawal of life support before organ procurement may start? Is

126 Deutscher Ethikrat 2015, 96–113. Some authors suggested abandoning the dead-donor rule (cf. e.g. Miller / Truog 2016, 113–152; Jox 2014; Sade / Boan 2014; Collins 2010; Truog / Miller 2008), which would, however, lead to a multitude of ethically problematical consequences (Deutscher Ethikrat 2015, 104–113; Bernat 2013). We will return to this point in section 5.4.

127 The donation of non-vital organs (one kidney, one lobe of the liver, parts of the lung, of the pancreas, or of the intestines) and of certain tissues would, of course, still be possible.

the moment after which autoresuscitation – the spontaneous recurrence of cardiac function – is deemed unlikely sufficient or should the protocols demand actual permanence or even irreversibility? How could one determine that the cessation of heartbeat is irreversible without trying to restart the heart? May one procure hearts under this scheme given that the loss of cardiac function is the very precondition of their explantation? Is the arrest of circulatory activity a criterion in itself or is it to be understood merely as a proxy indicating the loss of neurological function? Protocols vary considerably and there is no consensus regarding these and many other uncertainties.[129]

Irrespective of how one answers these questions, which are important but beyond the scope of this thesis, donations after circulatory death could not compensate the massive decline in available organs that abandoning donations after brain death (DBD) would cause. Since thousands of patients die each year because they do not receive an organ in time although many hospitals even apply the two schemes, DBD and DCD, in parallel, there should be great interest in finding a way to continue our current practice of organ donation after brain death despite the difficulties pointed out in the foregoing chapter.

Practical consequences, as far-reaching as they may be, must not be allowed to influence our metaphysical considerations, however. We are still trying to establish a conceptually sound and empirically accurate way of determining death, regardless of whether it would be beneficial to the practice of organ procurement or advantageous in any other practical regard. We have found that equating the destruction of the brain with the ceasing to exist of the organism as a whole is incorrect, and that one cannot simply pick a different organ instead: first, because the function of all other organs is even more easily replaceable by external means than that of the brain; and secondly, because equating the death of the organism as a whole with the isolated loss of function in *any* of its constituent parts is problematical since somatic unity is a holistic property.

When one can therefore not just switch to a new criterion of death which is immune to technological progress, that is, when there is no locus in an organism whose isolated destruction could indicate a global cessation of integrated functioning, the only alternative is concentrating one's efforts at a lower level, namely, the underlying metaphysical layer. Instead of asking what part of an organism must cease to function for the latter to die, one may also ask whether the organism is the appropriate subject of death to begin with. Put differently: if one cannot make any progress at the level of the *criteria* of death, a promising strategy may be moving to the level of the *definition* of death. That our ceasing to exist must consist in the deaths of our organisms is not self-evident. Historically, at least two other types of entities have been considered –

128 For recent discussions, see Miller / Truog 2016, 97–112; Deutscher Ethikrat 2015, 113–117; Keller 2011; Veatch et al. 2011; Veatch 2010; Bernat 2010; Jousset et al. 2009; Veatch 2008.

129 Miller / Truog 2016, 98 f.

souls and persons. The question that this chapter seeks to answer is therefore this: what are we essentially?

We will now enter the debate about personal identity. I shall begin by briefly presenting the three most influential views before suggesting a novel method of deciding which of them has the better claim, based on empirical evidence. We will be investigating whether mental features can persist in a head when it is neurologically, endocrinologically, and vascularly isolated from its original organism. The chapter closes with considering possible objections to our findings.

4.2 The Three Main Accounts of Personal Identity

4.2.1 Immaterial Substance Accounts

What we essentially are and wherein the conditions of our persistence consist is one of the great traditional philosophical problems that have been recurring throughout the centuries. Roughly speaking, the answers that have been given fall into three broad categories: we are taken to be immaterial substances, biological organisms, or psychological persons.

According to what is probably the oldest view of personal identity, we are immaterial things. Often referred to as *souls*, these non-physical entities can be conceived of in various ways. A useful distinction is the one between hylomorphic and non-hylomorphic conceptions of the soul. The conviction that our mental and moral capacities cannot be properties of ordinary matter is the basis of the *non-hylomorphic* views. Plato, who developed a comprehensive theory of the soul, believed souls to be eternal entities that neither come into existence together with the body nor cease to be when the latter decays.[130] The most prominent modern advocate of a non-hylomorphic account, René Descartes, identified the soul with the mind. He regarded it as a non-spatial entity that is responsible for our thinking, our feelings, and any other mental operations.[131] Being a substance in its own right, this *res cogitans* was taken to be capable of existing in a disembodied state and therefore to survive the disintegration of our physical bodies. The belief in a bodiless afterlife is also an important element of most religions.

Hylomorphic views originate in the Aristotelian doctrine of metaphysics. Aristotle held that the soul is the formative principle that guides the arrangement of matter in such a way that the particles compose our living bodies. Matter, he asserted, can constitute a human being only if it has a certain form – as a statue comes into existence only when the clay is given a particular shape.[132] Souls thus conceived are not substances in their own right. Just as the statue's form cannot persist without its un-

130 Plato 1997, 91 f. (*Phaedo*, 106d-107a).

131 Descartes 2008, 27. For a short analysis of the metaphysical underpinnings of Descartes' view, see Lowe 2006, 6 f.

132 Aristotle 1995, 1437–1440 (*On the Soul*, Book II.I).

derlying matter, the soul does not survive the destruction of the body.[133] Consequently, hylomorphic and non-hylomorphic views differ fundamentally and postulate divergent persistence conditions.

Although immaterial substance accounts have had many proponents throughout the centuries, we shall not explore the possibility of us being souls – not necessarily because it would be beyond any reasonable doubt that this doctrine must be false, but rather since it is questionable that one could ever scientifically establish that a soul has ceased to exist or that its separation from the body has occurred.[134] This, however, is what any definition of death seeks to do. To determine whether an individual is dead or alive, one must be able to detect a structural or physiological discontinuity. The cessation of which bodily function, if any, could be associated with the destruction or the departure of a non-physical entity must be pure speculation, however. The tendency of a postulated entity to evade scientific detection is, of course, not in itself a philosophical argument against any view of personal identity. Whether or not we are souls does not hinge on the suitability of such entities for the determination of death.

One of the findings that made immaterial substance accounts fall out of favour was that scientific evidence strongly suggests that mental attributes are somehow dependent on brain processes. We do not yet know how mind and neural substrate are related, but we do know that damage to certain areas of the brain corresponds to the loss of particular mental functions. And we are able to predict this correlation with considerable accuracy.[135] It seems therefore reasonable to assume that our mental capacities and psychological characteristics are in some way caused by, or realised in, brain tissue rather than emanating from the presence of an immaterial entity. Mechanistic models have replaced animistic ones.[136] For these reasons, the competing types of accounts of personal identity are nowadays biological and psychological views. We shall therefore, without further argument, focus on these two kinds of views and leave aside the metaphysically highly demanding doctrine according to which we are souls.[137]

4.2.2 Biological Accounts

If we are not immaterial substances like souls, the hypothesis that suggests itself is that we can exist only in conjunction with *material* substances. And the most plausible candidates for these substances are our living bodies. We usually identify, re-identify, and trace over time other individuals by criteria of outward appearance. There are no known cases of people existing without their bodies, and when we think of someone,

133 See also Lucretius (2014, III.838 f.), who believed that after the disbanding of the atomic complex of body and soul both entities decay (Meier 2019, 655 f.).
134 McMahan 2002, 9; Lowe 1995, 117.
135 Schneider / Ornstein 2019, 311 f.
136 Zeman 2001, 1284.
137 For very effective refutations of immaterial substance accounts, see McMahan 2002, 8–24 and Johnston 2016, 93 f.

we picture his or her body in our minds. Consequently, it seems only logical that living bodies are also the kind of entities that we essentially are.

Biological accounts of personal identity, which have recently been attracting a growing number of proponents, hold exactly this.[138] By far the most prominent variant of this view is animalism. Although, strictly speaking, animalism is just one among other biological accounts, we shall focus on this most influential view as the main rival of the psychological accounts. Each of us, animalists maintain, is essentially a human animal. The 'is' is not to be understood as a relation of constitution but as one of numerical identity.[139] Advocates of this account therefore assert that we are numerically identical with biological organisms of the species *Homo sapiens*.[140] Hence, we cannot cease to be human organisms without thereby ceasing to exist and we 'must be animals, and the selfsame animals, at all times we exist'.[141]

This is not to imply that *all* persons are necessarily animals. Animalism does not exclude the possibility of there being non-animal persons like extraterrestrials, angels, or robots.[142] Likewise, it does not mean that there could not also be human animals who are *not* persons – fetuses or the irreversibly comatose are commonly believed to belong to this category.[143] Consequently, animalism does not put forward necessary and sufficient conditions for the identity of persons per se; it is only concerned with persons of a particular kind, namely, human persons, and these are regarded as mere phases in the lives of organisms.[144]

According to animalism, we therefore 'have our criterion of identity by virtue of being organisms and not by virtue of being people. (…) [W]e persist, as other animals do, just in case our biological *lives* continue'.[145] In chapter 3, we investigated under which circumstances organisms continue to exist and under which circumstances they die. In most cases, an individual's psychological capacities and his or her biological life end simultaneously, for example, when a previously healthy patient suddenly dies from a heart attack. Sometimes, however, organisms lose their mental features at a time prior to their destruction. We have seen that such a situation occurs when a patient enters a persistent vegetative state. As follows from their criterion of personal identity, proponents of biological views hold that we survive profound changes like this; and they believe that, conversely, we would not survive situations in which solely our *mental* characteristics are preserved since they do not regard them as part of our funda-

138 See, for example, Blatti / Snowdon 2016; Blatti 2012; DeGrazia 2005; Wiggins 2003; Merricks 2001; Olson 1999; Olson 1997; Inwagen 1995; Snowdon 1990; Wiggins 1980.
139 Blatti 2012, 685. For authors who endorse constitutionalist positions, see note 160.
140 Noonan 2003, 196.
141 Snowdon 1991, 111.
142 Permitting non-human persons is, of course, not exclusive to animalism. See, for example, DeGrazia 2005, 4 and Baker 2000, 92.
143 Olson 1997, 96 f.
144 Noonan 2003, 197.
145 Olson 1997, 106.

mental nature. Advocates of psychological accounts, with which we will be dealing in the following section, deny both claims.

4.2.3 Psychological Accounts

Biological views of personal identity have been around for not more than three decades, which makes them newcomers among the other two centuries-old accounts.[146] Prima facie, this may seem surprising. Is it not obvious that we are organisms? Is this not what biology is telling us, and should this not therefore be the default position?[147] Consider the following widely shared belief.

> We appear to be able to imagine ourselves surviving the destruction of our bodies (…) as teletransported persons, who after a period of non-existence come to have a new body made from ambient matter. What is taken to survive in such imagined cases is something (at least partly) mental, either an entity or a series of events and states. In either case it is something that continues to exist even though the original organism, and the original animal, ceases to exist. So the imagination tells us that we do not have the same conditions of identity over time as mere organisms.[148]

Most people come to the same conclusion when they contemplate what would happen if two people exchanged heads. The subjects, it appears, would switch bodies, each one going with his or her respective brain. Considerations like these have convinced many authors that while we may not be immaterial souls, nor can we be *just* organisms. Rather, our persistence must consist in the continuity of some kind of psychological relation. According to this view, each of us was the past being whose mental features he or she has inherited; and he or she will be the future being who will be equipped with these mental features.

There is much disagreement, however, over what exactly these psychological characteristics may be. The most influential view is due to John Locke, who suggested that *memories* form the decisive relation as they enable a person to 'consider it self as it self, the same thinking thing in different places'.[149] Locke held that

> whatever has the consciousness of present and past Actions, is the same Person to whom they both belong. Had I the same consciousness, that I saw the Ark and *Noah*'s Flood, as that I saw an overflowing of the *Thames* last Winter, or as I write now, I could no more doubt that I, that write this now, that saw the *Thames* overflow'd last Winter, and that view'd the Flood at the general Deluge, was the same *self*.[150]

146 While I have allocated Aristotelian hylomorphism to the type of approaches that connect our persistence to that of souls, one may – depending on one's interpretation of Aristotle – also regard his view as an early ancestor of animalism. See Olson 2007, 172.

147 Olson 2007, 23.

148 Johnston 2016, 90.

149 Locke 2008, II.XXVII, § 9; see also Parfit 1987, 202.

150 Locke 2008, II.XXVII, § 16.

Hence, according to this criterion, a psychological subject *y* at t_2 is identical with a psychological subject *x* at t_1 if and only if *y* can remember having had *x*'s experiences, that is, if *y* can track herself through time and relive her past in a form of mental time travel.[151] Modified versions of this traditional account, so-called *neo-Lockean views*, continue to have many prominent advocates.[152] Unlike Locke himself, who only recognised autobiographical memories, they regard all kinds of standing mental states – beliefs, intentions, personality traits, and so forth – as constitutive of psychological continuity and, therefore, as the basis of personal identity.[153]

Instead of tying a subject's diachronic existence to long-term memories, one may also deem the persistence of mental *capacities*, like the capacity for consciousness, the crucial psychological relation. Such capacities are sometimes preserved even in the absence of the individual's memories, as we will see later. Examples of this type of view are Jeff McMahan's Embodied Mind Account and Peter Unger's Core Psychology Approach.[154] For the moment, the difference between these two subtypes of psychological views, the memory-based and the consciousness-based, need not worry us, but it will become important in chapter 5, where we shall draw finer distinctions.

The contrast between biological and psychological accounts is best explained by reference to the notions of substance sortal and phase sortal. The term *sortal*, originally coined by Locke and further developed by Wiggins, stipulates what the essence of a certain entity is, thereby specifying what changes it can undergo without ceasing to exist.[155] The essence of an entity is 'the very being of any thing, where-by it is, what it is'.[156] *Substance sortals* designate 'the sort of thing an entity *essentially* is – that is, the sort of thing it must be if it is to exist at all and thus the sort of thing it cannot cease to be without ceasing to exist'.[157] *Phase sortals*, on the contrary, specify a class of things to which the entity in question may belong for a certain period of time during its existence. It can become, or cease to be, a member of this class without ceasing to be what it fundamentally is, and without ceasing to exist altogether. Typical phase sortals are *teenager* and *kitten*.

According to biological accounts, the proper substance sortal for things like us is *organism*. As long as our organisms continue to exist, we continue to exist. Since human organisms can acquire mental characteristics, they can become persons. But they can also lose these characteristics again and still continue to exist. Persons are therefore merely phases in the lives of organisms.

Defenders of psychological accounts claim the exact opposite. They hold that the proper substance sortal for the kind of entity that we are is *psychological subject* or

151 Zeman 2006, 367.
152 These include Shoemaker 2008; Shoemaker 1999; Parfit 1987; Shoemaker / Swinburne 1984; Green / Wikler 1980; Lewis 1976; Perry 1972; Parfit 1971; Shoemaker 1970.
153 See section 5.2.
154 McMahan 2002, 66–94; Unger 1990; see also Gervais 1986.
155 Wiggins 1967.
156 Locke 2008, III.III, § 15.
157 McMahan 2002, 6.

person. We cannot come into existence until a psychological subject has developed, just as we cannot cease to be such a subject *and* continue to exist. Conversely, if it was possible somehow to sustain our mental characteristics in the absence of our organisms, we would persist, for being an organism is only a phase in our possible histories. Thus, proponents of psychological accounts usually accept the same persistence conditions for organismic life as animalists do. What they dispute is, rather, that these persistence conditions are *our* persistence conditions.

4.3 Deciding between the Accounts

In specifying different substance sortals, biological and psychological accounts of personal identity are mutually exclusive. We are *either* essentially organisms *or* essentially psychological subjects. How does one decide which view is correct? The traditional way of doing this is to construct hypothetical test cases that probe our intuition. As detailed in chapter 2, these scenarios usually take forms like: 'Can I coherently imagine myself existing as a brain in a vat?', 'If a blueprint of my body was sent to Mars, would I survive as the individual who leaves the teletransporter?', and so forth. We shall not follow this conventional thought-experimental approach, but make use of empirical facts and clinical data wherever possible.

To evaluate the accuracy of the neurological criterion of death, we have in the foregoing chapter investigated whether a body devoid of its brain can remain a functioning organism. To decide between biological and psychological views, we shall now pursue converse tactics by posing the opposite question: can a brain devoid of its original organism remain functional, that is, would it retain the mental features that psychological accounts of personal identity regard as the essence of our persistence? To reiterate, we are not asking whether this is *conceivable*, but whether this is in fact possible, technically and physiologically, in the actual world.

The rationale behind this approach is the following. As we have seen, biological accounts hold that each of us is numerically identical with his or her organism, whereas psychological accounts, assuming that each of us is identical with a mental entity, deny this. Hence, if one could show that an individual's mental capacities can continue detached from his or her organism in situations that do not differ much from those that we encounter in every major clinic around the world, this would be a very strong indication that psychological accounts have it right. If, on the other hand, it turned out that mental features can occur or continue only in conjunction with the individual's respective body, then biological views, according to which persons are only phases in the lives of organisms, would appear to have the better claim.

Needless to say, there are many more views on personal identity than the three that I presented, so that, strictly speaking, eliminating immaterial substance accounts does not reduce the available options to just two. Besides hybrid approaches that com-

bine features of several views into a single theory,[158] there are also views according to which we are material bodies,[159] persons constituted by organisms,[160] spatial or temporal parts of animals,[161] bundles of mental states,[162] or nothing at all.[163] While all of these accounts would be worth discussing, this is not the place to do so. The biological and the psychological views are not only by far the most prominent ones; they are also the only ones that have gained any recognition in the field of today's medicine and that are therefore of any potential practical relevance in the debate about brain death.[164] We will therefore limit our considerations to these two accounts.

For reasons of simplicity, we shall tentatively regard the preservation of consciousness as sufficient evidence of the persistence of a psychological subject, and thus of the entity that psychological accounts of personal identity postulate. As mentioned above, psychological views can be subdivided into those that emphasise the persistence of long-term memories and those that focus on the preservation of mental capacities, but we will temporarily ignore this difference while we are discussing the separation of brain and body. We shall, however, return to this division in chapter 5, where we will be focusing on cases in which the brain *itself* is modified.

Before we can establish whether consciousness can be preserved when the head is isolated from the organism, we must define what exactly we mean by this term. The Latin word *conscientia* is a compound of the prefix *cum*, which means *with*, and *scio*, which translates as *I know*. Originally, it referred to an 'inner witness', to someone with whom one shares one's knowledge.[165] Many authors from different disciplines have tried to give definitions of consciousness, but to date none of them can count as being universally accepted.[166]

In this thesis, we shall work with the definition that is used in clinical practice. It distinguishes two major aspects of consciousness that are easily conflated: its quantitative and its qualitative dimension. One may denote the former as *wakefulness* and the latter as *awareness*.[167] Unsurprisingly, the basis for someone to be conscious is for her to be awake – as opposed to being asleep or comatose. Under normal circumstances, wakefulness, which is also denoted as *arousal* or *vigilance*, is a daily recurring brain state. Wakefulness can be conceived of as the *level* of consciousness, which encompasses many intermediate stages: a subject may be somnolent, fully alert, or anything in between, just as one may be lightly or deeply anaesthetised.[168] Wakefulness is

158 Noonan 2003, 205–209.
159 Mackie 1999; Thomson 1997; Carter 1988; Williams 1970.
160 Baker 2007; Baker 2000; Johnston 1992.
161 Parfit 2012; Hudson 2007; Lewis 1976.
162 Campbell 2006; Hume 1896.
163 Unger 1990.
164 See, for example, Gervais 1986 and Green / Wikler 1980.
165 Lewis 2015, 181–191.
166 See, for example, Damasio 2010, 157–159; McGinn 2008, 237 f.; Zeman 2002a, 13–16; Block 1995, 227 f.; Nagel 1974, 436.
167 Multi-Society Task Force on PVS 1994, 1501.
168 Zeman 2006, 358.

therefore a graded notion. Neurophysiologically speaking, the process of awakening is the desynchronisation of cortical neuronal activity.

When one is awake, however, one is usually also aware *of* something. The stream of consciousness has a certain content, like a perception of an internal or external stimulus that one is having, an emotion that one is feeling, or a memory that one is recalling.[169] Being aware therefore means that in addition to having reached a sufficient level of wakefulness, certain processes are ongoing that give rise to a stream of consciousness.

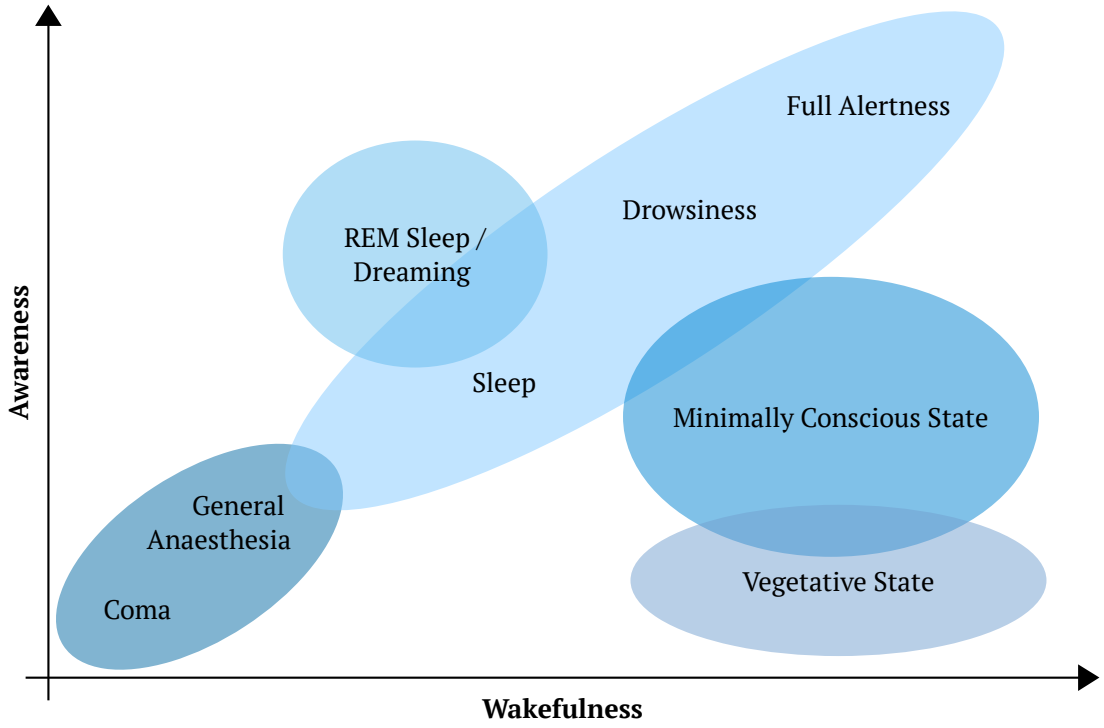


Fig. 1: The two dimensions of consciousness.[170]

How could one objectively measure these two dimensions of consciousness? In clinical practice, the most widely accepted standard for capturing the *quantitative* dimension of consciousness is the Glasgow Coma Scale. Eye, verbal, and motor responses that the patient exhibits are assessed and each assigned a certain score. The three values are then added up. Patients who achieve the highest possible score of 15 are considered fully awake. The lowest possible score is 3 and describes a deep coma.[171]

Awareness, the *qualitative* dimension of consciousness, is harder to grasp in objective terms. A subject's awareness can encompass very different things – whatever comes into the focus of her attention. Often, a subject is aware of many things simul-

169 Zeman 2001, 1265.

170 Illustration from Laureys 2005, 556; slightly modified.

171 How problematical any attempt at objectively assessing consciousness via clinical tests is becomes obvious when one considers conditions with severe motor impairments like locked-in syndrome (see section 3.3). Locked-in patients only achieve a very low score on the Glasgow Coma Scale although often neither the quantitative nor the qualitative aspect of their consciousness is diminished.

taneously. One can, for instance, watch the road while talking to the person in the passenger seat and at the same time be conscious of the pain in one's toe.

When the qualitative dimension of consciousness is slightly impaired, this content is altered in certain ways. The subject may be hallucinating, be disoriented, or be unable to act and communicate. The most prevalent causes of these conditions are intoxication, traumatic brain injury, and brain tumors. A very severe disturbance of the qualitative dimension of consciousness obtains when *any* experiential content is absent despite the subject being able to be aroused to a degree that would otherwise permit the manifestation of such content. This is what happens in the vegetative state. As detailed in section 3.3, patients in this condition open their eyes and exhibit sleep-wake cycles, but damage to the cerebral hemispheres precludes this arousal from being accompanied by awareness.

Thus, to count as conscious in our investigation, the subject shall be both awake (that is, not somnolent) and there shall be something *of* which she is aware. This may be any phenomenal or propositional content – the environment conveyed via the senses, mental imagery, feelings, memories, and so forth. This is one of the least demanding notions of consciousness. Unlike other conceptions, it does, for example, not require that the individual be self-conscious in the sense of being able to form a personal narrative or to relive her own past;[172] nor does it imply that the subject must possess higher-order mental states, like the reflection on a desire or on a perception.[173]

To determine whether consciousness according to this definition can occur in a head that is detached from its organism, we will be analysing all ways in which the brain and the rest of the body communicate, to ascertain what fraction of this exchange of information the brain can relinquish before either the quantitive or the qualitative dimension of consciousness breaks down. We will establish this by examining pathological conditions that result in drastic reductions of the amount of electrically and chemically coded information that brain and body exchange. Physiologically speaking, this approach is, in a sense, a mirror image of the analysis in chapter 3: while we were investigating the functional status of a brainless body, we are now looking into the functional status of a bodiless brain. Philosophically speaking, we are establishing the minimal supervenience base for consciousness.

172 Zeman 2006, 367; Baker 2000, 91 f.
173 Searle 1992, 142 f.

4.4 Isolating Consciousness from the Organism

4.4.1 Neurological Isolation

4.4.1.1 Anatomy and Terminology

The brain and the rest of the body exchange signals via two main channels: the nervous system and the endocrine system. Let us begin with an overview of the nervous system, which can be subdivided into the *central nervous system* and the *peripheral nervous system*. The former includes the three parts of the brain – the cerebrum, the cerebellum, and the brainstem – and the spinal cord. The latter comprises the somatic nervous system and the autonomic nervous system. The *somatic nervous system* carries information from the different sensory organs to the brain and relays motor commands to the muscles. Most of its actions are voluntarily initiated and conscious. Conversely, the *autonomic nervous system* operates predominantly without conscious direction. Its task is to control the many automatic functions that an organism has to perform through regulating the actions of its inner organs and adapting them to the requirements of different situations. The autonomic nervous system can be subdivided into the *sympathetic*, the *parasympathetic*, and the *enteric* nervous system. Roughly speaking, the sympathetic nervous system increases the activity of several organs, while the parasympathetic system decreases it. The enteric nervous system governs digestion.

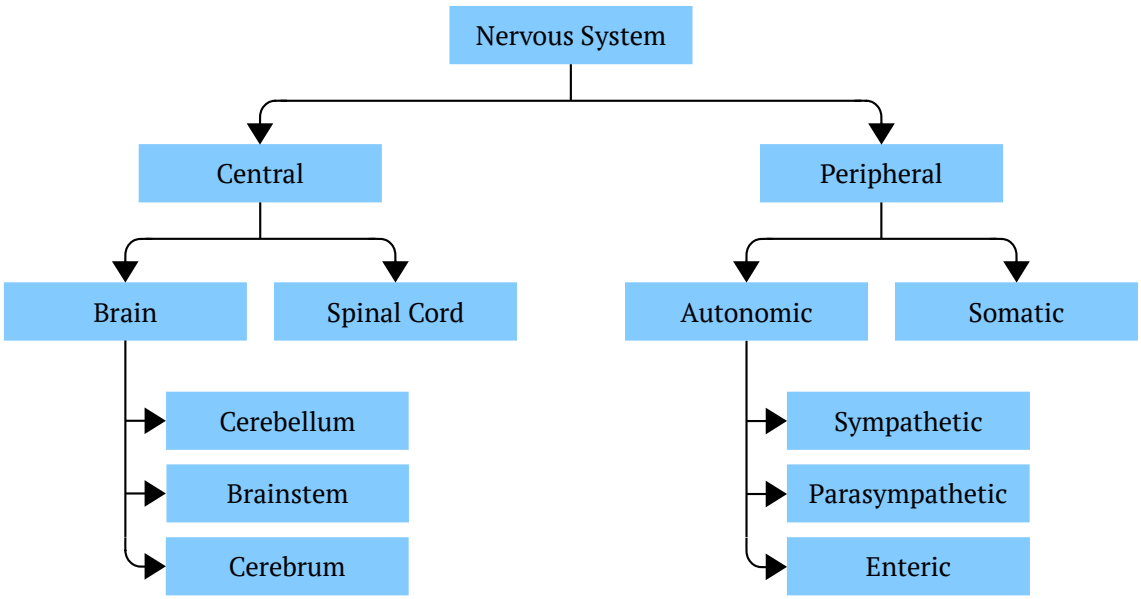


Fig. 2: The divisions of the human nervous system.

Information that the brain receives is called *sensory*. One may differentiate between two realms about which the brain receives sensory information: the environment (*ex-*

teroception) and states of the organism itself (*interoception*). While the former aids the organism's navigation in the world, the latter is important for maintaining homeostasis. Sensory data from the classical five senses, like seeing a tree or hearing the sound of a bell, belong to the realm of external perception, while being thirsty or having an elevated carbon-dioxide level in the blood stream are internal perceptions. The brain constantly receives signals about these two realms, which should not be conceived of as mutually exclusive: some stimuli, for example, the perception of a needle penetrating the skin, inform of a state of the organism (that a part of it is suffering damage) while simultaneously indicating a fact about the environment (that there is a pointy object present). Another common classification is that of somatosensory and viscerosensory input. *Somatosensory* information, which stems from receptors in the skin, the joints, and the skeletal muscles, enables the organism to perceive stimuli that arise from the external environment or from the position of the body. *Viscerosensory* signals are those that convey information about the status of the body's internal organs and help to maintain homeostasis.

Via the exteroceptive senses alone, the brain receives an amount of sensory data of 10^9 bits per second. Only 10^1 to 10^2 bits per second reach the level of conscious awareness. The remaining information is either processed unconsciously or immediately discarded.[174] The same is true of the interoceptive signals that the brain receives. Bodily needs that require the subject to act, like thirst or pain, reach awareness, while other internal parameters, like low arterial pressure, are being dealt with at an unconscious level. In many cases, stimuli are also handled unconsciously until their intensity reaches certain thresholds, at which point they finally become conscious.[175] Several experiments have shown that much of the data that is processed without the subject's being aware nonetheless exerts effects on judgements and actions. An example of this is the phenomenon of blindsight – a condition in which people who are cortically blind nonetheless respond to visual stimuli that they do not consciously see.[176] Hence, in investigating how much sensory input the generation of consciousness requires, one must not only take into account the impressions with which one is phenomenologically familiar but also all other stimuli that the brain receives.

The *motor* commands that the brain issues, that is, its output, also fall into two categories. They can be divided into those that trigger the contraction and relaxation of the skeletal muscles and thus direct locomotion and facial expressions (*somatomotor*), and those that govern the functions of the organism's internal organs, like breathing and heart rate (*visceromotor*). Somatomotor functions usually require voluntary de-

174 Silbernagl / Despopoulos 2009, 314. Bits per second (bit/s or bps) is a unit that quantifies in a binary digit (0 or 1) the amount of information that can be transmitted via a certain communication channel in one second. 10^9 bit/s equals 125 MB/s, which is the speed of the fastest internet connections available for private use today.

175 Fuchs 2017b, 301.

176 Zeman 2001, 1275–1277.

cisions to act on the part of the subject, whereas visceromotor functions are mostly executed automatically.

Type		Source | Target	Example
Input (*afferent*)	exteroception	about environment	hearing a noise
	interoception	about state of organism	monitoring the blood sugar level
Output (*efferent*)[177]	somatomotor	to skeletal muscles	raising one's arm
	visceromotor	to internal organs	increasing the heart rate

These are, in brief, the different domains about which the brain and the rest of the body exchange electrically coded information. For answering the question of whether consciousness can be generated without the brain being part of an organism, one must consequently ask how much, and what kind of, sensory input the brain requires and which motor signals it must be permitted to pass on. Once this minimum is established, one can determine whether or not the remaining anatomical structures still constitute an organism. Let us therefore examine what happens when this input and output is diminished.

4.4.1.2 Spinal Cord

The brain receives and sends electrical signals via two main routes: the spinal cord and the cranial nerves. We shall begin with the former. The spinal cord extends from the medulla oblongata in the brainstem to the lumbar region of the vertebral column and relays massive amounts of interoceptive and exteroceptive sensory information to the brain. These signals originate from the many tactile, thermal, pain, muscle, joint, and other receptors in the trunk and the limbs. In addition to these sensory pathways, the spinal cord also contains motor fibres for the voluntary control of the movements of trunk and limbs (somatomotor) and for the autonomic innervation of the inner organs (visceromotor).[178]

Does the status of consciousness change when these important routes of communication are severed and the brain is cut off from all the organs that the nerves arising from the cord supply? A lesion in the spinal cord can, for example, be caused by physical trauma. The higher the lesion, the more severely impaired is the communication between brain and body, and the greater is the number of functions that are lost. As discussed in section 3.3, cervical spine transection at the highest possible level (C1) leads to full quadriplegia: the paralysis of all voluntary movement below the neck. Since not only the motor pathways but also the somatosensory nerves that travel

177 *Afferent* nerves carry information from other parts of the body towards the brain. *Efferent* nerves transmit impulses from the brain towards other organs.

178 Crossman / Neary 2015, 69.

through the spinal column are transected, the brain is moreover deprived of all sensory information originating from the trunk and the limbs. Due to the fact that its fibres emerge from the thoracic and lumbar levels of the spinal cord, the defect also includes the sympathetic division of the autonomic nervous system. This leads to the brain becoming unable both to receive electrical information about the status of the inner organs (viscerosensory) and to control their function (visceromotor). In consequence, the organism loses, for example, the ability to breathe on its own and it becomes incontinent. As *para*sympathetic fibres, too, emerge from the sacral level of the spinal column (S2-S4), one of the two main branches of parasympathetic control is also abolished.[179]

In conclusion, the decrease in electrical signalling between the head and the rest of the body following high spinal-cord transection is immense. Naturally, the onset of such a condition changes a person's life in a profound way. However, there is no evidence suggesting that consciousness in quadriplegic people is diminished or altered in any way. Neither the tremendous deficit in sensory input nor the impossibility of sending motor commands to target organs via the spinal route alters the quantitative or the qualitative dimension of consciousness.

4.4.1.3 Cranial Nerves IX, X, and XI

This result, impressive as it may seem, does not yet show that neurologically separating head and body is possible since, in addition to the spinal cord, twelve cranial nerves attach to the brain that also transmit signals to and from several peripheral structures. Most of these structures are located in the head: the sense organs for the reception of visual, auditory, olfactory, gustatory, tactile, and vestibular stimuli as well as target organs for motor output, like the muscles for facial movement, the movement of the eyeball, or the elevation of the lid. Besides the sensory (I, II, V, VII-X) and motor (III-VII, IX-XII) pathways, cranial nerves III, VII, IX, and X also carry fibres that belong to the parasympathetic division. They enable automatic functions like lacrimation and salivation.

The spinal part of the accessory nerve (XI), the glossopharyngeal nerve (IX), and the nervus vagus (X), however, do not terminate in the head but pass through the neck. In order completely to neurologically isolate the head, one must therefore investigate how the brain reacts when these three nerves are severed. The accessory nerve contains exclusively motor fibres, which innervate the sternocleidomastoid and trapezius muscles. These muscles rotate and tilt the head and are responsible for shrugging one's shoulders. In radical neck dissections, which are surgical procedures to remove lymph nodes from the neck into which cancer cells may have migrated, the accessory nerve must sometimes be sacrified. The sternocleidomastoid and trapezius

179 The vagus nerve, the other main parasympathetic branch, does not travel inside the spinal column and is therefore unaffected by this type of injury.

muscles of these patients become paralysed, but apart from being unable to abduct their shoulders beyond 90 degrees, no other effects occur.[180]

Unlike the accessory, the glossopharyngeal nerve carries both motor and sensory information. It supplies the stylopharyngeus muscle and provides parasympathetic innervation of the parotid gland. Moreover, it receives sensory input from the carotid sinus and carotid body, the pharynx, the posterior third of the tongue, and the middle ear.[181] Bilateral glossopharyngeal nerve paralysis, which can occur as an extremely rare complication of tonsillectomy, therefore leads to dysphagia, numbness, decreased taste sensation at the posterior aspect of the tongue, nasal regurgitation of liquids, and a nasal voice quality.[182] There is no indication that the transection of this nerve exerts any influence on consciousness, however.

The case of the nervus vagus is more complicated. As described in section 3.3, the vagal motor pathways are responsible for the parasympathetic control of several organs, where they cause, in most cases, a decrease in activity. The majority of vagal fibres, however, are afferent and transmit sensory information about the state of the lungs, heart, liver, stomach, pancreas, and bowel from visceral receptors to the brain.[183] Since the vagus is not only the main parasympathetic nerve but also the largest visceral afferent nerve, its destruction entails a drastic reduction in the signals that reach the head.[184]

As the vagus belongs to the autonomic nervous system, the visceral information that it carries is usually processed unconsciously. This does not mean, however, that these signals do not exert any psychological influence. The vagus nerve is closely associated with the emotional state of mind. Temporary loss of bladder control when experiencing extreme fear, for example, is the result of an overly strong parasympathetic response, mediated by the vagus. That there exists a connection between an individual's mood and vagal input to the brain is also confirmed by favourable therapeutic outcome: some forms of depression have successfully been treated with vagus nerve stimulation.[185]

How is it possible that psychological parameters are susceptible to vagal input? The vagus nerve is the main connection between the central nervous system (CNS) and the enteric nervous system (ENS). The two divisions are 'welded together in a neural network in which outflow from the CNS may have pervasive effects on ENS operations, and the ENS may have extensive inputs to the CNS by way of the vagal afferents'.[186] The ENS, which controls the function of the gastrointestinal tract, contains as many

180 Montgomery / Evans / Gullane 2009, 514; FitzGerald / Gruener / Mtui 2012, 219.
181 Crossman / Neary 2015, 103.
182 Ford / Cruz 2004, 2196 f.
183 Zagon 2001, 671.
184 FitzGerald / Gruener / Mtui 2012, 221.
185 This procedure involves the subcutaneous implantation of a small device resembling a pacemaker, which stimulates the vagus electrically in fixed intervals (Corcoran et al. 2006, 282).
186 Powley 2000, iv32.

neurons as the entire spinal cord, which is why it is often referred to as the *gut brain*.[187] A simple example of the kind of influence that the ENS exerts on mood is the effect of food consumption:

> The commonest way in which food can affect behaviour is the change in mood and arousal that occurs from before to after eating a meal. (…) Many animals, including human beings, tend to be aroused, alert and even irritable when hungry. (…) After eating a satiating meal, we and other animals typically become calm, lethargic and may even sleep, and mood is more likely to be positive than negative. Nutrient absorption is rapidly detected by the brain, as afferent information is conveyed by the vagus nerve from the gut and liver. The potential influence of this internal information route on emotional behaviour is beginning to be acknowledged.[188]

A 2011 study investigated the interaction between nutrient-induced ENS signalling and sad emotion engendered by musical and visual cues. The subjects were given an intragastric infusion of either fatty acid solution or saline during neutral or sad emotion induction and were asked to rate sensations of hunger, fullness, and mood. The study found that the infusion of fatty acid attenuated both the neural as well as the behavioural responses to the induction of sad emotion.[189] There is also evidence suggesting that visceral vagal input modulates the responsiveness of forebrain neurones and thereby even influences the interpretation of exteroceptive stimuli.[190]

Consequently, unlike the transection of the spinal cord or of the other cranial nerves, severing the vagus is likely to modify psychological properties to a certain degree as it removes the influence that the ENS exerts on the brain.[191] However, this does not mean that the qualitative dimension of consciousness would be altered to an extent incompatible with the standards that we have set for assessing whether consciousness is still being generated. What we require is that there be something of which the subject is aware, that is, that her stream of consciousness has a certain content. Whether this content is somehow modified, for instance, so that changes in mood occur, is not decisive.

When the two main trunks of the abdominal vagus nerve are surgically divided in a bilateral truncal vagotomy, the ENS becomes disconnected from the parasympathetic nervous system, thereby drastically reducing ENS input to the brain. While vagotomies often cause digestive problems, there is no clinical evidence of significant psychological effects.[192] This appears to be the case even when the vagus is not transected

187 FitzGerald / Gruener / Mtui 2012, 155.
188 Gibson 2006, 54.
189 Oudenhove et al. 2011, 3094 f.
190 Zagon 2001, 672.
191 Although the digestive system also receives *sympathetic* innervation, namely, from the thoracal and lumbal segments of the spinal cord, it appears that 'gut-brain signaling occurs mainly via the neural route through CCK-induced activation of vagal afferents' (Oudenhove et al. 2011, 3097).
192 Clark et al. 1964, 902 f.; Martin 2015, 3088; see also note 100.

where it enters the abdomen, but at neck level.[193] Vagal input to the brain, it therefore seems, is not a necessary condition for the latter to generate consciousness.[194]

We have now (1) identified all neural pathways that transmit electrical signals between the head and the rest of the body, namely, the spinal cord and three cranial nerves; (2) shown that the transection of the spinal cord results in a drastic reduction of both sensory information that reaches the brain (somatic and visceral) and motor signals that the brain can send to the body (somatic and visceral), but that this reduction does not affect the generation of consciousness; (3) established that the influence on brain function of two of the three cranial nerves, the accessory and the glossopharyngeal, is negligible; and (4) determined that with the transection of the remaining nerve, the vagus, the brain loses an important source of autonomic nervous system input. While studies suggest that the absence of the latter may have an effect on the individual's mood, there is no indication that the generation of consciousness *per se* depends on vagal input or that the subject's mental experience is modified beyond changes in emotional hue.

Hence, even a drastic reduction in electrical communication between head and body would not lead to a breakdown of consciousness or to a marked reduction in wakefulness or awareness. The subject would not be somnolent or comatose (quantitative criterion); nor would the content of the stream of consciousness be eliminated (qualitative criterion). Presumably, the brain can relinquish such a large portion of its sensory input without significant changes in consciousness only on the condition that there is still a connection to the environment through the sensory pathways of the remaining cranial nerves. While disorders like blindness or loss of smell, or combined losses like deaf-blindness, show that not even all senses mediated by the cranial nerves are required for the brain to be awake and aware, this does not license the conclusion *all* sensory input would be dispensible simultaneously.[195] The generation of consciousness may well depend on incoming information, but not on this information stemming from organs below the neck.[196]

193 Ghaemmaghami et al. 2008, 846. In interpreting this article, however, one must take into account that a unilateral transection only affects half of the vagal fibres.

194 Interestingly, stimulating the vagus nerve appears to have the potential to improve the level of consciousness in the persistent vegetative state, however (Corazzol et al. 2017, R994).

195 Studies that compared the sensory construction of dreams in congenitally blind and late blind patients found that patients born blind lack visual content in their dreams, while those who went blind at a later point in their lives do experience such content. The longer an individual has been deprived of visual perception, the shorter and poorer the visual content becomes in his or her dreams (Meaidi et al. 2014, 592). 'Our own experience provides the basic material for our imagination, whose range is therefore limited', remarks Nagel (1974, 439).

196 A very illuminating source for understanding the effects of sensory deprivation on consciousness are John C. Lilly's experiments with isolation tanks. Lilly sought to create an environment that would achieve the greatest possible attenuation of all external stimuli that impact on the body, including light, sound, odor, taste, pressure, gravity-opposition forces, temperature, and so forth. To this end, he constructed a soundproof water-immersion tank. Before the subject enters the tank, internal stimuli like hunger, thirst, or pain are also minimised to the maximum possible degree. In this setting, both exteroception and interoception is dramatically reduced (Lilly 1977; Lilly / Shurley 1961).

The degree to which the brain can relinquish motor output and thus its ability to act on the world is even more astonishing. In section 3.3, we already discussed locked-in syndrome, which results from bilateral ventral pontine damage and is characterised by an inability voluntarily to move any muscles with the exception of the eyelid. If a code can be established, the patient can communicate with her environment via blinking, yet no other movements are possible. Still, wakefulness and awareness are preserved. Even *total* locked-in syndrome, which deprives the patient also of this last transmission channel, is characterised by 'preserved consciousness i.e. existing inner monologue and awareness of external and internal stimuli as far as the corresponding pathways for sensory perception are spared by the lesion'.[197]

Another case in point is British scientist Stephen Hawking, who was suffering from motor neurone disease, which causes muscle weakness. Victims of this disorder gradually lose the ability to initiate voluntary movement. In Hawking's case, the disease had paralysed nearly every muscle in his body. In the final years before his death in 2018, Hawking was only able to interact with the world through moving his cheek. The motion was detected by an infrared switch, which enabled him to select signs on a spelling card. The rate at which he could express himself via this single channel was one word per minute.[198] Given that the average word length in the English language is 5.1 characters, and that the alphabet comprises 26 letters, Hawking could communicate at a rate of only 0.4 bits per second.[199] Through speech, facial expression, and gesturing, a healthy human subject can convey up to 10^7 bits per second to the environment.[200] Consequently, Hawking was able to transmit to the world only 0.000004% of the information that ordinary individuals can convey within the same time span, yet his mind remained sharp.

From Hawking's case and from locked-in patients, one can learn that although coping with such a situation most certainly requires an extraordinary amount of patience and self-determination, the generation of consciousness continues even when the brain can command only minimal motor output. Being able to communicate just a tiny fraction of the information that a healthy organism, equipped with speech, facial expression, and body language can convey to the environment is sufficient. In the setup that we are considering in this section, 43 facial muscles would still be available

197 Bauer / Gerstenbrand / Rumpl 1979, 84. Total locked-in syndrome brings with it the epistemic problem of how one establishes the presence of consciousness in a fully de-efferented patient. A normal waking and sleeping EEG may not be sufficient evidence. Currently, researchers explore whether conveying information via f-MRI is feasible in these patients. Functional neuroimaging visualises brain activity by detecting changes in blood flow. By performing mental imagery tasks, thereby increasing the activity in a specified brain region, individuals that are deprived of all motor output could respond to questions when their brain activity patterns are translated into affirmative or negative answers (Naci / Owen 2013).

198 Lange 2012.

199 The result of this calculation depends on whether or not one includes digits and non-alphanumeric characters.

200 Silbernagl / Despopoulos 2009, 314.

for this task in the isolated head. This greatly exceeds the means of expression that both Hawking and locked-in patients have at their disposal.

Since we have now considered all channels for the transmission of electrically coded information that connect the head to the body and deemed them dispensible with regard to the generation of consciousness, we can conclude that neurologically isolating a conscious head is possible. Does this establish the separation between organism and consciousness that we require for deciding between biological and psychological views of personal identity? Not quite. *First*, the fact that head and body can be *neurologically* disconnected is not tantamount to a total standstill of communication. Even in the case of complete neurological separation, information is constantly being exchanged via another channel: the endocrine system. It may therefore still be the case that consciousness is dependent on hormones that the organism produces. *Secondly*, the body supplies the head with oxygen and nutrients. As long as this is so, one may argue that the brain is still part of the organism, regardless of whether or not the two entities can communicate. And *thirdly*, one might object that an isolated head is actually an organism in its own right. Although this position is not a very plausible one, it is what some animalists assert. We will now take these remaining steps of the argument, beginning with the first one.

4.4.2 Endocrine Isolation

Hormones are regulatory substances that are secreted by several glands, like the pituitary, the adrenal, or the thyroid. Via the blood stream, they are carried to their respective target sites, where they elicit various reactions. The production of hormones begins early in utero and continues throughout an organism's life. In section 3.3, we examined the effects of hormonal disturbances on the brain-dead body. Now we are interested in the influence that hormonal deprivation exerts on the brain or, more specifically, on consciousness.

The brain acts as both a secretor of hormones and an endocrine target. Hormones that are produced in the body can cross the blood-brain barrier, which separates the central nervous system from the peripheral tissues. Consequently, some of these hormones – those to which the brain is sensitive – can influence its function. Let us therefore consider cases in which the levels of these hormones are altered and assess their respective impact on consciousness.

A state during which hormonal imbalances occur naturally is pregnancy. In pregnant women, progesterone and glucocorticoid levels are usually elevated. In consequence, they may experience a temporary decline in memory performance, especially in the third trimester.[201] This does not license the conclusion that an absence of these hormones would therefore necessarily be beneficial, however. While chronic exposure to glucocorticoids results in hippocampal atrophy, the opposite effect – an increase in

201 Brett 2001, 354 f.

dendritic arborisation and spinogenesis of neurons – occurs in the amygdala.[202] This shows how difficult it is to assess the effects that hormones have on the brain.

Natural hormonal fluctuations are also related to menstrual cycles, which are known to exert an influence on mood and emotional processing. Low oestrogen levels are often associated with negative feelings, while high oestrogen levels have positive modulatory effects on neurotransmitter systems that are involved in the regulation of affective behaviour.[203] Correlations between mood and the levels of certain hormones were also found in men. It was, for example, shown that depression coincides with the age-related decline in testosterone levels.[204]

Moreover, there is some indication that an association exists between testosterone levels and aggressive behaviour, which would be an example of hormones modulating personality traits. Collected data from chemically castrated sex offenders demonstrated that the drastic reduction in testosterone that they experience resulted in a suppression of sexual drive and fantasies in the majority of cases. This correlation did not obtain in all subjects, however. In some individuals, even castration levels of testosterone did not lead to measurable behavioural effects.[205] A very comprehensive recent study judged the link between testosterone and aggression to be 'relatively weak'.[206]

Studies seeking to determine the influence of androgens on cognitive function yielded conflicting results – some indicating that low levels of testosterone can reduce cognitive performance across ages, others finding an inverse correlation.[207] Oestrogen, the primary female sex hormone, is believed to enhance cognitive functions that are mediated by the hippocampus and frontal lobes, like learning and the processing of memory.[208] It also appears to have neuroprotective effects as well as to encourage hippocampal neurogenesis.[209] A possible explanation for this is the modulating role that oestrogens exert on several neurotransmitter systems, including the acetylcholine, the catecholamine, and the serotonin systems.[210] Deprived of the influence of oestrogen, these positive effects are believed to decline.

These are, in short, the endocrine mechanisms whose modification can reasonably be expected to have the greatest impact on brain function. What are we to conclude from analysing them regarding our question as to whether a brain deprived of endocrine input from the rest of the body would be able to generate consciousness?

The first thing to note is that this question is ambiguous. If what is meant is whether, devoid of the influence of hormones, a brain could *develop* to a stage at which

202 Tatomir / Micu / Crivii 2014, 5.
203 Gasbarri et al. 2012, 599.
204 Mahmoud / Wainwright / Galea 2016, 137.
205 Koo et al. 2013, 565.
206 Geniole et al. 2020, 8.
207 Ali / Begum / Reza 2018, 37.
208 Voytko et al. 2009; Maki 2005.
209 Mahmoud / Wainwright / Galea 2016, 138.
210 Gasbarri et al. 2012, 587.

it can sustain consciousness, the answer must clearly be negative. Maternal hormones – oestrogen and progesterone – are required to stimulate neuronal growth in the fetus. Thyroid hormones are also essential for brain maturation as they promote neurogenesis, neuronal migration, glial cell differentiation, myelination, and synaptogenesis. Thyroid hormone deficiency during fetal and postnatal development, which occurs as a result of hypothyroidism, can therefore lead to intellectual deficits.[211] Importantly, without hormonal influence not only the brain itself but the entire organism cannot develop appropriately. Since brain maturation is intimately intertwined with the growth of the organism as a whole, a functioning endocrine system is absolutely crucial.

If, however, the question is whether a *mature* brain can relinquish constant hormonal input while still generating consciousness, the available evidence seems to suggest that the answer is yes. The hormones that aid in brain development are required during *specific time windows*. While, for example, severe thyroid hormone deprivation of early onset is associated with neurodevelopmental retardation, this is usually not the case when the condition manifests later in life.[212]

Taken together, the findings appear to indicate that when the developed human brain stops being subjected to hormones that are produced in other parts of the body, the two domains that are mainly affected are mood and memory. As some hormones are believed to have neuroprotective effects, their absence will likely increase the risk of neurodegenerative disorders like Alzheimer's or Parkinson's disease. Consequently, one should expect a brain that is detached from the rest of the body's endocrine system to deteriorate at a higher pace. While the long-term impact of hormonal deprivation on brain function can therefore indeed be drastic, there is no indication that the generation of consciousness would cease immediately.

It is important to emphasise that the empirical reliability of the conclusions drawn in this section does not match that of the preceding ones. This is so for two reasons. First, hormone actions in the brain are extremely complex and only relatively poorly understood. The effects that a single hormone has on the brain's fine-tuned chemistry depend on its dose, the concentration of other hormones that are also present, the receptor type and density, the time of exposure, and on a multitude of other factors.

Secondly, when we were considering the severing of neural pathways, our conclusions rested on unambiguous clinical cases. When, for example, the spinal cord is completely transected in a quadriplegic patient, no electrical signals that it would normally transmit can possibly enter the brain. This is a very reliable basis for drawing medical and philosophical conclusions. In contrast, there are no pathological conditions in which endocrine activity would be totally eliminated. Usually, certain types of hormones are still being produced even in patients who suffer from the most severe

211 Bernal 2015.
212 Prezioso / Giannini / Chiarelli 2018.

imbalances. Predicting the effects of the simultaneous absence of all chemical messengers that the brain does not secrete itself is therefore very difficult.

What one can safely conclude from the available studies, it seems to me, is that the main influences of the hormones to which the fully developed adult human brain is sensitive are in the domains of mood and memory. There is no indication that hormones are directly implicated in bringing forth consciousness. Consequently, a breakdown of chemical communication between the brain and the rest of the body – just like the transection of the main neural pathways – appears compatible with the generation of consciousness.

Let us pause here for a moment and take a step back. We have now established that a brain which is entirely cut off from all routes of information exchange with each and every organ below the neck would continue to give rise to consciousness. This is an astonishing result. The head would, of course, still be physically attached to the torso. But this may not be decisive since the brain would, as it were, be isolated *within* the organism that surrounds it. That, in spite of this fact, the brain would still be able to bring forth a stream of consciousness is a strong indication that a psychological subject, the entity that advocates of psychological accounts of personal identity regard as the essence of our being, can exist independently of his or her organism.

4.4.3 Vascular Isolation

Like any other organ or tissue in the body, the head, and especially the brain, is dependent on a constant supply of oxygen, nutrients, and a multitude of other substances. Even when one has eliminated all means of communication between head and body, the latter still supplies the former. Without being connected to the body's circulation, the cerebral tissues would irreversibly lose their function within a few minutes. Other structures in the head would follow. One may therefore argue that the head and the rest of the body would remain parts of the same organism even if the two entities could not communicate – neither via the nervous- nor via the endocrine system. The final step in our inquiry must therefore be to establish whether, in addition to neural and endocrine isolation, vascular isolation of the head from the original body could also be achieved. Two approaches suggest themselves: transplanting the head onto another human organism, and exchanging the body for inorganic devices.

4.4.3.1 Head Transplantation

The idea of equipping an individual's head with a different organic body has been fascinating people for centuries and has featured in numerous fictional stories and films.[213] In the late 19th century, French neurologist Jean-Baptiste Laborde was prob-

213 Mary Shelley's *Frankenstein* (1818), Alexander Beliaev's *Professor Dowell's Head* (1925), Thomas Mann's *Die vertauschten Köpfe* (1940), and Roald Dahl's *William and Mary* (1960) are just some examples among many.

ably the first systematically to experiment on severed human heads, which he obtained from executions in Paris. Laborde connected the transected arteries to the cardiovascular systems of living dogs in the hope of being able to restore circulation in the heads. However, no spontaneous movements or other signs of consciousness could be elicited.[214]

Transplanting a head onto another organism poses three main problems, which we will analyse in this section: achieving vascular anastomosis before the onset of ischaemia, connecting the severed nerves to the new body, and suppressing immunologic mechanisms of rejection. We shall conclude the section by applying the results to biological views of personal identity and considering objections that their proponents might raise. Let us begin with anastomosis.

Prima facie, it may appear as if the timely reconnection of the blood supply to a severed head would be an unsurmountable problem. Unlike most other organs, which, if cooled appropriately, can be conserved for several hours before they are finally implanted, cerebral tissue is extremely vulnerable to oxygen deprivation. The brain's energy requirements per kilogram are higher than those of any other organ in the human body, but it possesses virtually no reserves of nutrients that are critical to its functioning.[215] Through two large pairs of blood vessels – the common carotid arteries, which branch into the external and the internal carotid arteries, and the vertebral arteries – an adult's brain receives approximately 750 ml of oxygenated blood per minute. Via the internal and external jugular veins and the vertebral veins, carbon dioxide and other metabolic products are constantly being removed. When the brain's oxygen and glucose supply is cut off, irreversible damage to the cerebral tissues begins to ensue after only three minutes.[216] Achieving vascular anastomosis within this time frame is impossible.

Even the great metabolic demands of the brain, however, can be lowered by cooling. The time window that opens up is just shorter than in the case of other organs. For decades, doctors have been employing deep hypothermia in aortic surgery and other operations in which a total stoppage of the cerebral circulation is unavoidable.[217] At a temperature of 12.5 °C, often even higher, electrocerebral silence occurs.[218] In this state, the brain's metabolic rate – chiefly the consumption of oxygen and glucose – is drastically reduced. The interval available for keeping the brain unperfused before neurological damage ensues is now increased by a factor of ten compared to normothermia. Operating times of up to 40 minutes become possible.[219] If the procedure is carefully planned, this time window should be long enough to connect the carotid and vertebral arteries as well as the jugular and vertebral veins to the receiving body.

214 Brukhonenko / Tchetchuline 1929, 33.
215 Plum / Posner 1980, 193 f.
216 Silbernagl / Despopoulos 2009, 130; see also section 5.3.2.
217 An early application of this technique is described in Niazi / Lewis 1958.
218 Stecker et al. 2001, 20.
219 Chau et al. 2013, 1558.

Does any empirical evidence support this supposition, given that such an operation has never been performed in humans? In 1954, Soviet scientist Vladimir Demikhov transplanted a canine upper thorax to the neck of another dog, thereby creating a two-headed animal. Both heads appeared alive and were eating and moving separately. Demikhov repeated the operation several times, but since he could not control the immune rejections that occurred in the bodies, the longest recorded survival time was only 29 days.[220] In an equally gruesome and heavily criticised animal experiment, Robert White and colleagues exchanged the heads of rhesus monkeys. They observed that

> from 1 to 3 hours after transection, the cephalon awakened and demonstrated a persistence of awareness of its environment during an eight hour period of study. These preparations were capable of vocalizing, accepting and chewing food, tracking with their eyes, and biting if orally stimulated. Throughout the period of observation the EEG record reflected the established characteristics of an awake pattern provided arterial pressure was maintained. (…) Two of these models were prepared simultaneously and following the return of consciousness and normalization of the EEG, the cephalons were vascularly interchanged and connected to the other isolated body. (…) [N]o deteriorization in the performance of the brain was observed even though each cephalon was then being perfused by the other animal body.[221]

In interpreting the results of these ethically highly questionable experiments, one should bear in mind a difficulty that we have touched on earlier, namely, the problem of detecting consciousness from a third-personal perspective in the absence of verbal reports.[222] In this section, however, we are only interested in whether the temporary break in circulation that is required for performing a head transplantation can be bridged. Uneventful EEG readings combined with behavioural reactions typical of animals that are awake and aware are undeniably acceptable pieces of evidence of the successful preservation of cerebral tissue.

The timely vascular anastomosis of transplanted heads, we can conclude, will most likely be possible. This brings us to the second problem that we have identified: the reconnection of the nerves that must be severed when the head is removed. As described in section 4.4.1.3, these are the spinal cord as well as cranial nerves IX, X, and XI. In the animal experiments discussed above, no attempts were made at reattaching any neural pathways. Although medicine has made some progress at regrowing smaller nerves, rejoining transected spinal cords, especially when the stumps belong to different bodies, is still very far in the realm of fiction.[223] In practice, transplanting a human head would therefore not be a worthwhile procedure. The result would be a quadriple-

220 Konstantinov 2009, 456.
221 White et al. 1971, 602 f.
222 See note 197.
223 Only a very small minority within the scientific community (cf. e.g. Canavero 2013) believes that medicine will soon possess the means to achieve this.

gic individual who is forever dependent on a ventilator. From the pathological conditions analysed in section 4.4.1, we know that the nerves that must be severed in this procedure need not be intact for consciousness to manifest, however. The subject would therefore indeed be conscious, and since all sensory and motor functions that are located in the head would not be affected by the transection of neural pathways at neck level, she could even express herself using her facial muscles, albeit not verbally, for her vocal cords would, of course, be paralysed.

The final difficulty is that of managing the immune response. If the receiving body's immune system could not be kept from rejecting the foreign tissue, it would not tolerate the transplanted head any longer than it was the case in the discussed animal preparations. While one can nowadays successfully suppress immune reactions following the transplantation of most organs and tissues, including even in combined heart-lung transplantations, the obstacles to a full head transplantation will naturally be greater. Since the underlying mechanism is the same, however, these obstacles appear surmountable.[224]

In any case could one entirely circumvent this difficulty by using isografts instead of allografts. Allografts are organs or tissues transferred between two genetically different individuals of the same species. Isografts stem from genetically identical individuals like monozygotic twins. Due to the total histocompatibility of donor and recipient, adverse immune reactions to isografts occur virtually never, which means that the head would be tolerated even without any postoperative immunosuppressant therapy. Hence, at least for head transplantations performed on monozygotic twins, the third major problem disappears.

Very few people who are in need of a transplant enjoy the luxury of having at their disposal an identical twin whom they can sacrifice.[225] Our question, however, is one of principle. Consequently, it is not decisive whether a transplanted head would be functional for more than a few days, whether a cephalic transplantation would benefit the patient, or whether such an operation would be an ethically justifiable undertaking. The crucial point is solely whether consciousness would irreversibly cease upon the separation of the head from its original body or whether it could be transferred to the new environment. Even if the receiving body eventually rejected the head, the latter would still have been awake and aware for a significant amount of time – in the absence of the rest of the individual's original body.

We have now established that vascular anastomosis is achievable by means of hypothermia, that the nerves that need to be severed and cannot be reattached are not essential for consciousness to manifest, and that there are constellations in which no immune rejection would occur. Hence, we have eliminated all three problems that we

224 Barker / Frank / Leppik 2015, 614.

225 The idea of creating clones who act as involuntary donors of isografts has been explored in science fiction films like *The Clonus Horror* (1979) or *The Island* (2005) and in Kazuo Ishiguro's novel *Never Let Me Go* (2005).

identified and can therefore conclude that transplanting a head – for the purpose of maintaining consciousness – is indeed possible.

Proponents of biological views of personal identity will have a hard time explaining this empirically grounded result. If, as they maintain, each of us is numerically identical with his or her organism, our persistence conditions and those of our organisms must be congruent. Each of us continues to exist if and only if his or her organism continues to exist. Consequently, if patient *x*'s head was removed and successfully connected to the circulation of patient *y*'s organism while *x*'s body was left to die, patient *x* would, according to this doctrine, cease to exist – despite *someone* being awake and aware in his head. That this could be an accurate description of the result of this operation is hard to believe. The conscious head would, of course, emphatically deny that he has ceased to exist. If patient *x* was Albert Einstein and patient *y* was a young philosophy student, would it now really be *y* who effortlessly solves complicated field equations in Einstein's very own head? And would one really deem it appropriate to punish *x*'s head for a crime that *y* has committed?

There is only a single way for the advocates of biological accounts to evade this absurd implication of their view. They could submit that while it may indeed be the case that one can isolate a fully functional head from its original body, this does not show that a centre of consciousness and an organism can be separated, but rather that an organism can be diminished to its core, which is its head. When attached to a new body, this core does not itself become part of a different organism, but it is simply equipped with additional organs. Our intuition, animalists and other proponents of biological accounts could concede, is correct: we go wherever our heads go; but it is correct for the wrong reason. Not the brain's *mental* features are responsible for our continued existence in the absence of the rest of the body, but the fact that the head is the core of the *organism*. The organism goes where its head goes. Hence, all that we have established, defenders of biological approaches may insist, is what the essential anatomical parts of organisms are.

This manoeuvre, if succesful, would completely remove the challenge that the foregoing sections pose. Unsurprisingly, leading animalists Eric Olson and Peter van Inwagen indeed take this position. Olson maintains that 'a detached head is a (…) severely debilitated animal. (…) [T]he head seems to be a living organism because it retains the capacity to coordinate its vital functions'.[226] Similarly, van Inwagen asserts that a 'severed head is a genuine living organism and the headless body is not'.[227] Thus, when patient *x*'s head is severed and successfully connected to the circulation of patient *y*'s headless organism, patient *y* ceases to exist while patient *x*'s organism just becomes larger.

226 Olson 1999, 133.

227 Inwagen 1995, 177. Van Inwagen's claim is based on his revisionist ontology, according to which organisms are the only genuine complex wholes because only particles that constitute a biological life compose things. All other supposed entities, including brains, are only simples, arranged in a certain way. We cannot here discuss this ontology. For a short critique, see Madden 2016, 183.

We should reject this proposal for two reasons. The first is a physiological one; the second reason has to do with the transitivity of identity. Let us begin with the physiological objection. The explanation that Olson provides for equating the head with the organism is that after having been detached from the body, the head would 'continue to regulate its internal metabolic activities in its characteristic way until the "control centers" shut down owing to lack of oxygen'. The severed head would thus 'behave like a dying organism', whereas the rest of the body would not.[228] Van Inwagen makes a similar claim.[229] Not only is it questionable, however, whether the cascade of pathological events that occur in a dying organ should be regarded as indicative of this organ's mereological status in an otherwise larger system; the claim is also physiologically false, for it is not the case that a detached head would be able to regulate any metabolic activities since nearly all target organs responsible for carrying out such tasks are located in the trunk. To give just one example: in order to compensate for the sudden drop in arterial pressure when the neck is severed, the cardiovascular centre in the brainstem would attempt to increase the heart rate by sending nerve impulses to the cardiac pacemaker – but obviously to no avail.

We have seen that organisms are characterised by integrated functioning and mutual dependence among their different organs and subsystems. Each organ is tailored to a certain purpose. While some of the processes that an organ carries out, like the consumption of oxygen or glucose, benefit the respective organ alone, the main objective of every organ is to contribute to the functioning of the organism as a whole. The heart maintains circulation, the liver detoxifies the blood, the kidneys remove waste products from the blood stream, and so forth. It is this reciprocal dependency and interrelatedness among its various components that constitutes a living body. Consequently, no function or capacity of one organ alone, however important, can be equated with the organism in its entirety. The brain, or even the whole head, are no exceptions.

Furthermore, even if an organism could indeed be pared down to one of its components, it is not as obvious as Olson and van Inwagen claim that this component must be the head. Why not, for example, regard the heart as the core of the organism? When compared to the brain, one might think that from a physiological viewpoint the heart has in fact the better claim. Since it houses the sinoatrial node, it does not only neurologically direct circulation – and thereby the body's most important integrating function – but it also carries out this task itself. Although the heart is therefore self-sustaining in a way that the brain is not, no one would want to hold that an isolated heart is an organism shrunk to its core. So much for the physiological reasons.

The second reason why we should reject Olson's and van Inwagen's reply has to do with the character of identity relations in general. To reiterate, animalism has the implausible consequence that an individual would cease to exist when her body is des-

228 Olson 1999, 134.
229 Inwagen 1995, 177 f.

troyed even if her conscious head was successfully maintained by the circulation of another body. Olson and van Inwagen try to avoid this unwanted implication by asserting that the severed head remains an organism, only diminished to its core. We can now use our findings from previous sections to defuse this reply in another way.

Identity is a transitive relation. If *a* is identical with *b*, and *b* is identical with *c*, then *a* is identical with *c*.[230] Consequently, if the original organism indeed continued as the severed head, as Olson and van Inwagen submit, the rest of the body could not also be this organism; for otherwise the head and the decapitated body would have to be identical with each other, which is not the case. One can therefore not hold that a head is a pared-down organism without simultaneously being committed to the claim that a headless body is *not* an organism, even if it is being maintained on external life support. We have spent a large portion of chapter 3 establishing that brain-dead bodies *are* functioning organisms. If this conclusion was correct, then the laws of logic dictate that it is impossible for the original organism to continue as the severed head.

Moreover, we have established that consciousness would manifest in a head even if no structures except for the large blood vessels of the neck were connected to the receiving body's circulation. This finding enables us to show the absurdity of the animalist's position even more clearly: since the severed nerves would not need to be reattached, one could simply graft the head onto the receiving body *without* first removing the body's original head – just as Vladimir Demikhov did in his cruel experiments on dogs.[231] In this case, it would be even more obvious that the head that is transplanted does not acquire new body parts to complement its supposed core when it is grafted onto the pristine body on the operating table. For this latter body, breathing and conscious, would already possess its own core since it is not decapitated. Animalists would now be incapable of accounting for the additional head. Given that *two* alleged cores of organisms are present, they cannot describe this configuration as a case of fusion.[232]

If detached heads are so unsuited to count as a pared-down organisms, why do Olson, van Inwagen, and other animalists make this claim? The first reason is certainly the desire to escape challenges of the kind that we have explored in his section. Animalists must avoid situations in which cephalic transplantations force them to hold that

230 For an analysis of whether identity is necessarily transitive, see Perry 1972. The vast majority of authors believe that this is so. See, for example, McMahan 2002, 39; Martin 1998, 12; Parfit 1971, 206.

231 It is likely that the body, maybe with some external assistance, would indeed be able to support the second head. In 1954, doctors began using a technique referred to as *cross circulation* to achieve a total cardiopulmonary bypass in order to be able to perform open heart surgery. In this procedure, the patient who was to be operated upon received blood from another individual with the same blood type – usually a relative – who served, as it were, as an organic oxygenator. To this end, the donor's superficial femoral artery and saphenous vein were canulated and connected to the recipient's circulation via an interposed pump. The inventors of this procedure described it as simulating a 'temporary placenta' (Lillehei 1982, 9). With the advent of heart-lung machines, an easier way of oxygenating blood became available, and cross circulation, which poses serious risks for both participants, was abandoned.

232 Campbell and McMahan (2010) used conjoined twinning to construct a very effective argument against animalism, which exploits the fact that in this condition two centres of consciousness appear to be associated with only one organism. See also Savulescu / Persson 2016.

the newly attached head has become a part of the organism onto which it was grafted. For if the transplanted head was not an organism in its own right, their criterion of personal identity would entail that the individual associated with this head has ceased to exist, while the conscious head would unreservedly be dismissing this interpretation. The best way to circumvent this implausible contention is to proclaim that organisms possess a core to which they can be diminished and to choose as this entity the very anatomical structure on which depends the persistence of mental features, thereby making the two inseparable.

Another reason may be that transplantation medicine has, to a certain extent, made the question of an organism's identity resemble Theseus's paradox. As Thomas Reid remarked, due to the constant exchange of material with the environment in which organisms must engage to uphold their life processes, 'the identity (…) which we ascribe to bodies, whether natural or artificial, is not perfect identity; it is rather something which, for the conveniency of speech, we call identity'.[233] Advocates of biological views of personal identity disagree, of course. That they need to allow for a certain exchange rate of organic components to handle the naturally occurring process of cellular turnover is indeed not normally considered problematic.[234] As McMahan observes,

> our understanding of the continued existence of a physical object normally tolerates a replacement of the object's constituent elements or parts, provided that no one of the parts is essential to the existence of the whole and that the turnover is sufficiently gradual that each new tissue would coexist for a significant period of time with substantially greater amounts of older matter.[235]

But the speed rate cannot be indefinitely high, and the amount of matter exchanged cannot be indefinitely great. Endowing one of the organs, preferably an indivisible one, with special significance avoids this difficulty and immunises biological accounts against all sorts of transplantation scenarios in which large parts of two or more bodies are exchanged in short intervals, which would make it hard, if not impossible, to trace a particular organism through time. Protecting their account against this type of attacks, rather than a special physiological status of amputated heads, may be the true motivation behind stipulating that there is one organ that forms the organism's core. Since organismic life is an irreducibly holistic phenomenon, animalists cannot convincingly retreat to this position, however. A headless (or brain-dead) body on life support is an organism; and a bodiless head (or brain) is not an organism shrunk to its core, but an organ.

I do not see any way for defenders of biological views to remove or sidestep these enormous difficulties. An even harder case for them to attack, however, would be one in which consciousness is preserved while it seems that *no* organism is present. Let us therefore consider how far away from such a scenario we are today. Although we

233 Reid 1878, III.III, § 2.
234 See also note 396.
235 McMahan 2002, 71.

already possess some empirical data regarding this matter, I do not yet consider what follows a sufficient basis for constructing a proper philosophical argument. But it may become one in the future.

4.4.3.2 Inorganic Body

Must a head be supported by an *organic* entity for consciousness to manifest? Instead of transplanting the head onto another human body, one may also consider whether all the supplying functions could instead be carried out by some sophisticated machinery. As early as in 1812, French doctor César Le Gallois speculated about how one could sustain a severed head by artificial means. Le Gallois supposed that if one could artificially perfuse a body part, replacing the function of the heart by a blood injection, even the brain in a detached head would be able to function normally.

> Mais si l'on pouvoit suppléer au cœur par une sorte d'injection, et si en même temps on avoit, pour fournir à l'injection d'une manière continue, une provision de sang artériel, soit naturel, soit formé artificiellement, en supposant qu'une telle formation soit possible, on parviendroit sans peine à entretenir la vie indéfiniment dans quelque tronçon que ce soit; et par conséquent, après la décapitation, on l'entretiendroit dans la tête elle-même avec toutes les fonctions qui sont propres au cerveau.[236]

It took more than a century until apparatuses were finally devised that could not only circulate blood but also oxygenate it, thus replacing both heart and lung function. After dedicated research efforts had also led to a better understanding of the various characteristics of blood, like clotting mechanisms and the compatibility of different types, the first successful cardiopulmonary bypass operation eventually took place in 1953.

Heart-lung machines have since become very sophisticated devices. One of the most popular techniques is extracorporeal membrane oxygenation (ECMO). Via a cannula, usually inserted into the right common femoral vein, a pump drains blood from the body and transports it to a membrane, where the red blood cells pick up oxygen while carbon dioxide is being removed. The blood is then rewarmed to body temperature and returned to the right femoral artery.[237] ECMO is so effective that it enables doctors to bridge the time until a suitable donor organ can be found even when the explantation of the diseased organ has already taken place. In 2016 the case of a Canadian woman hit the headlines whose lungs had to be removed following an infection before a transplant became available. After she had undergone bilateral pneumonec-

236 Le Gallois 1812, 134 f. Translation (by L. J. M.): 'But if one could replace a heart by a kind of injection and if, to deliver the injection in a constant manner, one also had a reservoir of arterial blood, be it natural, be it artificial, on the assumption that such a synthesis would be possible, one would achieve without any problem an indefinite continuation of life in this body part; from which it follows that, after decapitation, one could even maintain the head itself with all the functions that are typical of the brain.'

237 In a sense, ECMO is the inorganic equivalent of cross circulation. See note 231.

tomy, she was being kept alive with an empty chest cavity for six days until she finally received a donor organ and made an astonishing recovery.[238]

Today, inorganic substitutes for vital functions need not even be extracorporeal. Artificial hearts can be implanted into the patients' bodies, allowing them to be discharged from hospital. Such mobile devices are usually employed as short-term bridges to cardiac transplantation, but in some cases patients have survived with artificial hearts for more than thousand days before the latter were finally exchanged for allografts.[239]

Cardiopulmonary activity, the intake and distribution of oxygen, is certainly the bodily function that is most intimately connected to life. But what about the replaceability of the tasks that other organs execute? Keeping a head functional would, at least if consciousness is to be maintained for more than a few minutes, also require that important nutrients be added to the blood stream and that waste products that the brain releases be removed from it. The former task is predominantly performed by the gastrointestinal tract, the latter by the kidneys and the liver. Are there any artificial substitutes available for these organs?

The kidneys filter the blood to remove toxins and take part in maintaining homeostasis. Renal failure can be compensated by haemodialysis. Although dialysis puts strains on the body and carries some risks, it is so effective that patients could even live without both kidneys for extended periods, provided only that they go for dialysis regularly.

The number of functions that the liver performs is enormous. They include the removal of toxic substances, energy storage in the form of glycogen, the manufacturing of blood proteins, and the secretion of gall. Medicine has not yet succeeded in replicating all of these functions artificially. To date, the only effective treatment for acute liver failure is the transplantation of a donor organ. However, extracorporeal liver support systems, like albumin dialysis, are being developed and have shown first promising results.[240] It is likely that in a few decades doctors will be able to replace at least the most important functions of the liver for a certain amount of time.

The provision of nutrients to the head, and thus the substitution of the digestive tract, is more straightforward. Total parenteral nutrition is a technique of intravenously feeding a patient who, due to intestinal failure, cannot take in food via the enteral route. As the nutrients are injected directly into the blood stream, the ordinary process of digestion can be completely bypassed. The patient receives substances like glucose, amino acids, salts, and vitamins in the appropriate concentrations to meet all nutritional needs. The survival rate after 1.5 years of continued total parenteral nutri-

238 Cypel et al. 2017. Overall, the survival rate in adults who require veno-arterial ECMO for more than two weeks is 37.5% (Posluszny et al. 2016, 578). When only pulmonary function needs to be replaced, doctors may employ veno-venous ECMO instead, which is associated with lower risks than its veno-arterial counterpart (Lorusso et al. 2017, 1390) and results in an even higher survival rate of 48.6% (Posluszny et al. 2016, 578).

239 Torregrossa et al. 2014, 626.

240 Maiwall et al. 2014, S511.

tion is 58%.[241] The pancreas, spleen, gallbladder, thyroid, and the reproductive organs can be removed without mechanical replacements when adequate medication is provided.

All the devices and procedures that we have listed are only imperfect substitutes for the respective organ, their application often carries enormous risks, has problematical side effects, and the duration of their application is limited. Most importantly, they are usually not employed in a combined way, which mechanically maintaining a detached head would require. Science is very far from being able to substitute a whole body, and this goal is also not going to be achieved in the near future. I have, therefore, not been trying to sketch an actual medical possibility, that is, an operation that could be carried out with the means available today, as I did in the foregoing section.

Still, we are living in the first century in which we can remain alive without our hearts, lungs, and most other organs for several days. What had been unthinkable for millennia has come true: it is irrelevant whether a bodily function is performed by organic or inorganic means. If this does not only apply to every organ in isolation, but also when the mechanical substitutes are combined, then supporting a conscious head in the absence of the rest of its organism becomes possible. This would make biological accounts of personal identity extremely hard to defend, especially given that animalists cannot just adopt a functionalist position and argue that it is irrelevant whether an underlying physical medium is organic or inorganic as long as the respective functions are performed.[242]

Functionalists would, of course, welcome the prospect that non-biological systems may one day be equally effective at supporting a conscious head as systems comprised of biological organs. Even the generation consciousness itself, they may submit, could in principle supervene on inorganic components, so that the 'cyborg' – the human head maintained by sophisticated machinery – could equally well be turned into an entirely inorganic device without thereby losing its mental features.

A proper treatment of functionalism would take us too far afield, and objections to this doctrine were raised elsewhere.[243] The important point is that biological accounts of personal identity, against which the arguments in this chapter are directed, cannot retreat to this position. Eric Olson, like most other animalists, maintains that

> no biological organism could come to be partly or wholly inorganic. If you cut off an animal's limb and replace it with an inorganic prosthesis, the animal only gets smaller and has something inorganic attached to it. It doesn't acquire prosthetic parts. If you were to replace all an organism's parts with inorganic prostheses, it would no longer be there at all.[244]

241 Oterdoom et al. 2014, 1105.
242 Putnam 1967.
243 See, for example, Levin 2018; Chalmers 1996; Block 1978.
244 Olson 2007, 25.

This differs markedly from the functionalist standpoint. Animalists may have several reasons for holding such a view. An intrinsic motivation could be that organisms, due to their cellular organisation, are autopoietic: they are able to produce the structures of which they are composed and thus to maintain themselves. Inorganic devices, in contrast, are allopoietic. They execute a certain function, but they do not fully participate in the organism's metabolism, for example, by utilising as energy source the glucose that other organs provide or by contributing to the body's immune responses.

An extrinsic motivation for holding that mechanical artifacts can never truly become parts of organisms may be that animalism would undermine itself as an account of personal identity if its advocates permitted that human organisms could gradually be turned into machines and thereby stop being animals without going out of existence. Whether these and other possible reasons that animalists may have for excluding inorganic components from partaking in organismic life are sound need not concern us.[245] What is important is that proponents of biological accounts, according to which we are *biological organisms* and cannot cease to be such entities without thereby ceasing to exist altogether, cannot make use of functionalistic descriptions of inorganic prostheses to block objections to their view.[246]

4.5 Conclusion

The previous chapter ended with a pressing problem: due to the growing sophistication of intensive-care medicine, brain death is no longer an appropriate criterion of the ceasing to exist of an organism. Since all criteria of death that are based on human *physiology* are vulnerable to technological progress, the most promising alternative was looking for a solution at the underlying *metaphysical* level. Thus, instead of modifying the *criterion* of death by trying to find another physiological discontinuity that is seemingly better suited to indicate the organism's death than the destruction of the brain, we chose to re-examine the *definition* of death. The question of this chapter was therefore one of personal identity. We were asking whether we are indeed essentially organisms and, hence, require a biological definition to determine our death, or whether the essence of our being consists in something else.

We began by briefly sketching the three main accounts of personal identity, only to exclude immaterial substance views from our investigation because they evade scientific scrutiny. To decide between the two remaining accounts – the biological and the psychological – we investigated whether an individual's consciousness could manifest in the absence of his or her organism. Since advocates of biological views hold that

245 See also note 227.

246 As noted in section 4.2.2, animalists permit the existence of non-human persons – for example, of extraterrestrials or of God. These creatures could in principle have a metaphysical nature entirely different from ours. They might be inorganic or even wholly immaterial. However, the question of animalism 'is not about the basic metaphysical nature of people as such, but only of ourselves' (Olson 2007, 9). And *human* persons, animalists and other proponents of biological accounts agree, are necessarily *organic* entities.

we cannot cease to be organisms while still continue to exist, preserved consciousness detached from the original body would be compelling evidence against biological accounts and in favour of psychological views.

Rather than asking whether we can *conceive* of ourselves as disembodied conscious entities, which is the ordinary philosophical approach, we analysed actual clinical data from cases in which organismic structures were modified in ways that severely impair the flow of information between the brain and other parts of the body, and evaluated the effects of these modifications on the generation of consciousness. We began by considering pathological conditions and medical procedures that compromise the exchange of electrical signals via all neural pathways that travel through the neck – most notably the spinal cord and the vagus nerve. It emerged that neither the severe reduction in somato- and viscerosensory information that reach the brain under these circumstances nor the brain's inability to send motor commands to any structures outside of the head terminate the generation of consciousness.

We proceeded to analyse the organism's second main channel of communication, the endocrine system, and found that mood and memory are indeed susceptible to hormonal influences. Moreover, several hormones have neuroprotective effects, so that their long-term absence would likely be detrimental to brain function. There is no evidence of a direct implication of hormones in the generation of consciousness, however.

That the brain would still bring forth consciousness despite these neural and chemical signalling pathways being disrupted testifies to a remarkable independence of mental capacities from extracephalic structures – for at this point the head would only be *supplied* by the organism, while otherwise there would be a total standstill of communication. Some may already regard this as sufficient evidence that a locus of consciousness can be detached from the original organism; others may object that strict spatial or temporal separation would be indispensible to prove the metaphysical independence of the structures that support consciousness from the rest of the organism.

In a second step, we therefore investigated whether a conscious head could be supported by another human organism. We identified three main problems with transplanting a head: achieving anastomosis before ischaemia of the cerebral tissues sets in, reconnecting the severed nerves, and controlling the immune rejection. Of these procedures, the first one is possible even with the means available today, the second one is medically impossible in the foreseeable future, but not required for consciousness to manifest, and the third one is probably already medically possible, but unnecessary in genetically very similar individuals like monozygotic twins, which is all that the argument requires.

We considered an often-raised objection from the animalist camp according to which a detached head is in fact a pared-down organism and replied that such a contention is not only mereologically untenable given that organisms are systems of mu-

tually dependent components that work in concert to compose a unified whole, but also incompatible with the transitivity of identity if our findings in chapter 3 were correct.

Finally, to make the case against biological accounts even more compelling, we examined whether, instead of connecting the head to another organism's circulation, one could also employ sophisticated machinery to sustain it. While I regarded the currently available empirical evidence as insufficient to deliver the premises of a proper philosophical argument to this effect, this may become a convincing line of reasoning in the future. From today's perspective, it at least appears as if no important physiological reason precludes such an undertaking, provided that the required technology can be devised.[247]

In this section it was not our aim to show – nor did we succeed in doing so – that the detached head would have 'the same kinds of subjective experiences or states of phenomenal consciousness as an embodied brain', which is a common requirement of thought experiments involving brains in vats.[248] In the metaphysical variant of these hypothetical situations, the envatted brain has subjective experiences that are *similar* to those of an embodied brain; in its stricter epistemic form, the vat is even an environment that is qualitatively *indistinguishable* from proper embodiment from the first-personal perspective. The standards of our argument are much lower. It is immaterial whether, disconnected from its original organism, the brain in the detached head would give rise to exactly the same kinds of subjective experiences as before, when it was still fully embodied. Given the now limited means of sensory input and motor output, the actual content of the stream of consciousness would certainly deviate from that of a brain that is still part of its original organism – just as patients undergo changes in their mental experience following spinal-cord injuries that leave them quadriplegic and deprive them of a comparable amount of sensory and motor functions. Like these patients, however, the detached head would be consciousness, which is all that the argument requires.

It is also important to emphasise that we were not making the claim that conscious human subjects could *come into existence* without developing in conjunction with organisms. One must differentiate between the withdrawal of the organismic environment during embryonic development and early infancy, and its removal later in life. That, for example, the lack of tactile stimulation alone has dire consequences in newborns is well known.[249] The same may be true of the deprivation of motor output. Being able to act in the physical world may be necessary for acquiring the capacity for consciousness during the time of early brain development.[250] A brain that was never

247 But see note 333.

248 Thompson / Cosmelli 2011, 164; see also Damasio 1994, 229.

249 Ardiel / Rankin 2010.

250 See also Strawson's (2010, 251–289) hypothetical case of the *Weather Watchers*, with the help of which he examines whether beings that have mental properties are necessarily behavioural beings. Strawson reaches the conclusion that this is not so.

presented with the environmental influences that an organic body provides, and forced to react to the challenges that they pose through moving the latter in various ways, may be greatly impaired in its development.

This is not an essay in ontogenesis, however. The question of this chapter was not under which conditions a centre of consciousness can *develop*, but whether, once an adequate level of mental activity has been reached, a certain percentage of the incoming and outgoing signals are dispensible. The pieces of empirical evidence that I presented show that this is the case. While we were not able to determine where exactly this threshold is located, hopefully a convincing case was made that it must lie distinctly below the amount of sensory input and motor output that necessitates the presence of a whole organism. Drawing the line at the level of the neck is perfectly sufficient to ensure this.[251] It is overwhelmingly plausible that when an individual's consciousness persists, this individual has not ceased to exist. And if this holds true even in the absence of the body, we must be psychological subjects, not organisms.[252]

251 An additional reason for drawing the line where we did is this: the neck region is a natural weak spot where pathological separations like spinal cord transections are prone to occur. Consequently, the availability of clinical data is particularly good in this case.

252 In this chapter, I was not meaning to suggest that consciousness would be the only important aspect of human existence and that the feeling of being embodied and the pleasure and pain that one derives from possessing an (organic) body are negligible. Becoming wheelchair-bound and losing all sensations that originate from below the neck can change an individual's life in a profound way. But quadriplegic people, or those suffering from any of the other conditions that we have analysed, do not differ from bodily unimpaired members of our society in any *relevant* way – despite being afflicted with diseases that have partly isolated their brains from the rest of their bodies.

The Brain — is wider than the Sky —
For — put them side by side —
The one the other will contain
With ease — and You — beside —

(Emily Dickinson)

5 Memory-Based or Consciousness-Based Account?

5.1 Introduction

A final piece is left of the puzzle to our ultimate goal of determining under which circumstances we die. We need to ascertain what exactly psychological subjects are. So far, we have mainly determined what we are *not*. Rather than only criticising rival views, we should also offer an account ourselves. In the previous chapter, we have only obtained a very rough sketch of our persistence conditions, mainly in opposition to biological views of personal identity. We will now draw finer distinctions, this time not between the brain and the body, but between different areas and functions of the brain itself.

If the functional status of the brain was a binary affair, that is, if the brain was an organ that either functioned in its entirety or else was completely dysfunctional, the question as to when we cease to exist could be answered on the basis of what we have already established: our existence would end with the irreversible destruction of the brain. Matters are not so simple, however. The brain has different parts, which do not necessarily lose their function simultaneously. Cases in point are the vegetative state, in which the cerebrum is severely damaged while the brainstem remains intact, and brainstem strokes, which affect the lower brain but spare the cerebral hemispheres. Conditions like these are particularly challenging to any attempt at defining death psychologically, yet most in need of scientific clarification.

In this chapter, we shall therefore try to determine what the persistence conditions of psychological subjects are and how they relate to the different pathological states that the brain can assume. The relevance of particular brain functions to the diachronic persistence of psychological subjects shall be assessed by considering cases in which these functions are impaired. In the end, we will hopefully be in a position to specify what sort of damage we can survive, and the destruction of which brain areas terminates our existence.

In the first part of the chapter, we shall evaluate the classical memory-based accounts of personal identity of the type that Locke, Parfit, or Shoemaker defended. We will do this by applying these views to medical conditions in which long-term memory is compromised to see whether they can account for such changes. That memory-based views of personal identity may conflict with conditions like amnesia and dementia is, of course, not a novel insight.[253] But this relation has not yet been explored vis-à-vis

253 See McMahan 2002, 43 f.; McMahan 1995, 110 f.; Gervais 1986, 117 f.

the different classes of memory and the different types of mental states that the respective disease affects or spares.

We shall then move on to the opposite permutation, that is, to a condition in which long-term memories are preserved while other mental capacities are irreversibly absent, to probe what memory-based accounts may have to say about this configuration. Patients who have sustained an isolated lesion in the ascending reticular activating system (ARAS) of their brainstem cannot awake and remain forever comatose, yet they retain the neural correlates of their mental states. Are these subjects still in existence? Analysing amnesia, Alzheimer's disease, and ARAS defects will enable us to establish whether it is plausible to assume that our existence is based on a continuity of specific mental states, as memory-based accounts of personal identity maintain.

The second part of the chapter will explore whether the essence of our being could consist in the persistence of the capacity for being conscious *per se* rather than in the retention of long-term memories. We shall be considering how consciousness-based views could solve the problem that the stream of consciousness is periodically interrupted and explore the relation that consciousness bears to our brains. Finally, following an analysis of the results of certain neurosurgical operations, we will hopefully be in a position to specify what circumstances must obtain for us to cease to exist.

5.2 Memory-Based Accounts

Ever since the publication of the second edition of Locke's *Essay Concerning Human Understanding* in 1694, views that focus on the diachronic preservation of memories have been dominating the scene. The great majority of those who advocate psychological accounts of personal identity still follow Locke in believing that we persist over time by virtue of being connected to our former selves through our long-term memories.[254] We have already seen that Locke stipulated that

> whatever has the consciousness of present and past Actions, is the same Person to whom they both belong. Had I the same consciousness, that I saw the Ark and *Noah*'s Flood, as that I saw an overflowing of the *Thames* last Winter, or as I write now, I could no more doubt that I, that write this now, that saw the *Thames* overflow'd last Winter, and that view'd the Flood at the general Deluge, was the same *self*.[255]

According to this criterion, individual *y* at t_2 is identical with individual *x* at t_1 if and only if *y* can remember having had *x*'s experiences. Hence, direct memory connections are both necessary and sufficient for a person to continue to exist. This condition, however, comes with a difficulty: the criterion of diachronic personal identity must

254 Some of the most prominent modern works include Shoemaker 2008; Shoemaker 1999; Parfit 1987; Shoemaker / Swinburne 1984; Lewis 1976; Perry 1972; Parfit 1971; Shoemaker 1970. There is also a constitutionalist variant of Lockeanism, which I will not discuss (cf. e.g. Baker 2007; Baker 2000; Johnston 1992).

255 Locke 2008, II.XXVII, § 16. By *consciousness* Locke meant *memory*, as Thomas Reid (1878, III.III, § 2) noted. See also Noonan 2003, 43 f.

have the same logical form as relations of identity in general.[256] Identity is the relation that a thing bears only to itself. It is transitive (if *a*=*b* and *b*=*c*, then *a*=*c*) and it is all-or-nothing, which means that it does not admit of degrees.[257]

Psychological connectedness satisfies neither of these conditions. First, it is intransitive: if an individual *x* at t_1 is psychologically connected to an individual *y* at t_2, and *y* is psychologically connected to an individual *z* at t_3, it does not follow that *x* is psychologically connected to *z*. This cannot be right, however, since when their respective psychological connections link both *x* and *z* to the same biographical event at t_2, they are supposed to be stages of the same person.[258] Secondly, psychological connectedness does admit of degrees: individual *x* at t_1 can be connected to individual *y* at t_2 by just a single psychological connection or by millions of these. Consequently, the intransitive, non-binary concept of psychological connectedness cannot satisfy the requirements of identity relations.

It was therefore suggested instead to rely on a *continuity* of memories. According to this broader criterion, individual *x* at t_1 is identical with individual *y* at t_2 if and only if there is an overlapping chain of memories between the two points in time.[259] This yields a transitive notion as all person stages that are part of the chain of psychological connections can now be considered to belong to the same mental life: individual *x* at t_1 is identical with individual *z* at t_3 even if there is not a single direct psychological connection between *x* and *z*, provided only that *x* at t_1 is connected to *y* at t_2, and that *y* at t_2 is connected to *z* at t_3.

Like psychological connectedness, psychological continuity, too, admits of degrees since the number of psychological connections between t_1 and t_2 may vary. However, as soon as one specifies a certain number of psychological connections that must obtain between the two points in time for the relation to count as strong enough to ground diachronic identity, the concept becomes a binary one. Thus, psychological continuity satisfies both requirements of identity relations.

Modern proponents of Lockean views, so-called *neo-Lockeans*, have modified Locke's original account even further. Memories understood as recollections of autobiographical events constitute merely one of many different kinds of mental states. Other types of states include desires, intentions, perceptions, values, attitudes, beliefs, and personality traits. Nowadays, most Lockeans hold that a person's diachronic existence is not only constituted by a continuity of memories that depict autobiographical events but also by the persistence of all these other types of mental states.[260]

256 McMahan 2002, 39; Parfit 1987, 206.
257 Carruthers 1995, 207; Reid 1878, III.III, § 2.
258 Reid 1878, III.VI, § 15-30.
259 Parfit 1987, 205; Shoemaker / Swinburne 1984, 90.
260 See, for example, Nagel 2013, 276; Shoemaker 2008, 316; Noonan 2003, 10 f.; Shoemaker 1999, 288; Wollheim 1999, 2; Parfit 1987, 205; Nagel 1986, 45. In the English language, the word *memory* has two different meanings that must not be confused. It can refer to the recollection of a *singular* autobiographical event, like the memory of one's first day at school, or to the brain's capacity for storing information in general.

We can now specify what Lockean accounts put forward as the criterion of personal identity across time: an individual *x* at t_1 is psychologically continuous, and hence identical, with an individual *y* at t_2 if and only if there is a sufficient number of psychological connections that need not extend from t_1 to t_2, but that form overlapping chains between these two points in time. A psychological connection is established by individual *x* at t_1 and individual *y* at t_2 both possessing the same mental state *a*. This mental state may be an autobiographical memory or any other type of state.[261]

Two parameters of this definition are in need of further specification: *how many* psychological connections are necessary and sufficient to reach the threshold above which psychological continuity is regarded as strong enough to ground personal identity? And – since neo-Lockeans have expanded Locke's original criterion to include not only autobiographical memories but also other types of mental states – do all forms of psychological connections carry the *same weight* or are some more important to an individual's persistence than others?

Where the threshold is below which the number of remaining psychological connections is no longer sufficient for a person to continue to exist is a moot point among Lockeans. Parfit, for example, placed it at half the number of direct connections that hold over any day in the life of an ordinary person.[262] In view of memory disorders, some of which we shall be considering later, this is a rather stringent requirement. Moreover, the number of psychological connections is not constant throughout an individual's existence. It increases as we grow older, and it decreases again at the end of our lives. An infant has fewer connections than a person in her forties, who has been spending many years learning facts about the world, forming intentions, developing personality traits, and so forth. Specifying a threshold for diachronic persistence is therefore a difficult task. However, one can also regard the theory's vagueness about this point as a virtue rather than a failing, as McMahan points out. It would be very implausible to assume that a sharp cut-off point exists below which a person suddenly ceases to be with the extinction of just one additional mental connection.[263] Nonetheless, we usually require of an account of personal identity that it is able to deliver a verdict regarding whether or not a certain individual exists at a particular point in time – especially if identity is conceived of as an all-or-nothing relation.

Besides stipulating how many psychological connections must underlie psychological continuity for a person to continue to exist, Lockeans also need to specify how much weight is given to each type of connection. Are all psychological connections of equal importance to a person's persistence or does, say, a belief count more than an intention? Parfit maintained that 'more weight should be given to those con-

261 Since there exists the theoretical possibility that a person undergoes fission and is henceforth psychologically continuous with two distinct individuals, which would be incompatible with the transitivity of the identity relation, a no-branching clause is often added to this definition. See, for example, Parfit 1987, 207.

262 Parfit 1987, 206.

263 McMahan 2002, 44. One might also find it suspicious that mental connections are supposed to be countable. See the considerations about vagueness in section 3.5.

nections which are distinctive, or different in different people'. Since, for example, all English-speakers share many undistinctive memories of how to speak English, these connections should be regarded as less important.[264] Shoemaker expressed a similar thought, noting that while different individuals can have the same personality traits, autobiographical memories seem to be much more intimately related to a particular individual.[265] Thus, Lockeans may hold that those psychological characteristics that contribute the most to making someone the very individual who she is are also of greater importance to her persistence than psychological features that many other people also possess.[266]

Having recapitulated the fundamental assumptions of memory-based accounts of personal identity, we shall now confront these views with medical conditions in which memory is impaired. We will begin with retrograde amnesia and then move on to Alzheimer's disease. Our aim will be to establish whether Lockean accounts have to offer convincing descriptions of these conditions in which the psychological relation that forms the basis of their criterion of diachronic persistence is substantially modified.

5.2.1 Consciousness without Memory

5.2.1.1 Retrograde Amnesia

Retrograde amnesia is a pathological impairment of memory that results from damage mainly to the hippocampus, the temporal lobes, and the diencephalon. Its causes are as diverse as traumatic head injury, infection, alcoholism, or brain surgery, and the symptoms range from mild temporary memory loss to total deletion.[267]

A particularly well-documented case of amnesia is that of Clive Wearing. Mr Wearing is an accomplished musicologist and conductor, known for editing the works of Orlande de Lassus. At the height of his career, he contracted herpesviral encephalitis. The virus attacked his central nervous system, which resulted in a very severe case of retrograde and anterograde amnesia, leaving Wearing unable both to retrieve old memories and to form new ones. His recollection does not exceed the capacity of short-term memory, which spans approximately thirty seconds.[268] Consequently, Mr Wearing perceives every minute of his life as if it was the moment of first awakening. If asked what he has been doing a few seconds ago, he always replies: 'It is like death. No

264 Parfit 1987, 206 n. 6; see also McMahan 2002, 74 and 81.
265 Shoemaker / Swinburne 1984, 89.
266 Giving more weight to those psychological connections that are rarer in the overall population makes identity relations partly dependent on external factors – which means that identity can in certain situations become comparative. Many views are incompatible with such a consequence. I shall not elaborate on this point, however, as I will ultimately not endorse memory-based accounts.
267 Squire / Alvarez 1995, 169 f.
268 Silbernagl / Despopoulos 2009, 340.

thoughts of any kind. No dreams. No sight, no sound, no taste, no smell.'[269] When his wife re-enters the room after having been gone for only a few minutes, he often believes that they had never met before.

This behaviour is clear evidence that the number of psychological connections to his past that Mr Wearing still possesses must be extremely low. His wife, who described living with him in a book entitled *Forever Today*,[270] characterises her husband's condition as follows: 'It's like being a wife and a widow simultaneously. I lost Clive, or most of Clive', she says. 'He's in many senses dead.'[271] Mr Wearing's case is a paradigm example, albeit a very dramatic one, of the damage that amnesia can inflict upon an individual's memory. How would Lockean views interpret such a condition? Consider the following argument.

(P_1) A person *x* persists if and only if in the future there will be someone with whom she is psychologically continuous.

(P_2) In retrograde amnesia, most of the mental states that facilitate the psychological connections which form the basis of psychological continuity are irreversibly lost.

(C_1) *X*, who is amnesic, has ceased to be a person.

(P_3) We are essentially persons (substance sortal).

(C_2) *X* has ceased to exist.

P_1 is the Lockean criterion of diachronic existence as described in the previous section. P_3 is the central assumption of psychological views of personal identity, to which memory-based accounts belong. Lockeans must hold on to this premise, for if it was possible for us to cease being persons *and* to continue to exist, then Lockean views would only postulate the existence conditions of *phase* sortals. As we have seen in section 4.2.3, however, *person* is supposed to denote a substance sortal. In other words: while we can become, and cease to be, teenagers or adolescents (phase sortals), we cannot cease to be persons without thereby ceasing to exist altogether. When the retention of her memories is a necessary condition for a person to continue to exist, then their loss is a sufficient condition for this person to go out of existence.

The second conclusion (C_2), however, seems hard to swallow. As the capacity for consciousness and all other mental faculties are retained in amnesia, there seems to be *someone* there, a conscious subject, who clearly has experiences and communicates. If, as some authors claim, this viewpoint does not belong to the original person,[272] it must belong to a different individual who came into existence after the number of psycholo-

269 Ellis 1999.
270 Wearing 2005.
271 Treays 2005.
272 See, for example, Green / Wikler 1980, 125.

gical connections had fallen below the specified threshold. In addition to multiplying entities beyond necessity, this would be an utterly implausible consequence.[273]

Hence, assuming that the argument is valid, the only option open to Lockeans is to attack the empirical premise (P_2). Given the variables that we have introduced in the foregoing section, they may either maintain that amnesia does not affect large enough a *number* of mental states, so that, in consequence, the number of psychological connections from day to day does not fall below the specified threshold; or they may focus on the *types* of mental states involved and contend that, while the number of psychological connections is indeed rather low, this does not matter since those states that are uninfluenced by amnesia are of greater importance than the affected ones.

As detailed in the foregoing section, there is no consensus regarding the required number of psychological connections. Given this vagueness, it may not be possible to pass judgment on cases that lie in the middle of the spectrum. In cases like Clive Wearing's, however, it is fairly obvious that the threshold has been crossed. If a subject's psychological connectedness to the past is so weak that he experiences every minute of his life as the moment of first awakening, any attempt at incorporating such a condition into Lockean views by arguing that the remaining psychological connections are still numerous enough to account for transtemporal psychological continuity would be rather unconvincing. It would also differ in several orders of magnitude from Parfit's original suggestion according to which at least half the number of direct connections that hold over any day in the life of an ordinary person must be retained.[274] Moreover, many higher animals would meet such a low threshold with ease, which would make it a rather imprecise criterion for the persistence of human persons.[275]

Since, therefore, the prospects of escaping the argument by tinkering with the cut-off point appear bleak, advocates of memory-based accounts can only use the uniqueness of particular psychological connections to build a strategy of defense. They could maintain that the deficit in the total number of psychological connections can be offset if certain *types* of connections are preserved that are especially important to the persistence of persons. To evaluate this possible reply, we must establish what types of memory there are and which of them, if any, remain unaffected in amnesia.

Human memory falls into two broad categories: *declarative* (or explicit) and *non-declarative* (or implicit) memory. Declarative memory encompasses all the information that a subject can actively recall. It is 'memory for facts, ideas, and events – for information that can be brought to conscious recollection as a verbal proposition or as a visual image. (…) [I]t is conscious memory for the name of a friend, last summer's vacation, this morning's conversation'.[276] Declarative memory can be further divided into

273 McMahan 2002, 47.
274 Parfit 1987, 206.
275 Personally, I would not regard this last consequence as particularly problematical. In fact, one should welcome the moral implications that it would have for our treatment of animals.
276 Squire / Kandel 1999, 15.

episodic and *semantic* memory. Memories of autobiographical events and the emotions associated with them belong to the former category. If someone relives her wedding day or thinks of a rendezvous that she had a few weeks ago, she is retrieving episodic memories. It is this type of memory that enables an individual to reconstruct her past in a serial form and to create an autobiographical narrative. Semantic memory, on the other hand, refers to the general factual knowledge that an individual accumulates throughout her life. Facts about the world, concepts, and the meaning of words, for example, fall into this category.

Non-declarative memory, too, admits of subcategories. A common classification is that of *procedural memory*, *priming*, and *classical conditioning*.[277] For the present purpose, only the first of these categories is of potential relevance. Procedural memory enables certain cognitive and motor skills that are executed without the need for conscious control or attention. It is acquired slowly through practice and repetition over a longer period of time.[278] Procedural memory does not contain previous experiences as conscious knowledge on which the subject could report, as in the case of its declarative counterpart, but as skills that can be activated when a certain situation re-occurs. Examples are riding a bicycle, dancing, or playing the piano. We can effortlessly ride a bicycle without paying attention to the movements that we execute while focusing on the traffic or even having a conversation. If we consciously had to coordinate every single motion, simultaneously engaging in other tasks would be impossible. Procedural memory ensures that this is not so.

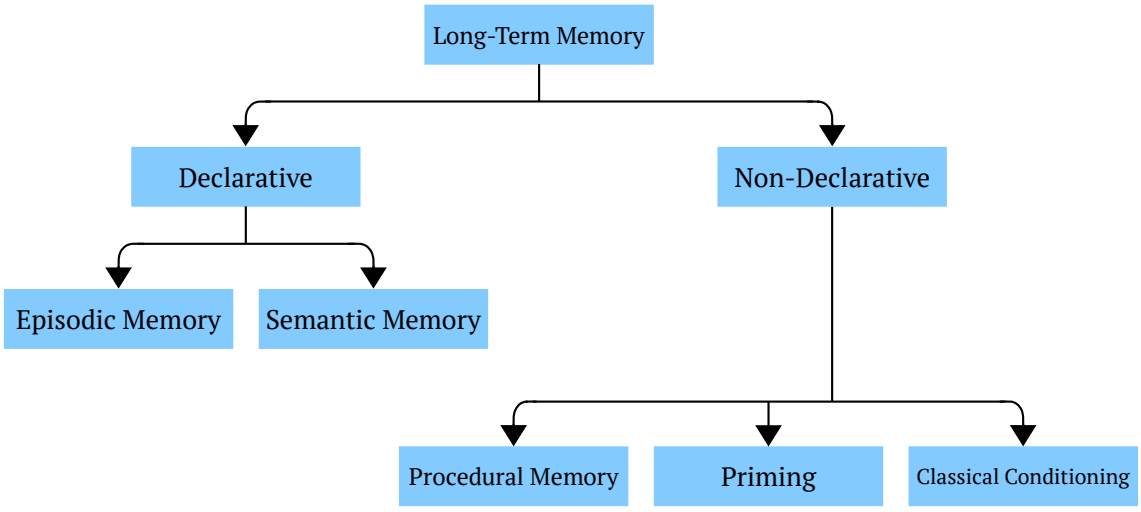


Fig. 3: The different types of memory.

Let us look at semantic memory first. Could Lockeans argue that the retention of semantic memories can compensate for the loss of parts of their episodic counterpart if

277 *Priming* denotes changes in belief and perception that are caused by previous experiences. *Classical Conditioning* refers to the programming of an involuntary, automatic response to a previously neutral stimulus.

278 Fuchs 2017a, 129.

they are sufficiently unique? Semantic memory, we have seen, consists of factual knowledge. Much of this knowledge is shared by a great number of individuals. Many people will, for example, have the conviction that London is the capital of the United Kingdom. Unlike a certain episodic memory, which stems from an autobiographical experience that only very few people could have had, a semantic memory can therefore hardly be unique. But Lockeans may argue that the entire *collection* of these memories can be quite distinctive. Since individual biographies lead to the acquisition of very different pieces of knowledge, their respective combination must differ significantly between subjects.[279] A professor of physics, for instance, possesses vast factual knowledge in an area where most other people have little. If she also happens to be, for example, an avid philatelist, the combination of her semantic memories will be rather unique. While certain other groups of people, too, will know what the theory of critical opalescence entails and that the famous Mauritius 'Post Office' stamps were issued in 1847, only very few individuals will be aware of the two facts in conjunction. The same may be said of a person's vocabulary, which is another important element of semantic memory. While being in command of a particular word is in itself very common, the overall composition of a person's vocabulary, with its strengths in some areas and its weaknesses in others, can be quite distinctive. Clive Wearing may, for instance, still command a large number of specialist musical terms unknown to the average individual.

There are, consequently, two dimensions of uniqueness that one must distinguish: the uniqueness of a *single* psychological connection and the uniqueness of the *set* of psychological connections that a person possesses. As we have seen, these notions can be in conflict with each other. Unexceptional psychological connections can form an extremely rare collection when combined; conversely, a single connection can be highly unique while being part of a rather conventional set.

While most forms of retrograde amnesia target both episodic and semantic memory, often approximately to the same degree,[280] procedural memories – the kind of mental states that guide one's movements when tying one's shoelaces or using knife and fork – are completely spared.[281] These tasks are not at all distinctive, but some people possess motor skills that are relatively unique. Clive Wearing was a gifted pianist, and he is still able to execute complicated sequences of movements on his piano at a speed unattainable to untrained individuals (although he is, of course, ignorant as to why he possesses this ability since only episodic memory would enable the recollection of his musical education). Pool players handle their cue stick so dexterously as if it was a proper part of their body. Excavator drivers control multidirectional movements of their machine simultaneously via two joysticks.

279 For a discussion of the question of whether knowledge can be classified as a mental state, see Nagel 2013.

280 Squire / Zola 1998, 210.

281 Machado et al. 2009, 334; see also Squire / Kandel 1999, 14.

The execution of these highly complex tasks presupposes sets of sophisticated procedural memories that smoothly guide the minute manipulations of the body. Consequently, there can be a high degree of uniqueness even to some procedural memories. In a certain sense, these memories also reflect the individuals' biographies: one could not have acquired them without repeated training over longer periods of time although, unaccompanied by the corresponding episodic memories, these skills would not be embedded in any biographical narrative.[282] For Clive Wearing, one may argue, his procedural memories 'enabl[e] him to transcend the present moment which he can no longer escape by reminiscence, thus guaranteeing a continuity of the self and a kind of identity that is still implicitly felt though not reflectively known'.[283]

In conclusion, it seems that episodic memory, which amnesia affects particularly severely, and which is completely extinguished in Clive Wearing's case, is indeed the most unique type of memory. There can hardly be anything more personal than an individual's recollection of events in her life and the emotions associated with them, seen from a viewpoint that nobody else could possibly occupy. Other types of mental states that belong to the episodic category, like wishes, intentions, or attitudes, are of a highly personal nature, too. Since episodic memories are accessible to conscious recollection, subjects can actively engage with their past, draw on experiences that they have had, and combine the manifold jigsaw pieces into an autobiographical narrative. Only these links between the present and the past enable an individual to lead a structured life. When accounts of personal identity assign different weight to different types of psychological connections, they should therefore put emphasis on episodic memory.

According to Locke's original account, which was based solely on these episodic memories, amnesic persons would go out of existence. Neo-Lockeans, however, who recognise various types of mental states as constitutive of psychological continuity, may use the distinction between the different categories of memory that we have introduced to argue that semantic and procedural memories, which are often spared in amnesia, can compensate for the absence of their episodic counterparts. Certain characteristics of these remaining psychological connections, especially their respective uniqueness, may therefore enable neo-Lockeans to balance an amnesiac's low total number of psychological connections. For this reason, amnesia may not provide a fully convincing counterexample to memory-based views – at least not on a charitable interpretation of these accounts.

282 I do not mean to imply that the ability to play the piano or to operate an excavator resides in the brain *only*. As Wheeler (2018, 238 f.) and Haugeland (1998, 223–227) point out, the connection between the instructions that the brain sends and the bodily factors involved in executing them is so intimate that motor abilities must always be conceived of as holistic properties of the body. 'Der Begriff des Vermögens bezeichnet (…) eine Form integraler Potenzialität, die dem Lebewesen als solchem zukommt und nicht in Einzelprozesse zerlegt werden kann', observes Thomas Fuchs (2017a, 127). Translation (by L. J. M.): 'An ability is (…) a form of integral potentiality that is inherent in a being as such and cannot be split into individual processes.'

283 Fuchs 2017b, 311.

5.2.1.2 Alzheimer's Disease

We have ended the foregoing section concluding that the loss of episodic memory would not entirely disrupt the psychological continuity of Lockean persons provided that the remaining types of memory – semantic and procedural – are at least partly unaffected and sufficiently unique. Lockeans can therefore evade the implausible conclusion that amnesic patients cease to exist by reference to these two types of memory. There is, however, another medical condition in which the deletion of memories is usually even more extensive: Alzheimer's disease.

Alzheimer's is an age-related progressive neurodegenerative disorder that is characterised by cognitive impairment and a variety of neuropsychiatric symptoms, which result in severe restrictions of daily living.[284] It is the most common form of dementia, affecting millions of people worldwide. The effects that the disease exerts on the brain are excessive: cell loss, senile plaques, and neurofibrillary tangles occur in the neocortex, the hippocampus, the amygdala, the basal nucleus of Meynert, and, to a lesser extent, also in the thalamus and the dorsal tegmentum.[285]

Like retrograde amnesia, Alzheimer's leads to extensive deficits in episodic memory. Patients cannot recall autobiographical events and often fail to identify even close family members. However, Alzheimer's differs from amnesia in that the disease ultimately targets semantic and procedural memory, too. Let us therefore explore whether memory-based accounts of personal identity can give a convincing description of this condition.

We have seen that semantic memory comprises general concepts and facts about the world that are not necessarily linked to the individual's biography. Examples are knowing the titles of Plato's most important works, what a red traffic light signifies, or to which entity the word *university* refers. In demented patients, a rapid decline in this kind of knowledge occurs as the disease progresses. They begin to experience difficulties with identifying everyday objects, until they can no longer recall which item to use for which purpose and thus become unable to execute most tasks. The patients' visuospatial orientation is often also drastically compromised, so that they are unable to navigate even familiar surroundings.[286] This development may be accompanied by aphasia: in the beginning, patients only exhibit word-finding deficits, but over time their vocabulary becomes increasingly restricted until it is, in the later stages of the disease, 'often limited to half a dozen words'.[287] These symptoms give a clear indication that, in addition to episodic memory, the number of the semantic psychological connections is also drastically reduced.

The other type of memory that saved Lockeans from the force of the argument of the foregoing section because it is not normally affected in amnesia is procedural

284 Lleó / Blesa 2017, 27.
285 Wenk 2003, 7.
286 Alzheimer's Society in the United Kingdom 2012, 1 f.
287 Lleó / Blesa 2017, 31 and 28.

memory. How does Alzheimer's impact on these unconscious recollections of motor skills? Patients who are in the advanced stages of the disease experience great difficulty in using even the most common objects.[288] They struggle to handle cutlery and require assistance to dress or even to chew the food that they are fed. While Clive Wearing can still play the piano, a patient in the later stages of Alzheimer's disease would not be able to do so. This apraxia is a direct consequence of the destruction of those areas in the brain that used to contain the motor information that specified how to carry out the respective task. Without these mental states, the movements cannot be performed any longer.

The destruction of the different types of memory does not proceed concomitantly but sequentially. Episodic memory is targeted first, so that the patients fail to recall a growing number of autobiographical events. Then the disease begins to affect semantic memory, leaving them increasingly ignorant of general facts about the world and also severely limiting the patients' vocabulary. Procedural memories are the last to go. Thus, as Alzheimer's advances, *all* types of long-term memory eventually fall prey to the disease.

The strategy of defense that we suggested Lockeans could launch against retrograde amnesia being employed as a counterexample to their view is therefore unavailable in the case of Alzheimer's disease. It seems implausible to assume that among the few remaining psychological connections there is one so unique that its retention could reasonably be regarded as compensating for the loss of all the other connections, so that it alone can carry the weight of ensuring the person's diachronic persistence.

Alzheimer's disease is ultimately fatal. Before their organisms cease to function and biological death occurs, there is, in the advanced stages of the disease, a period during which the subjects are still conscious in the basic sense of being awake and aware that we have specified in section 4.3. With nearly all of their psychological connections to the past disrupted, however, they are at this point no longer able to consider themselves as themselves, as Lockean accounts usually demand, nor do they recall general facts about the world.[289] The decline in memory can be so severe that some patients even become unable to recognise their own reflection.[290]

Still, these patients experience emotions, have certain perceptions, feelings, phantasies, or hallucinations, and enjoy other metal states that do not depend on the retention of long-term memory. It would therefore be very implausible to suppose that these patients have ceased to exist. This, however, is what memory-based views of personal identity inevitably imply. Quantitatively, that is, as far as the sheer number of connections is concerned, as well as qualitatively, that is, regarding the types of connections in question and their respective uniqueness, there is absolutely no basis for Lockeans to permit diachronic persistence through Alzheimer's disease. All mitigation

288 Lleó / Blesa 2017, 28.
289 Locke 2008, II.XXVII, § 9.
290 Alzheimer's Society in the United Kingdom 2012, 1 f.

strategies that were successful in the case of retrograde amnesia fail when confronted with this condition.[291]

Advocates of memory-based views could, of course, assert that a *new* subject comes into being when the disruption of psychological continuity supposedly puts an end to the existence of the original person. It is true that demented individuals often behave in ways that appear irreconcilable with the demeanour that they had been displaying prior to the onset of the disease. Alzheimer's, like other neurodegenerative diseases, may affect emotional processing and sexual behaviour, sometimes even leading to antisocial action and violence in individuals who had no history of such conduct.[292] The changes can be so profound that even close friends and relatives may find it challenging to recognise the patients as the individuals they used to be.

Concluding from these changes in character that the original subject was supplanted by someone else is confusing the concepts of qualitative and numerical identity, however. When we say of someone that she is 'not the same as she was before the onset of dementia', we do not really mean that we are dealing with a numerically different individual. Rather, with this *façon de parler* we only denote a significant qualitative change. We are not expressing a fact about diachronic personal identity, nor would we be justified in doing so.

As long as thoughts, emotions, and perceptions persist, *someone* must be present – a subjective perspective, a centre of awareness, to which these mental states are attributable.[293] We have no reason to suppose – not only to avoid multiplying entities beyond necessity – that a new individual, another psychological subject, comes into being as Alzheimer's disease progresses. Hence, it must be the original individual who remains present, from which it follows that the retention of long-term memory cannot be the basis of our diachronic existence and that memory-centered accounts of personal identity must be false.[294]

---

291 It is tempting to make a similar argument that targets the other end of life: is it not the case that psychological continuity is too weak in fetuses and early infants to fulfil the Lockean condition of diachronic persistence, while simultaneously there is clearly a psychological subject present? 'Any understanding of the conditions of our ceasing to exist or ceasing to be alive', observes McMahan, 'implies a corresponding understanding of the conditions of our beginning to exist, or beginning to be alive. The beginning to exist and the ceasing to exist of a certain kind of thing should be symmetrical, mirror-images of one another' (McMahan 2002, 436; see also Birnbacher 2017, 19 f.). Hence, if psychological continuity from day to day is still very weak in fetuses and infants, they are in a situation similar to that of amnesic and demented individuals, and their condition should equally well serve as an objection to Lockean views. The problem with this line of reasoning is that, although conceptually correct, it rests on shaky empirical grounds. Recent studies that employed non-verbal test procedures revealed previously overlooked impressive memory performance in this age group (Schneider / Ornstein 2019, 308 f.).

292 Liljegren et al. 2015; see also Frings / Jox 2015, 104.

293 See also McMahan 1995, 110.

294 Proponents of *biological* accounts of personal identity would, of course, have an effective reply to the disorders analysed in this chapter: amnesia and Alzheimer's disease, they would submit, just show that insisting on psychological criteria is futile in general. Amnesiacs and demented individuals persist due to the simple fact that these conditions do not threaten the survival of the *organism*. We have, however, already excluded biological accounts as suitable candidates for what we essentially are for other reasons. In this chapter, we are solely concerned with deciding between different *psychological*

5.2.2 Memory without Consciousness

5.2.2.1 Dysfunctional Ascending Reticular Activating System

In the previous section, we were analysing conditions in which psychological continuity is disrupted while other mental capacities are largely retained, and concluded that we would survive this configuration. We shall now turn to the opposite permutation, which may present an even more compelling reason to reject memory-based accounts of personal identity: a condition in which the neural correlates of psychological continuity remain intact – and the Lockean criterion of personal identity continues to be fulfilled – while the brain's capacity to support any mental activity is irreversibly lost. This configuration occurs in the rare case of an isolated lesion that disables the ascending reticular activating system (ARAS) but spares the rest of the brain.

I shall begin by explaining the role that the ARAS plays in the generation of consciousness before examining how damage to this structure engenders a situation that is incompatible with Lockean accounts as they are currently formulated. We will then consider three possible objections that defenders of memory-based views might raise to this argument: that a person's mental states persist only when the ARAS remains functional, that a dysfunctional ARAS exerts a qualitative influence on mental states, and that the ARAS is a replaceable structure.

In section 4.3, we defined consciousness as the product of a quantitative and a qualitative dimension, namely, of wakefulness and awareness. Wakefulness, the level of consciousness, encompasses many intermediate stages ranging from somnolence to full alertness. The brainstem structure that is responsible for the generation of wakefulness will be the focus of the following sections. We have also seen that that being awake does not necessarily entail being conscious since a stream of consciousness can manifest only when there is also certain content of which the subject is aware. Awareness is predominantly a function of complex interactions between several areas in the cerebrum.[295]

The relation between wakefulness and awareness is a hierarchical one. Although the degree of wakefulness does not predispose the content of the stream of consciousness, being awake is a necessary condition for awareness to manifest. The converse is not the case: wakefulness can occur without awareness, and when it does, the result is the vegetative state. As discussed in section 3.3, patients in this condition open their eyes and exhibit sleep-wake cycles, but their arousal remains empty and is not associated with any experiential content.[296]

The quest for the mechanism that underlies the regulation of wakefulness and sleep began in the last quarter of the 19th century. Initially it was thought that sleep is

views.

295 Pallis / Harley 1996, 10; Plum / Posner 1980, 11.

296 Giacino et al. 2018, 2; Multi-Society Task Force on PVS 1994, 1500 f.

a passive phenomenon that results from a lack of adequate sensory stimulation, which was considered necessary to maintain wakefulness. Sleep was therefore conceived of as a resting state that the brain enters whenever the input from a number of sensory modalities drops below a certain threshold.[297]

In the middle of the 20th century, animal experiments revealed that this cannot be true, however. It was found that when a specific area in the brainstem of a sleeping animal is stimulated with an electrode, diffuse EEG desynchronisation and behavioural arousal ensues; and that a slow synchronised EEG coma occurs when the same area is destroyed. Sensory stimulation does not lead to a reversal of this effect, even when the main sensory pathways that connect the periphery to the cortex are entirely preserved.[298] The conclusion drawn from these observations was that the onset of sleep following damage to the brainstem was not, as previously thought, the result of deafferentation, that is, of the interruption of sensory conduction to the brain;[299] rather, it was now interpreted as the physiological expression of the elimination of a yet undiscovered structure that exerts a waking influence on the cerebrum, of a 'subcortical pacemaker diffusely affecting the cortex but lying outside of it'.[300] In 1949 neurophysiologists Moruzzi and Magoun eventually discovered this structure: the ascending reticular activating system.[301]

Of the two dimensions of consciousness – wakefulness and awareness – the ARAS only regulates the former:

> The ARAS influences cortical activity to produce alertness, or cyclical wakefulness, including sleep. Awareness on the other hand, is principally a product of the thalamus and cerebral cortex, although a functioning ARAS is necessary for it to manifest.[302]

What the animal studies showed is also true of human physiology. Put simply, the ARAS functions as a controller for consciousness. When its activity level is high, neural oscillations in the cortex desynchronise and the subject awakes; when the neurons of the ARAS reduce their firing rate, cortical oscillations synchronise again, and the subject becomes drowsy until she finally falls asleep.

Originating from the reticular formation of the lower midbrain and upper pons, the ARAS connects to the cortex via three different pathways.[303] Even very small structural lesions in these strategically important brainstem areas are sufficient to render the ARAS permanently dysfunctional and to induce catastrophic global changes in the electrophysiological activity of the cerebral hemispheres. Given that being awake is a prerequisite of being aware, affected patients are going to remain forever comatose.[304]

297 Mendelson 1987, 20.
298 Plum / Posner 1980, 12; Hassler 1971, 27 f.; Moruzzi / Magoun 1949, 471.
299 Moruzzi / Magoun 1949, 470.
300 Plum / Posner 1980, 12.
301 Moruzzi / Magoun 1949; see also Dempsey / Morison 1941 and Mauthner 1890.
302 Horne 2009, 12.2.
303 Jang / Kwon 2015, 200 and 202.

This is the configuration in which we are interested in this section. Its relevance as a counter-argument to memory-based accounts of personal identity lies in the fact that tissue damage extensive enough to disable the ARAS is strictly confined to a brain area of which we have reason to assume that it does not house[305] the neural correlates of long-term memories. Consequently, the persistence of the respective mental states, and thus psychological continuity in the Lockean sense, is *structurally independent* of the persistence of the capacity for wakefulness.[306] From this it follows that individuals whose ARAS is destroyed and who are therefore never going to regain consciousness nonetheless continue to fulfil the criteria of diachronic existence that memory-based accounts endorse.

Let me explain this in greater detail. If the anatomical loci of the ARAS and the brain areas that contain the neural correlates of our memories were entirely congruent, a functional dissociation of arousal and the storage of mental states could not occur. Damage to the respective brain regions would always result in the loss or modification of both the capacity for wakefulness *and* the person's beliefs, desires, intentions, and so forth. But this is not the case. We have already dealt with two medical conditions that give an indication of this separation: Alzheimer's disease and amnesia. Alzheimer's has a profound effect on a subject's memory, but at least in its earlier stages the disease is not accompanied by a reduction in arousal.[307] The same is true of retrograde amnesia. Amnesic individuals are unable to recall details from their previous life, yet their wakefulness level remains entirely unaffected and consciousness itself is not diminished. This configuration is possible owing to the fact that memories reside in the cerebrum, whereas large parts of the ARAS – most notably the reticular formation – are located in the phylogenetically older brainstem.[308]

While we are here concerned with the opposite permutation, namely, the persistence of mental states in the absence of the capacity for arousal, it is the same functional segregation that accounts for both scenarios. When the wakefulness component of consciousness is rendered dysfunctional as a result of lesions in non-cerebral brain areas, that is, in the reticular formation or its projections, the neural correlates of memories, and thus of psychological continuity in the Lockean sense, remain unaffected. However, owing to the fact that being awake is a prerequisite of being aware,

304 Koch 2010, 20; Hassler 1971, 29 and 33; Hassler et al. 1969, 309. Even in REM sleep, in which the subject is aware of the contents of her dreams while not being fully conscious, a certain degree of wakefulness, and thus of cortical activation, is required for these dreams to manifest (Fuchs 2017b, 300; Windt / Nielsen / Thompson 2016; Laureys 2005, 556; Evans 2003, 2 f.).

305 This is to be understood in the most neutral sense possible. I shall withhold judgment on which account – identity theory, double-aspect theory, property dualism, functionalism, or any other of the many models that have been proposed – best describes the relation between mental properties and their neural substrate. Especially in the phenomenological tradition, the notion of the brain 'containing' any information is often rejected (Fuchs 2017a, 58–64 and 129 f.). Since Lockeans generally do not dispute this assumption, however, we shall not question it here.

306 We will consider objections to this claim in the subsequent sections.

307 Plum / Posner 1980, 5.

308 Crossman / Neary 2015, 166 f.; Plum / Posner 1980, 13.

the respective mental states cannot become conscious any longer since 'without arousal mediated by the ARAS, awareness is not possible. Even if cortical awareness networks are intact, they remain quiescent without activation by the ARAS'.[309] Beliefs, intentions, and other states continue to exist – but they will forever be beyond reach.

In practice, this situation can only occur when the lesion is not extensive enough to destroy other areas of the brainstem, so that the cerebrum is being kept oxygenated and supplied with glucose.[310] For if perfusion of the cerebral hemispheres stops and anoxia persists for longer than approximately three minutes, irreparable ischaemic damage to the tissues begins to ensue.[311] In the course of this process, the neural correlates of all memories encoded in the delicate synaptic circuitry are extinguished and psychological continuity is disrupted, in which case Lockeans need no longer regard the comatose entity in the hospital bed as numerically identical with the original person.

Lesions confined to the reticular formation or its projections do indeed occur, their most common causes being ischaemia, hemorrhage, direct trauma, and tumours. Edlow et al. describe the case of a 62-year-old woman who was admitted to hospital after having suffered severe traumatic brain injury. Spontaneous movement in the extremities was absent, but brainstem-mediated reflexes remained intact.[312] A connectivity analysis revealed that only the subcortical fibre tracts of the ARAS were completely transected, which strongly implicated the latter as the defining substrate of coma.[313] Other case studies have confirmed the hypothesis that coma can result exclusively from compromised ARAS structures.[314]

If these clinical reports are correct, and if it is true that the neural correlates of the mental states that constitute psychological continuity do not extend to the brainstem area that is the core of the ARAS, then it follows that the fact that wakefulness cannot be regained owing to a dysfunctional ARAS does not entail that these neural correlates themselves are extinct. They are preserved as long as the corresponding brain areas are being oxygenated.

Those who advocate psychological continuity as the criterion of the diachronic existence of persons must consequently conclude that this condition is still fulfilled. For *x* at t_1 (prior to the destruction of the ARAS) and *y* at t_2 (subsequent to the destruction of the ARAS) possess the same number of identical mental states and are therefore sufficiently psychologically connected. The person can *per definitionem* not have ceased to exist when the sole relation that Lockeans endow with significance is not compromised. Hence, the selective destruction of the ARAS must be a case of survival

309 Edlow et al. 2013, 506; see also Jang / Kim 2015, 671 and Parvizi / Damasio 2003, 1531 f.

310 A small number of cases have been reported in which the cerebrum continued to be perfused even though the brainstem was nearly entirely destroyed (Schlake / Roosen 2001, 69 f.; Zwarts / Kornips / Vogels 2001; Kosteljanetz et al. 1988; Deliyannakis / Ioannou / Davaroukas 1975).

311 Silbernagl / Despopoulos 2009, 130; see also section 5.3.2.

312 Edlow et al. 2013, 506.

313 Edlow et al. 2013, 511 and 516.

314 Jang / Kim 2015, 669; Koch 2010, 20; Parvizi / Damasio 2003, 1530.

according to memory-based accounts of personal identity. For a psychological view, this is a highly unattractive result given that consciousness can never again materialise in the affected brain.

5.2.2.2 Quantitative Dependence of Mental States on Cortical Activation

We shall now consider three objections that proponents of memory-based accounts may raise to this argument. An important premise is that the persistence of mental states is quantitatively and qualitatively unaffected by the status of the ARAS, for if it were not, there would be no reason to suppose that psychological continuity still obtains when the ARAS is dysfunctional. By *quantitative* independence, I mean that the entities that persist in the brain after the destruction of the ARAS are indeed still the *kinds* of things that Lockeans recognise as constitutive of personal identity across time. *Qualitative* independence denotes the fact that an ARAS defect does not alter the *content* of these states – like, for example, modifying a memory.

Let us begin by examining the quantitative part of the objection. Is a permanently unconscious neural correlate of a mental state still the correlate of a mental state? Or, put differently: is the ARAS an integral component of a mental state's realising base? And are defenders of memory-based views obliged to accept these unconscious entities as the constitutive elements of psychological continuity?

The essential feature of mental states, it seems, is that they are conscious. Prima facie, it therefore appears absurd to suppose that they could persist without this defining attribute. However, even when we are fully awake only a tiny fraction of our mental states are brought to consciousness, while the great majority of them remain unconscious. And during episodes of dreamless sleep, coma, and general anaesthesia, *none* of the mental states that we possess – or rather their neural correlates – figure in a conscious process.

In the philosophy of mind, authors therefore traditionally distinguish between occurrent and standing mental states.[315] *Occurrent* (or transient) states are those that are being entertained, that is, those that feature in the subject's mental processes at a given moment in time – for example, a sudden sensation of cold or an overwhelming feeling of joy. These states are active.[316] *Standing* (or dispositional) states are those that are not currently part of the stream of consciousness. A standing mental state, stored in an individual's long-term memory, only becomes occurrent when it is integrated into a mental process. One may, for instance, have the standing belief that the moon revolves around the earth, but only when one looks out of the window during a starry night, this belief becomes consciously endorsed and thereby occurrent.[317]

315 Buckwalter / Rose / Turri 2015, 753; Farkas 2008, 40; Braddon-Mitchell / Jackson 2007, 303; Gertler 2007; Wollheim 1999, 1 f.; Shoemaker 1997, 295; Block / Fodor 1972, 168; Goldman 1970, 86 f.; Broad 1945, 135.

316 Bartlett 2018a, 11 f.

317 Standing mental states must not be confused with what is commonly referred to as *subconscious* states. Subconscious states are emotions, wishes, and so forth, of which the subject is *also* not aware,

Standing mental states can therefore be conceived of as dispositions to have certain occurrent states;[318] or, in other words, (some) occurrent mental states are conscious manifestations of standing states.[319] Conversely, occurrent mental states can initiate the formation of standing dispositions when they are laid down in memory: a close encounter with a large spider, for example, may lead to long-lasting arachnophobia in some people.[320]

When a subject is dreamlessly sleeping or comatose, she possesses, *ipso facto*, only standing mental states. Some states that were occurrent before the stream of consciousness ended will now be retained in their dispositional form, while others will simply be lost.[321] None, however, will continue to be occurrent given that wakefulness is globally absent. Since memory-based accounts of personal identity endeavour to formulate the diachronic persistence conditions of persons for intervals longer than a single day, that is, exceeding periods of awareness that are uninterrupted by sleep, it follows that standing mental states, rather than occurrent ones, must form the basis of the relation of psychological continuity that they put forward as the criterion of a person's persistence.

Although usually not expressed in the above terminology, this is indeed the neo-Lockean position. Shoemaker, for example, maintains that mental states 'need not be conscious; most of them will exist in the way my beliefs about Argentina exist when I am giving no thought to Argentina, or in the way my memories of my schooldays exist when I am sound asleep'.[322] To say of a sleeping subject that she retains most of the beliefs, wishes, and personality traits that she has during her waking hours is also in accordance with most people's intuition.[323] As John Searle submits, we understand the notion of an unconscious mental state

> only as a possible content of consciousness, only as the sort of thing that, though not conscious, and perhaps impossible to bring to consciousness for various reasons, nonetheless is the *sort of thing* that could be or could have been conscious.[324]

Thus, unless Lockeans want to stipulate that persons momentarily cease to exist during episodes of dreamless sleep or coma, these unconscious neural correlates of standing mental states are the kinds of entities whose integrity they must accept as suffi-

---

but which are nonetheless active and may therefore still exert an influence on her actions or on her other mental states (Bartlett 2018a, 7 f.; Freud 2016, 7; Farkas 2008, 40 f. and 45–48; Mele 2003, 30 f.). When, for example, a repressed feeling of guilt guides someone's behaviour, the subject may not be aware of its existence as it operates below the level of conscious reflection.

318 Mele 2003, 31; Goldman 1970, 86.

319 Bartlett 2018b, 2.

320 Wollheim 1999, 2; Price 1969, 244. The exact nature of the relation between these two basic modes of mental states has been the subject of a long debate in which we cannot here engage. See, for example, Bartlett 2018a; Farkas 2008, 40–45; Crane 2001, 105–108; Wollheim 1999, 2 f.; Block / Fodor 1972, 168 f.

321 Buckwalter / Rose / Turri 2015, 753.

322 Shoemaker / Swinburne 1984, 96.

323 Mele 2003, 30; Price 1969, 244.

324 Searle 1992, 156.

cient condition of psychological continuity whenever the individual is not awake. And since the function of the ARAS is only to raise the global wakefulness level, it is not a necessary component of these states' physical realisation base; only *occurrent* states require that it be intact. The microstructural configurations in the cerebral tissues that underlie *standing* mental states are carried forward unimpaired through periods of absent wakefulness. Psychological continuity in the Lockean sense must therefore indeed be quantitatively independent of cortical activation and thereby of a concomitantly operating ARAS. Hence, the quantitative objection seems to be unsubstantiated.

Sleep, anaesthesia, and transient coma, however, are temporary phenomena. That Lockeans regard standing mental states as constitutive of psychological continuity, they may retort, is really only with the proviso that the former are going to be integrated into a stream of consciousness at a later point in time – that they will become occurrent in the future. If consciousness is not regained, they may assert, the person has already ceased to exist. Consequently, the ARAS must remain intact. Would this be a convincing reply? Consider the following situation.

> Hypnos and Thanatos are brothers who share a genetic predisposition to cardiovascular disease. One evening they go to bed at midnight. Hypnos immediately falls into a peaceful dreamless sleep. At the same time, Thanatos, while still awake, suffers a heart attack and dies quickly. Hypnos continues to sleep until dawn, by which time his heart fails, too. He does not wake up, and eventually his breathing stops.[325]

Did Hypnos and Thanatos cease to exist at the same time? Both brothers experienced their last conscious moment at midnight – Thanatos because this was when his circulation collapsed and Hypnos because at that time he fell into a dreamless sleep from which he did not awake.[326] Although from Hypnos's perspective it must appear as if he had died at midnight, too, it seems reasonable to assume that the psychological subject Hypnos did not stop existing before dawn. A sleeping person who does not dream has not gone out of existence, irrespective of whether or not consciousness is later regained; for what happens to someone at the end of the night does not retroactively determine his status at any prior point in time. Although Hypnos slipped into unconsciousness at the same time at which Thanatos's biological life ended, the brothers did not cease to exist simultaneously, that is, with their last conscious moment at midnight. Rather, Hypnos's psychological continuity extended beyond midnight since the unique structuring of his brain tissues, and thus the neural correlates of his standing mental states, remained intact until, at dawn, his cerebral blood flow broke down and autolysis began to dissolve his hemispheres, thereby irrevocably disrupting psychological continuity. Although Hypnos did ultimately not regain consciousness, there must have been a person until dawn.

325 Example taken from my BPhil thesis (Oxford, 2016, 43 f.).

326 Should it be objected that even in a dreamless sleep the cerebrum is not totally idle, one could exchange this scenario for pharmacologically or hypothermically induced burst suppression followed by a lethal injection.

Sleep and coma are characterised by very low or even completely absent ARAS activity. If after midnight Hypnos had suffered a small ischaemic brainstem stroke that had caused localised damage to the reticular formation, thereby destroying the core of his ARAS, the ontological status of the neural correlates of his mental states would not have changed in relation to their status during dreamless sleep. Searle observes that

> the possibility of interference by various forms of pathology does not alter the fact that any unconscious intentional state is the sort of thing that is in principle accessible to consciousness. It may be unconscious not only in the sense that it does not *happen* to be conscious then and there, but also in the sense that for one reason or another the agent simply *could not* bring it to consciousness.[327]

Mental states are not only inaccessible when the ARAS is *destroyed*, but whenever it interrupts the stream of consciousness by lowering the level of wakefulness. If one regards the dispositional form that mental states then assume as sufficient for ensuring a person's persistence during dreamless sleep and transient coma, as Lockeans undoubtedly do, one must *ex hypothesi* also arrive at the same conclusion when the ARAS is defective.

To summarise, proponents of memory-based views can account for transient periods of absent consciousness only by postulating that mental states persist independently of cortical activation, that is, in their standing form as mere configurations of oxygenated brain tissue that retains a particular microstructuring. The case of Hypnos and Thanatos shows that this even holds when consciousness is in fact *never* regained. Psychological continuity therefore obtains not only irrespective of whether the neural correlates of mental states are currently incorporated into a conscious process, but also regardless of whether they will *ever again* become so incorporated and become occurrent.

Under ordinary circumstances, this does not present any problem for Lockean accounts since situations of the latter kind are normally either at least potentially reversible (Hypnos, for instance, did not awake, but he could have had), or the defect that keeps the brain from generating consciousness simultaneously also extinguishes the neural correlates of the mental states themselves (as in Thanatos's case). The sole deviation from this principle occurs when the ARAS is damaged while the rest of the brain remains intact.

Without the potential ever to awake, there is clearly no person present. The neural correlates of psychological continuity in a brain that has become incapable of reaching an adequate level of wakefulness no more constitute a living person than an indecipherable blueprint for a house constitutes a building. Consequently, if advocates of memory-based accounts adhere to their criterion of personal identity, they are forced either to hold that the existence of persons end when they fall into a dreamless sleep, are anaesthetised, or enter a transient coma (and to assume that numerically

327 Searle 1992, 160.

different subjects come into being upon awakening), or to classify irreversibly comatose individuals as Lockean persons. Neither option seems tolerable.

5.2.2.3 Qualitative Dependence of Mental States on Cortical Activation

The proponents of memory-based accounts may accept that a defect of the ARAS does not extinguish the neural correlates of standing mental states, but argue instead that it changes their respective content. This is the *qualitative* objection. Since, according to these views, transtemporal psychological connections are established only between mental states that are qualitatively identical between t_1 and t_2, such as the same belief or the same intention, any substantial alteration of these states' content would disrupt psychological continuity, in which case the Lockean condition of diachronic persistence would cease to be fulfilled. The Lockean person would then indeed no longer exist and our argument against memory-based views would be unfounded.

Psychologist Hans Eysenck famously suggested that dissimilarities in cortical arousal are responsible for behavioural differences between introverts and extraverts. He hypothesised that the latter exhibit lower arousal levels, which prompt them to seek a greater amount of external stimulation than introverts.[328] Arousal is a function of the ARAS. Hence, if Eysenck was right, the envisaged situation in which the neural correlates of standing mental states persist unaltered while consciousness is irreversibly absent could not occur – for a comatose individual's mental states would not only be withdrawn from any conscious process but, devoid of the influence of the ARAS, actually be *different*. Does Eysenck's theory therefore refute the argument?

As noted in section 5.2, Locke only considered what one would nowadays refer to as *episodic memory* as the basis of psychological connectedness. Personality traits, with which Eysenck was concerned, do not fall into this category. Since arousal levels cannot interfere with episodic memories in the way that Eysenck took them to impact on personality traits, Locke himself could not employ any dependence that might exist between ARAS activity and temperament against our argument.

As also previously mentioned, however, neo-Lockean views of personal identity endeavour to correct the imbalance towards episodic memory by recognising many different classes of mental states as forming the basis of psychological continuity. Since personality traits are among these recognised subtypes, there would, if Eysenck's theory was correct, indeed be one class of mental states that is constitutive of a person's persistence while not being qualitatively independent of the prevalent degree of arousal.

However, there is no indication that the level of wakefulness modifies any other class of standing mental states: a person's *belief* that London is the capital of the United Kingdon persists unchanged regardless of whether she is drowsy or highly alert, her *memory* of seeing the Tower Bridge remains stable through periods of sleepiness,

328 Eysenck 2006.

and her *wish* to meet Queen does not change during general anaesthesia. Whether these states can be *accessed*, that is, whether they can become conscious or play a subconscious role, does depend on whether wakefulness is being generated by the ARAS. A comatose individual cannot put to use her beliefs or intentions, just as she cannot relive her memories. Qualitatively, however, the mental states remain unaltered. It is this very dichotomy on which our argument against memory-based accounts rests. We can therefore conclude that since personality traits are only one of the many types of mental states that these views recognise, Eysenck's theory – which remains controversial even among psychologists[329] – does not present a successful objection.

5.2.2.4 Reversibility of ARAS Defects

'Irreversible' is a polysemous term. There are different types of irreversibility and it may be critical to which a dysfunctional ARAS belongs. The impossibility of restoring wakefulness in a brain may be logical, nomological, metaphysical, or merely technical.[330] If an ARAS defect fell into the last of these categories, one could argue that Lockean persons would survive this condition since the standing mental states, which remain physically realised, could *in principle* still become conscious – if the appropriate technology were available. There would, at a future time, be an event that would count as the arousal of numerically the same person who is related through psychological continuity to the one who once suffered the destruction of her ARAS. The subject's mental states would in this case only be *de facto* inaccessible, that is, in relation to the technology available today.[331] Since death is *per definitionem* an irreversible state, the person would have to remain in existence as long as there is the possibility for her to regain consciousness at any subsequent point in time.[332] Do we have reason to regard the ARAS as a potentially replaceable structure?

Two considerations, it appears, are relevant in this regard. The first, which one may term the *technological* problem, is the question as to whether we can expect to develop a functional prosthesis. The second, which one could call the *metaphysical* problem, is the question as to whether someone who awakes with the help of this implanted device would still count as numerically the same psychological individual. If

329 See, for example, Gray's (1981) classic critique and Hagemann and Naumann's (2009) more recent study.

330 The different notions of impossibility were defined in section 2.3.

331 I owe the idea of applying this distinction to the question at hand to Jeff McMahan (1995, 105). Besides repairing or replacing the ARAS, a further, highly hypothetical way of accessing the mental states might be considered by some, namely, extracting the latter from the cerebrum and feeding them into an emulation. In *Whole Brain Emulation: A Roadmap*, Sandberg and Bostrom (2008, 83) paint a very optimistic picture of the technological feasibility of such a project. As a thought experiment, Bernard Williams (1970, 162 f.) imagined a similar situation. The assumption that a mental state is an entity that could simply be detached from its neural carrier substance and, like a digital file, be transferred to different media, is extremely speculative. We will therefore not pursue this idea.

332 Under *tertium non datur* conditions, that is. Two alternative options, which we will not be considering, are postulating a third, 'suspended' state between existence and non-existence, and deeming metaphysically sound what Locke (2008, II.XXVII, § 1) explicitly denied, namely, that one and the same thing can have two beginnings.

the answer to this latter question is negative, then the subject can *in principle* not return to consciousness, even if one answers the first question in the affirmative, that is, even if a replacement is *in practice* technologically feasible.

Both problems are extremely hard to resolve, and I am not pretending to have answers to them. The precariousness of the technological problem stems from the fact that long-term scientific progress is largely unpredictable.[333] The metaphysical problem, on the other hand, is vexed because we have not yet understood how nervous tissue brings forth mental phenomena. Whether the realisation of a person's consciousness is conditional on its supervening on the same brainstem substrate for generating wakefulness is something we do not know – just as we do not know this in the case of awareness and the cerebrum.[334] The best we can presently do is, first, carefully to consider the relevant physiological facts and, secondly, to take into consideration the relatively scarce clinical case reports that we have. This is what we are going to do next.

The ARAS is often portrayed as an on/off-switch for consciousness, which gives the impression that it is a relatively primitive structure. As we have already determined, however, wakefulness is not a binary affair but comes in many degrees. The ARAS should therefore rather be conceived of as a dimmer.[335] It is a complex network that originates from multiple brainstem source nuclei, projects to the cortex via thalamic and extrathalamic pathways, and releases various types of neurotransmitters.[336] The ARAS does not simply monodirectionally stimulate the cerebrum, but is itself influenced by areas of the upper brain: the hypothalamus and the basal forebrain reciprocally innervate the reticular formation, thereby providing a feedback mechanism that modulates ARAS activity.[337] Some authors even suspect the existence of several ascending activating systems working in parallel[338] and argue that the role of brainstem structures in the generation of consciousness has been underestimated.[339]

How successful have doctors been at treating defects in this complex structure? In the late 1960s, Rolf Hassler and his team tried to restore consciousness in an unresponsive patient by inserting electrodes into the basal part of the right pallidum and

333 To give just two examples of world-leading experts in their fields whose predictions turned out to be utterly misguided: no less a person than Albert Einstein remarked that 'there is not the slightest indication' that nuclear energy 'will ever be obtainable' (*Pittsburgh Post-Gazette* 1934). And when Thomas Edison was asked what technological progress the coming hundred years would bring, he answered that iron could soon be turned into gold since scientists were 'already on the verge of discovering the secret of transmuting metals' (*The Miami Metropolis* 1911).

334 See also Unger's (1990, 143–145) thought experiment featuring the *Taped Brainstem*. His conclusions are, however, marred by a conflation of the presence and absence of wakefulness with the distinction between conscious and unconscious information processing in the brain. For the latter, see Block (1995). See also McMahan's (2002, 442 f. and 1995, 104–108) excellent discussion of the relevance of the ARAS to his own account of personal identity.

335 Damasio 2010, 159; see fig. 1 in section 4.3.

336 Edlow et al. 2012, 531; see also Jang / Kwon 2015, 200 f.; Zeman 2006, 363; Parvizi / Damasio 2003, 1525.

337 Moll et al. 2009, 126, 140; Plum / Posner 1980, 13 f.

338 Machado 1999, 157; Robbins 1997.

339 Merker 2007.

the left latero-polar nucleus of the thalamus.[340] In reaction to this stimulation, the patient opened his eyes and exhibited spontaneous movements. Unintelligible vocalisation occurred and 'the level of consciousness was definitely improved'.[341] More recent studies have largely confirmed both the therapeutic value of electrical ARAS stimulation as well as the limitations of this method.[342] Has the destruction of the wakefulness component of consciousness become reversible?

It is important to understand that neither Hassler's group nor those who conducted the subsequent studies in fact substituted an electrical circuit for a defective ARAS. There is a big difference between stimulating a certain brain area and actually *replacing* its function. In the former case, electrical pulses interfere with neural networks at the target site to produce a desired outcome. This necessitates that the respective structures be largely intact. In the latter case, however, the entire structure would have to be exchanged for a substitute. What Hassler et al. achieved was restoring the function of an only partly compromised ARAS from a lower to a somewhat higher level of arousal:

> The *anatomically undamaged* neurones or parts of this non-specific system should be induced to take up again their spontaneous ascending activation of the cortex, necessary for EEG arousal and awareness, by long-term stimulation.[343]

Although a remarkable accomplishment, especially in view of the fact that the study was conducted as early as in the 1960s, this intervention is still very far from actually substituting a dysfunctional ARAS, and not much progress has been made since. In sum, it is at present at least technically impossible to replace or otherwise restore the function of an entirely destroyed reticular activating system. Whether this also presents a metaphysical impossibility, for instance, because the very arousal network that developed along with the cerebrum is needed for a particular individual to return to consciousness, we cannot say. The high complexity of the ARAS could be an indication that this may well be so.

What can one conclude from these considerations? Previously, we were concerned with conditions in which the disruption of psychological continuity apparently *precedes* our ceasing to exist: we established that disorders like Alzheimer's disease pose threats to memory-based views as the criterion of personal identity that the latter put forward entails that demented individuals have ceased to be although they still have thoughts, perceptions, and emotions. Then we investigated the opposite permutation. We asked whether there can be retention of psychological continuity that *succeeds* our ceasing to exist. We were therefore looking for a medical condition in

340 Hassler et al. 1969, 306.

341 Hassler et al. 1969, 308 f.

342 Koch 2010; Moll et al. 2009; Schiff et al. 2007; Yamamoto / Katayama 2005; Cohadon / Richer 1993; Sturm et al. 1979.

343 Hassler et al. 1969, 309; emphasis added. Due to the redundancy of connections from the ARAS brainstem nuclei to higher brain regions, the ARAS appears to have the capacity to recover when only a certain fraction of its pathways are severed (Edlow et al. 2013, 521).

which the Lockean criterion for diachronic persistence is fulfilled despite it being obvious that the subject can no longer be present.

We found that the persistence of the neural correlates of a person's standing mental states, and consequently of psychological continuity, is conditional on the cerebrum being oxygenated and supplied with glucose. Conversely, the retention of the capacity for consciousness is dependent on the integrity of *two* anatomically distinct loci: the cerebrum, which contributes the awareness component to consciousness, and the ARAS, which is responsible for wakefulness and originates in the brainstem. Since wakefulness is a precondition of awareness, the destruction of the ARAS alone results in the permanent loss of consciousness. However, due to the fact that the neural correlates of the mental states that constitute an individual's long-term memory do not overlap with the core area of the ARAS, psychological continuity remains unaffected by damage to the latter. The Lockean condition of a person's diachronic existence is therefore still fulfilled – although it is obvious that, devoid of the structural prerequisites for wakefulness, the subject must have ceased to exist.

Hence, as in the case of Alzheimer's disease, memory-based accounts of personal identity have untenable consequences when confronted with neurophysiological facts and clinical data. At least in their present form, these views carry the absurd implication that psychological persons continue to exist even when they neither can, nor ever will, awake or even dream. When there are real-life cases in which the Lockean condition for diachronic existence is fulfilled in spite of the irreversible absence of any conscious mental activity, all motivation for holding on to what purports to be a *psychological* view vanishes.[344]

5.3 Consciousness-Based Accounts

5.3.1 Disruption of the Stream of Consciousness

Long-term memories cannot be the essence of our being and, consequently, nor can psychological continuity be the criterion of our diachronic persistence – for this continuity is disrupted in conditions during which we are evidently *still in existence*, like Alzheimer's disease, whereas it persists undisturbed in circumstances in which we have apparently *already ceased to exist*, like an ARAS defect.

We have now gradually excluded all possible candidates for the essence of our being, bodily and mental, with the exception of one: consciousness. The preservation of conciousness is what appeared to be a strong indication of the presence of a subject when nearly all other mental relations were extinguished in Alzheimer's disease; and the permanent absence of consciousness is what made us doubt that a subject is still in

344 A modified version of this chapter is forthcoming in *The Journal of Medicine and Philosophy* under the title *Memories without Survival: Personal Identity and the Ascending Reticular Activating System* (Meier 2020b).

existence even though other important relations continued to obtain after an ARAS defect. In the remainder of this chapter, we shall therefore try to ascertain whether consciousness could be the basis of our persistence.

Views that identify our diachronic existence with the mere persistence of consciousness while simultaneously denying that we are immaterial substances are newcomers to the debate about personal identity. Only a handful of accounts of this type have been proposed that gained major recognition. In her 1986 book *Redefining Death*, Karen Gervais argues for a position that she terms Conservative Mentalism, according to which an individual remains in existence as long as the capacity for some mental life is preserved, regardless of whether this mental life continues to be associated with a personal history in the form of specific memories.[345] Only when the cerebral substrate ceases to support consciousness, personal identity is disrupted and the individual goes out of existence.[346]

Four years later, Peter Unger published his Core Psychology Approach. Unger suggests that personal identity consists in the continuity of an individual's basic cognitive capacities. These capacities are of the kind that all entities must have in order to count as human mental subjects, but they do not permit the individuation of a particular subject as, for example, specific memories would do. The psychological core capacities are underpinned by a sort of physical continuity, which is why Unger allocates his view to the physical rather than the psychological types.[347]

The most prominent and comprehensive example of a consciousness-based approach is Jeff McMahan's Embodied Mind Account. According to this view, what grounds a subject's diachronic persistence as well as her rational concern for the future is the sameness of mind. A mind is individuated by reference to its physical embodiment, on which its persistence is ontologically dependent. Hence, 'the continued existence and functioning (…) of enough of the same brain to be capable of generating consciousness or mental activity' is both a necessary and a sufficient condition for each of us to continue to exist.[348] Like Unger and Gervais, McMahan does not deem the retention of standing mental states essential.

The three authors arrived at their respective positions through considerations that deviate significantly from those presented in this thesis, and we cannot here go into the details of the differences between these subtypes of consciousness-based views. Rather, we shall ask the more general question of whether accounts of this *sort* are more plausible than the views of personal identity that we have already rejected. And if so, how they could be constructed in a way that conforms to the empirical facts that we have established.

Once again, we shall start with an historical author. Wondering what our essence might be, Descartes famously came to the conclusion that 'it is thought; this

345 Gervais 1986, 127 f.
346 Gervais 1986, 126.
347 Unger 1990. For a critique, see Shoemaker 1992.
348 McMahan 2002, 68.

alone cannot be stripped from me. (…) I am therefore, speaking precisely, only a thinking thing'.[349] In modern terminology, one may paraphrase this position as highlighting the capacity for being conscious *per se*, while deeming unimportant the retention and recurrence of specific mental content. As in the three contemporary accounts just presented, it is then not decisive whether an individual can remember, for example, seeing an overflowing of the Thames, as Locke held, but rather that he or she can enjoy *any* mental experiences to begin with, regardless of whether they are of past autobiographical events or not.

Defining personal identity in this way raises an obvious problem, however: we are not always thinking. What makes someone the same individual in view of the many interruptions in the stream of consciousness if not standing mental states that persist through these intervals?[350] When we go to bed, we anticipate what we will be doing the next day. When we awake in the morning, we remember most of the plans that we had made before we fell asleep, and we still possess the same beliefs, desires, personality traits, and many other mental states. Our memory seems to be, as it were, the glue that holds together successive episodes of conscious awareness. This is why Lockean theories of personal identity appear so convincing. Most of us believe, as Parfit contends, 'that it does not matter if there are interruptions in a stream of consciousness. This does not matter because these interruptions do not destroy psychological continuity'.[351] In the light of what we have established in the foregoing sections, however, we must now reply that if indeed these interruptions do not matter to our diachronic persistence, this cannot be attributed to psychological continuity in the sense of the persistence of standing mental states; for just as we have reason to believe that our awakening in the morning is a continuation of the life of numerically the same individual who went to bed on the day before, we also have reason to believe that this must equally be true of amnesic and demented individuals. If psychological continuity in the Lockean and Parfitian sense was responsible for our persistence through periods of unconsciousness, those who suffer from Alzheimer's disease would cease to exist at night while the rest of us would survive. This is too grotesque a position to adopt.

Descartes was aware of the problem that sleep and other intervals of absent consciousness pose to our diachronic persistence. He wondered:

> I am, I exist, this is certain. But for how long? Certainly only for as long as I am thinking; for perhaps if I were to cease from all thinking it might also come to pass that I might immediately cease altogether to exist.[352]

Descartes therefore assumed that the mind never stops thinking, not even in a dreamless sleep.[353] Today we know that this hypothesis is almost certainly false.[354] While even

349 Descartes 2008, 19.
350 Fuchs 2017b, 299–302; James 1890, 1:237–239.
351 Parfit 1987, 292.
352 Descartes 2008, 19; see also Gallie 1936, 35.
353 Descartes 2008, 188.
354 Strictly speaking, Descartes' conjecture is not falsifiable. Being an immaterial entity, the *res cogitans* might still continue thinking without this activity being accompanied by any bodily manifestations.

during seemingly dreamless sleep the mind is not necessarily idle,[355] there are intervals in which, judging from EEG-readings, f-MRI scans, and other third-personal evidence, mental activity apparently indeed comes to a complete halt. Such situations can occur naturally, for example, in hypoxic or hypoglycemic coma, but one can also induce them artificially by pharmacological means or by lowering the brain's temperature until electrocerebral silence is observed.[356]

Any account of personal identity that can only describe these periodical absences of mental activity by stipulating that the individual goes out of existence and subsequently comes into existence again, or that the original individual ceases to exist and is replaced by a numerically different subject upon regaining consciousness, would not perform any better than memory-based views when applied to Alzheimer's disease. Not only would such a description be in conflict with our intuitions; many authors believe that it is also metaphysically unsound to assume that one and the same thing could have two beginnings.[357] We are convinced that the intermittent phases of unconsciousness are not periods of nonexistence but form part of our lives, that we continue to be, even when we neither think nor dream. If only fully conscious moments counted, most people would, given that humans spend approximately one third of their time sleeping, not even reach the age of sixty.[358] Any account of personal identity that has these implications would rightly be regarded as unconvincing.

Biological views are, of course, not plagued with this difficulty. Dreamless sleep, coma, and other situations that induce gaps in the stream of consciousness do not pertain to the organism's integrity, and as long as our organisms persist, advocates of these views hold, so do we. It is worth noting, however, that biological accounts face an analogous problem, namely, cryopreservation. The idea of cryopreservation is entirely to suspend an organism's life processes for a potentially indefinite time span by subjecting it to temperatures below -130 °C and later to rewarm it in the hope of revival. While attempts to apply this procedure to complex organisms have failed, cryopreservation has successfully been achieved in very simple creatures such as drosophila melanogaster.[359] Proponents of biological accounts must be able to explain what happens to these animals. Is the organism that was frozen numerically identical with the one that is later thawed despite the fact that all biological functions that animalists usually identify with life and with the diachronic persistence of organisms came to a temporary halt? Has a new organism come into being that just happens to share all of its matter with the original body?[360] Or should animalists permit a state of suspension between life and death?

355 Windt / Nielsen / Thompson 2016. For an overview of the similarities and differences between slow-wave sleep and coma, see Plum / Posner 1980, 18 f.
356 Stecker et al. 2001, 20.
357 See, for example, Locke 2008, II.XXVII, § 1.
358 Birnbacher 2017, 24 f.
359 Koštál et al. 2012. For an evaluation of the future prospects of cryonics, see Merkle 1992.
360 Quante 2002, 130 n. 13.

Thus, biological accounts, too, must provide a rationale for phases of functional inactivity. Admittedly, while this is a vexed problem, it is not as pressing as the analogous difficulty for views that deem consciousness the essence of our existence since cryopreservation of humans is not yet a feasible procedure and may never even become one, whereas sleep, anaesthesia, and coma are everyday phenomena. If one wants to argue for a consciousness-centered account of personal identity, one must therefore provide a good explanation of how it is that psychological subjects persist through their unconscious phases.

Whatever is responsible for our existence during the interruptions of the stream of consciousness must, it seems, fulfil three conditions. First, it must persist even when the stream of consciousness is terminated. Occurrent mental states do not meet this requirement.[361] Secondly, it must be present in all individuals of whom we can safely assume that they are not relevantly different from ourselves and thus include amnesiacs and demented people. Standing mental states fail in this regard, as we have established. Finally, it must not comprise the whole organism for we have determined that we can exist as isolated heads.

What entity could satisfy these requirements? Consciousness cannot occur without a carrier substance. Descartes thought that this substance was an immaterial soul, so that 'the corruption of the body does not cause the mind to perish'.[362] Today it is universally accepted and supported by an array of empirical evidence that the carrier substance of consciousness is material rather than immaterial, that this substance is the brain, and that its corruption does cause the mind to perish.[363] The brain also conforms to all three requirements. Unlike the stream of consciousness, it persists through episodes of temporary mental inactivity. It also remains present in conditions in which long-term memory is lost, and it seems to be capable of functioning in the absence of the original organism. Referring to the brain therefore solves a major difficulty of consciousness-based views of personal identity. It is numerically the same consciousness that emerges after periods of sleep or coma by virtue of this consciousness being generated by the same brain.[364]

5.3.2 The Brain as the Substrate of Consciousness

While the postulation of immaterial substances comes with the luxury of a certain mysteriousness that surrounds their properties, referring to physical things like brains does not. When one invokes the brain, one must also specify what relation we are supposed to bear to it. This is what we shall be trying to do in this section.

361 Price 1969, 247.

362 Descartes 2008, 11.

363 If the human brain could be emulated on a non-neural basis and if we had reason to regard the resulting mental phenomena as relevantly similar to our own, the existence of a brain would, of course, not be a necessary condition for there to be consciousness, but we will not entertain this highly speculative possibility here. See also note 331.

364 McMahan 2002, 68; Shoemaker 1992, 140; Unger 1990, 25 f.

One option that comes to mind is that it may simply be a relation of identity. If the persistence of the substance brain to which consciousness is somehow attached is what accounts for the fact that we continue to exist through intervals of absent mental activity and ensures that numerically the same individual re-emerges, it does not seem entirely implausible to assume that we just *are* our brains. Let us briefly consider this idea.

Lenin died in 1924. Normally, a body begins to decompose shortly after circulation stops, but in Lenin's case, Soviet authorities decided to have his body preserved in formaldehyde and his brain extracted for subsequent study. The brain was cut into thousands of slices to facilitate microscopic analysis.[365] One may pretend that the slicing did not take place. If this was the case, and assuming that the application of a tissue fixative is compatible with the persistence of a physical thing,[366] the brain would still exist in its entirety. If we were identical with our brains, it would follow that Lenin did not cease to be in 1924 but that he still exists today. This is clearly absurd. Lenin has gone out of existence, and so have the thousands of people whose brains are floating around in the jars of anatomical collections in numerous university cellars.

What, then, has disappeared from these brains that ended those people's existence when it is not the molecules of which the brains consist? Let us analyse the process that dying brains normally undergo and see if we can ascertain at which point in the chain of events our existence diverges from that of our brains. The pathophysiological cascade begins when an event interrupts the subject's ongoing stream of consciousness. This disturbance may happen either in the brain itself or in another part of the organism. *Intracranial* events include hemorrhage, swelling, and infection; *extracranial* events include cardiovascular and respiratory diseases. The brain reacts to oxygen and glucose shortage by terminating the stream of consciousness after three to five seconds, thereby lowering its metabolic demands. Electrocerebral silence occurs fifteen seconds thereafter, at which point not only conscious activity but all mental activity ceases, albeit not yet irreversibly.[367] If the cerebral circulation can be restored within approximately three minutes, the subject is likely to regain consciousness.[368] So far, our persistence conditions and those of our brains seem to overlap since death requires irreversibility, which the mere absence of the stream of consciousness does not entail.[369]

365 No conclusive correspondence between Lenin's cognitive abilities and his brain's morphology could be found (Adrianov et al. 1993).

366 It is quite plausible to assume that the things stored at the Moscow Brain Institute are nothing else than slices of Lenin's brain. Those who doubt this interpretation should consider cryopreservation instead of the use of formaldehyde.

367 Schlake / Roosen 2001, 25.

368 Under normothermic conditions. As described in section 4.4.3.1, hypothermic conditions prolong this period.

369 In section 5.2.2.2, we have moreover determined that even the *final* slipping into unconsciousness, that is, the beginning of the comatose phase from which the subject will *de facto* not re-emerge, is not any different in this regard.

If anoxia persists longer than this time span, ischaemic damage to the tissues ensues. Only ten minutes following the onset of ischaemia, a significant number of cells already show clumping of nuclear chromatin. After thirty minutes, cytoplasmic swelling increases, mitochondria change in shape, microtubules disappear, and ribosomes detach from the cisternae of the endoplasmic reticulum.[370] In the course of this process, the brain is losing the structural prerequisites for generating consciousness. Even if doctors manage to remove the initial cause of the disorder and succeed in restoring cerebral perfusion, it is now too late. Consciousness is not going to return.[371] This appears to be the point at which our persistence conditions diverge from those of our brains, only the latter continuing to exist.

Eventually, decomposition sets in.[372] Cells lose their structural integrity as autolysis breaks down tissues. One can suspend this process by applying an embalming agent, as in Lenin's case, or by freezing the organ. These procedures, however, cannot preserve any functionality in the brain. Even if they could be carried out before anoxia inflicts irreversible damage, their invasiveness would still destroy the brain's delicate synaptic circuitry beyond repair. Tissue fixatives are toxic and the formation of ice crystals in cryopreservation results in mechanical damage. Consequently, irrespective of whether the brain's constituent matter is scattered or preserved, no psychological subject will henceforth be attached to it.

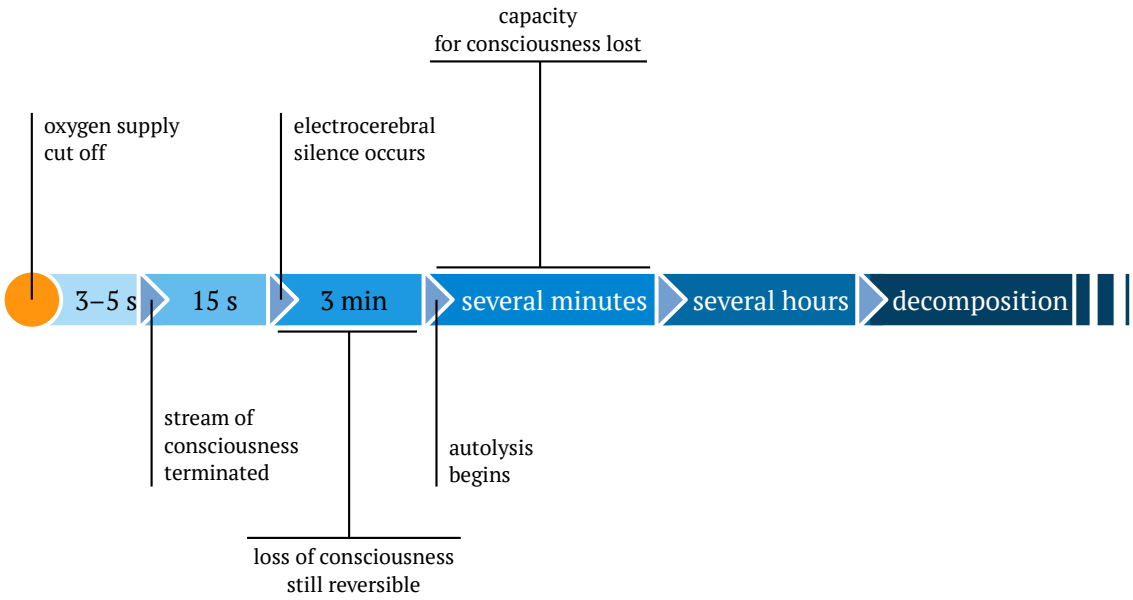


Fig. 4: Timeline of events in a dying brain.

370 Aggarwal et al. 2010, 14.

371 Since the brainstem is more resistant to anoxia than the cerebrum, extracranial events will affect the former slightly later. If circulation is restored at a point in time at which the cerebrum has already irreversibly lost its function while the brainstem is not yet fatally damaged, the result is the persistent vegetative state, characterised by a dissociation between wakefulness and awareness (see section 3.3).

372 Decomposition of the brain may even occur while the rest of the body is still successfully being externally ventilated, in which case one speaks of *intravital autolysis* or *respirator brain* (Moskopp 2015, 72 f.).

Of these three major events that occur in an unoxygenated brain – the termination of the stream of consciousness, the irreversible loss of the capacity for consciousness, and the decomposition of the macrostructure – it is therefore clearly the loss of the capacity for consciousness that marks the end of our existence.[373] This capacity amalgamates two elements: the potential to generate the stream of consciousness and the physical basis required to bring forth this phenomenon. Thus, it refers to both the immaterial phenomenon and the material substrate. Invoking the capacity instead of the stream of consciousness means not requiring that the actual phenomenon be continuously realised, which solves the gap problem with which we began. Rather, it necessitates that the structural prerequisites for the phenomenon to occur be physically manifested without interruption, so that consciousness remains poised to be realised. The material carrier in which consciousness is realised, or on which it supervenes, can acquire, retain, or lose this disposition. Fetal brains have yet to acquire it. Developed brains that are constantly being oxygenated and supplied with glucose retain it. And brains that are subjected to anoxia for more than three minutes are beginning to lose it.

Specifying exactly what kind of relation the capacity for consciousness bears to the brain is virtually impossible without first having solved what David Chalmers termed the *hard problem* of consciousness, namely, how mental phenomena arise from a physical basis.[374] Despite significant efforts, neither neuroscience nor biology, psychology, philosophy, or any other discipline have succeeded in this task. Attempting to decide between the many models that have been suggested would therefore warrant a thesis on its own.[375] Instead, I shall – as an entirely preliminary proposal – just give an *example* of how one could approach describing the general metaphysics of the relationship between the capacity for consciousness and the brain.

Roughly speaking, there are three main ontological positions that one may use to describe metaphysical relations: the traditional Aristotelian substance-accident framework, four-dimensionalism, and the constitution view. The *Aristotelian framework* distinguishes between substances and modes. A *substance* is the material substrate of which a thing is made – like a lump of bronze. A *mode* (or accident) is a quality that this substance has – like having the shape of a statue.[376] In our case, the capacity for consciousness, and thus the psychological subject, would be a mode of the substance brain.

373 Only when the brain is abruptly destroyed by a massive mechanical force, for example, in a severe transportation accident or due to a gunshot wound to the head, is it that the *stream* of consciousness and the *capacity* for consciousness cease simultaneously. In all other cases, the loss of consciousness precedes the loss of the capacity for consciousness by a short time span – the time span during which consciousness could potentially be restored. And the loss of the capacity for consciousness, in turn, precedes the deformation of the brain's macrostructure by hours or even by years, as in Lenin's case.

374 Chalmers 1995, 212; see also Nagel 1974, 435–437 and Leibniz 2014, 70.

375 See note 305.

376 Aristotle 1995, 3492–3537 (*Metaphysics*, Book Z).

Four-dimensionalism conceives of time as analogous to space. According to this doctrine, things have temporal parts just as they have spatial parts, and thus they extend through both time and space. Unlike three-dimensionalists, who believe that objects persist by *enduring* (from t_1 to t_2, objects are always present in their entirety; while they have spatial parts, they have no temporal parts), four-dimensionalists hold that objects persist by *perduring* (at any point between t_1 and t_2, only a temporal part of the object is present).[377] Applied to the question at issue, four-dimensionalism would conceive of the psychological subject as a proper temporal part of her brain that extends in time for as long as this brain possesses the capacity for consciousness.

According to the *constitution view*, there exists a relation of unity between certain objects that is intermediate between identity and separate existence. It is neither a mereological relation, that is, one that things bear to their parts, nor simple spatial coincidence, nor supervenience. Rather, if *x* constitutes *y*, then *x* and *y* are of different kinds and can survive different sorts of changes. A lump of bronze, for example, may constitute a statue. Melting it makes the statue disappear, but not the bronze itself.[378] If one used the constitution view to explain the relationship that we bear to our brains, one would hold that we are constituted by our brains.

Thus, one can in principle describe the relation between psychological subject and the underlying physical substrate with the help of any of these three ontological positions. They each have their virtues and vices, which are discussed elsewhere and which we do not need to rehearse.[379] From a practical standpoint, the aim of this thesis is to achieve the highest possible clinical applicability of the conclusions drawn. This meant relinquishing the use of fantastical thought experimentation, and it may also mean relying as little as possible on controversial metaphysical frameworks and revisionist ontologies.

The four-dimensional framework may be deemed unappealing for medical purposes since both clinical personnel and relatives conceive of patients as *enduring* entities, that is, as things that are wholly present rather than as mere time slices of such things. Especially in a profession that places great weight on cross-temporal ways of thinking, for example, when establishing case histories and making prognoses, regarding patients as temporal parts will appear unnatural.

In the case of the constitution view, on the other hand, it is very difficult to understand to what the character of the relation of constitution really amounts. While we have a firm intuitive grasp of the concepts of numerical identity and spatially separate existence, it is nebulous to many people, not only to non-philosophers, how there could be an intermediate ground between the two and wherein it could possibly consist. It therefore seems that the constitution view is not a framework that would be suited for clinical implementation either.

377 Hawley 2014, 49; Hawley 2001, 11–16.
378 Baker 2000; Baker 1999.
379 See, for example, Hasker 2004; Sider 2001; Merricks 1999.

Avoiding revisionist metaphysics simply for practical reasons is not convincing argumentatively, especially since the remaining option, the traditional Aristotelian framework, has its own flaws. Ultimately, however, none of the arguments presented in this thesis hinge on the ontological position that we choose for describing the relation between the capacity for consciousness and the brain. Unless we know more about the mechanism by which nervous tissue gives rise to mental phenomena, any account of this relation is bound to be speculative. Let us therefore, without further discussion, select the framework that promises to yield the best clinical applicability and that is most easily compatible with most people's pre-reflective understanding of the world.

As mentioned above, Aristotelian ontologies distinguish between material substances and their qualities, referred to as *modes* or *accidents*. Substances exist independently of other entities. Conversely, modes cannot exist unattached. They depend in their persistence on the substance whose modes they are.[380] The form of a certain statue, for instance, cannot exist without the lump of bronze from which the statue is made. Mapped onto this framework, the relation between the brain and the psychological subject could be described as follows. There is a substance – the brain – that has various modes. One of these, namely, the capacity for consciousness, is of direct relevance to the persistence of the psychological subject, while others, for instance the ability to direct voluntary motor tasks, are not. All of these modes are ontologically dependent on the brain. When the brain is destroyed, they, too, cease to exist. As we have established in the previous section, the converse is not always the case. The brain can exist devoid of some of its modes. Stripped of the capacity for consciousness, for example, the brain can still persist in a coma or, as Lenin's brain, even fixed in formaldehyde. The respective ontological dependencies, just like those between the statue and the lump of bronze, are consequently asymmetrical.[381]

A psychological subject persists if and only if the capacity for consciousness is retained. As we have determined in section 4.3, this capacity is a product of the two factors wakefulness and awareness. Consequently, a brain only retains the capacity for consciousness if it possesses both of these accidents. If it merely possesses the capacity for wakefulness, the result is a persistent vegetative state; if only the capacity for awareness remains, the result is an irreversible coma. The capacity for consciousness may therefore be described as a second order mode of the brain that is only preserved when these two more basic modes are retained.

Generally speaking, modes are explicable in terms of the way in which their underlying substance is structured.[382] While we do not yet know how exactly nervous tissue brings forth mental phenomena, it is very likely that this also goes for the modes to which the brain gives rise. Thus, what is required for a psychological subject to persist is not just the substance per se, but this substance microstructurally organised in

380 Locke 2008, II.XXIII, § 1.
381 Lowe 1995, 68.
382 See also Locke 2008, II.XXIII, § 3.

such a way that it has two specific modes which, combined, constitute the capacity for consciousness.

The difference between a brain whose structuring entails the preservation of these modes and a functionless brain that only retains its macrostructural shape is best viewed as to some extent analogous to the difference between an organism and a fresh corpse. Organism and corpse are composed of the same matter, have the same weight, and share most other properties. An organism turns into a corpse when integrated functioning has irreversibly ceased. This irreversibility requires that certain structural changes have taken place, for otherwise the loss of function would only be temporary. Initially, that is, shortly after death has occurred and decomposition has not yet set in, these changes do not exceed microstructural modifications that are invisible to the naked eye, while the body's macrostructure remains intact.[383]

Similarly, a brain that has undergone the transitional period of oxygen and glucose starvation in the course of which it is rendered functionless remains the same substance until it is cremated or autolytic decomposition finally liquifies it. For some time, it continues to possess many of the properties that it also had before the onset of ischaemia. But just like in the case of a dying organism, the above-named microstructural changes turn the organ from a functional into a dysfunctional one. These changes entail the loss of certain accidents, one of which is the capacity for consciousness.

The idea that our persistence as psychological subjects shall consist in another substance retaining an accident will not appeal to everyone. Whether or not we must be substances in their own right is a question that has been debated extensively throughout the history of philosophy.[384] Especially authors who work in the tradition of Descartes, Reid, or Butler will find it misguided to conceive of ourselves as mere modes of a separate entity. They believe, as many other philosophers do, that whatever we essentially are must be ontologically independent.[385] Reid wrote:

> My personal identity, therefore, implies the continued existence of that indivisible thing which I call *myself*. Whatever this self may be, it is something which thinks, and deliberates, and resolves, and acts, and sutlers. I am not thought, I am not action, am not feeling; I am something that thinks, and acts, and suffers. My thoughts, and actions, and feelings, change every moment; they have no continued, but a successive, existence; but that *self*, or *I*, to which they belong, is permanent.[386]

383 Compare also the difference between animalism and Bernard Williams' Bodily Continuity View: while animalism sees our identity conditions in the persistence of organisms, that is, of *living* bodies (Olson 1997, 106), the Bodily Continuity View stipulates that we are bodies simpliciter, irrespective of whether they still perform any biological functions (Williams 1970). Another proponent of the latter account is David Mackie, who holds that 'biological organisms persist as long as this organisation of constituent parts remains sufficiently nearly intact' (Mackie 1999, 237). See also Carter 1988. The contrast between these two views is therefore also sometimes expressed as one between *organic* (or biological) and *somatic* (or non-biological) animalism.

384 Gallie 1936, 28 and 31.

385 Shoemaker 1997, 287.

386 Reid 1878, III.III, § 2; see also Butler 1927, 257–263.

Like Descartes, Reid and Butler were, of course, substance dualists – a view that we have excluded in the very beginning.[387] Those who endorse accounts that rely on psychological continuities of some sort will be more sympathetic to the idea that our persistence need not necessarily be the persistence of a substance. Locke, for example, submitted: 'Place that *self* in what substance you please'.[388] And Parfit, like other modern reductionists, held that persons are not separately existing entities, but that their persistence is fully analysable in terms of the continuity of mental states.[389] After all, the identity conditions that he put forward are supposed to permit teletransportation in which only information but no matter is exchanged.[390] Hume took an even more radical stance, arguing that we 'are nothing but a bundle or collection of different perceptions, which succeed each other with an inconceivable rapidity, and are in a perpetual flux and movement'.[391] Conceiving of ourselves as mere accidents is therefore not unprecedented historically.

When we introspect, this position seems impossible to be true. It appears seducingly obvious that our existence cannot simply be a mode of another entity but must consist in something so foundational as the independent substances that Descartes, Reid, or Butler postulated. One reason for this conviction may be that we can only reflect on our existence *while* being conscious. Introspection thus presupposes the presence of the very property whose persistence conditions are under investigation. There is no external viewpoint that one could occupy. That epistemic access coincides, without exception, with the presence of consciousness is deceptive and may persuade one to view this property as much more stable than it in fact is. Therefore, through the biased lens of introspection, our mental world seems like a substance in its own right, independent in its existence of any other entity. The idea that a minute structural change in the brain can extinguish this world forever is bound to strike us as nearly absurd.

Things appear in a different light when we assume a third-personal perspective, as it permits us also to witness the *absence* of consciousness, even if only from the outside. Seeing someone lapse into an irreversible coma after a small localised brain event is a sobering observation. Suddenly, we appreciate how fragile our mental existence is: a few minutes of cerebral anoxia, the exposition to a tiny amount of certain chemical substances, or a lesion of the size of a sugar cube in a strategically important area[392] are all sufficient to put a sudden end to it. None of these interferences necessitate that the brain's macrostructure be significantly altered. To conceive of ourselves as accidents of this substance that outlasts all these changes becomes much more plausible

387 See section 4.2.1.
388 Locke 2008, II.XXVII, § 16.
389 Parfit 1987, 251.
390 Parfit 1987, 199.
391 Hume 1896, 252.
392 Koch 2010, 20.

from the third-personal perspective than it appears from one's introspective viewpoint.

Our persistence may therefore indeed not be nearly as ontologically independent as traditional accounts of personal identity that postulate substances like souls,[393] organisms,[394] or bodies,[395] make us believe. In the case of the statue made of bronze, radical interventions are required to eliminate its substance. Only much smaller modifications are needed, however, to make the bronze lose its specific form, so that the statue, but not the substrate, disappears. Analogously, the minuscule alterations in the brain that are sufficient to deprive it of its capacity for generating consciousness need not extinguish the substance itself to make the psychological subject disappear forever. If one is to defend a psychological (rather than a biological) view of personal identity, it is therefore indeed appropriate to conceive of ourselves as accidents rather than as substances. And if such an account is correct, each of us is a psychological subject whose diachronic existence consists in the persistence of the capacity for consciousness realised in a particular brain.

5.3.3 Disruption of the Substrate of Consciousness

In contrast to the immaterial substances that Descartes, Reid, and Butler postulated as the essence of our existence, physical substances like brains can be modified in certain ways. Examples of particularly radical modifications are physical trauma, large-vessel strokes, and brain surgery.[396] To be able to establish which modifications of their substrate psychological subjects can survive, it is imperative to determine how much of the brain's constituent matter must be retained for the capacity for consciousness to be preserved. In this section, we shall therefore examine three medical procedures in which the functional areas of the brain are significantly diminished and try to establish to what extent these interventions interfere with the generation of consciousness. That large parts of the brainstem – especially the reticular formation from which the ARAS originates – are required for wakefulness to manifest, which also makes their integrity a prerequisite for being conscious, we have already noted in section 5.2.2.1. Here, we will focus on the upper brain. Is the whole cerebrum necessary for retaining the capacity for consciousness?

393 Plato 1997, 61 (*Phaedo*, 70c-d).

394 Blatti / Snowdon 2016; Blatti 2012; DeGrazia 2005; Wiggins 2003; Merricks 2001; Olson 1999; Olson 1997; Inwagen 1995; Snowdon 1990; Wiggins 1980.

395 Mackie 1999; Thomson 1997; Carter 1988; Williams 1970.

396 Since our brains consist of organic tissues in which various life-sustaining chemical reactions take place, small and gradual modifications occur all the time. The cerebral metabolism is highly active and consumes a significant amount of the organism's oxygen resources. Substances are transported within and between cells to ensure energy supply and to facilitate the elimination of wastes. Compounds are synthesised to maintain various structures. However, the brain's cellular renewal rate is lower than that of most other organs. Neurogenesis ceases even before birth. With the exception of the hippocampus and the olfactory bulb, neurons in the brain do not regenerate during an individual's life (Lledo / Alonso / Grubb 2006, 190).

One of the few real-life medical phenomena that have been discussed in the debate about personal identity with significant attention to empirical detail is the so-called *split brain*.[397] Split brain allegedly occurs as a side effect of commissurotomies, which are neurosurgical operations to treat pharmacoresistant epilepsy. This side effect may provide important insights into how much cerebral tissue is required to bring forth a stream of consciousness.

We shall begin with a brief description of the operative procedure. Our cerebral hemispheres are connected via several fibre tracts that cross the midline. The largest of these so-called *commissures* is the corpus callosum, which contains more than 200 million axonal projections that that mediate cross talk between homologous cortical areas.[398] The surgical division of these fibres prevents the spreading of seizures, but it also appears to have the consequence that some information processed in one hemisphere becomes unavailable to the respective other.[399] While under ordinary circumstances the patient's behaviour reveals nothing unusual, controlled lateralised testing yields puzzling results: when, for instance, optical stimuli are arranged in a way that they just reach a single eye, only the hand that is controlled by the contralateral hemisphere will be able to draw the perceived object. Likewise, if a certain scent is presented to one nostril, the patient will be incapable of re-identifying it when it is subsequently offered to the other nostril. In each case, the sensory information seems to be accessible to consciousness only in one hemisphere while being withheld from awareness in the respective other one.[400] In very rare cases, there even appear to be conflicts between the hemispheres, like in one patient whose hands allegedly chose different clothes to wear in the morning.[401]

Various interpretations of this side effect that commissurotomies yield have been proposed. Authors suggested that the person's phenomenal consciousness is briefly split under certain experimental conditions, whereas otherwise it remains unified; that the operation results in two persons with two separate minds; that the patient remains a single person who undergoes two streams of consciousness simultaneously; that the procedure gives rise to a single mind and an unconscious automaton; and that it only happens to reveal the fact that the patient possessed two minds all along. These are just some of the more popular of the many explanations that were offered.[402]

There is much to be said both for and against each of these interpretations. However, the puzzle with which split brain presents us, fascinating as it is, has been

397 See, for example, Schechter 2015; Bayne 2008; Noonan 2003, 5 f.; Tye 2003, 109–132; Wilkes 1999, 132–167; Gillett 1986; Nagel 1985; Puccetti 1981; Puccetti 1973; Parfit 1971. Recently, Elizabeth Schechter published a whole monograph on this topic (Schechter 2018).
398 Park et al. 2017, 710.
399 Sperry 1984, 661 f.
400 Sperry / Preilowski 1972, 922.
401 Wilkes 1999, 140.
402 Tye 2003, 111–113.

discussed extensively for decades.[403] For the purpose of this section, we are only interested in whether the side effects of commissurotomies can be helpful in determining how much of the brain is required for consciousness to manifest. The idea is this: if one could conclusively show that split brain is a condition in which two centres of consciousness are associated with one brain, each of them residing in one hemisphere, it would follow that one half of the brain is sufficient for retaining the capacity for consciousness. The problem, as mentioned, lies with the antecedent. That the operation yields two centres of consciousness that operate concomitantly is just one of several unconfirmed hypotheses. Moreover, one should bear in mind that operations as invasive as commissurotomies are only ever performed on patients whose epilepsy is so severe that all other treatment options have already been exhausted. These patients typically suffer from significant brain damage even before the hemispheres are divided, which may skew the results of post-operative testing.[404]

Furthermore, and more importantly, even when the hemispheres are completely disconnected along the midline, they still attach to the same brainstem via subcortical connections.[405] While these pathways may not be crucial for the cross-hemispheric exchange of information, this fact is noteworthy in the light of our findings from section 5.2.2. Consciousness, we determined, is a product of the two dimensions wakefulness and awareness, of which the former is a function of the ARAS, which originates in the brainstem. Since the division that commissurotomies induce does not extend to the lower brain, the brainstem structures that are responsible for providing wakefulness are not severed but remain shared between the hemispheres. As we have established in section 2.3, one *could* not even divide the brainstem while preserving its function since, unlike the cerebral hemispheres, it is not a binate structure. Consequently, irrespective of how one interprets the patient's postoperative behaviour, one should, for neuroanatomical reasons, not accept the procedure as evidence of a structural separation of the neural correlates of *both* dimensions of consciousness; yet this is what would be required to regard centres of consciousness as truly independent.[406]

Fortunately, there are two other medical procedures, a surgical and a pharmacological one, that deliver the desired empirical data but avoid this difficulty: hemispherectomy and the Wada test. Like commissurotomies, anatomical hemispherectom-

403 See note 397.

404 Wilkes 1999, 140.

405 In the vast majority of operations performed, there is not even a total separation along the midline of the *upper* brain because doctors only transect the corpus callosum and spare the smaller anterior and posterior commissures.

406 It is often argued that split brain poses a serious threat to psychological views of personal identity for the same reason that we identified as problematical for biological accounts when considering the separation of head and body in section 4.4: the transitivity of identity would neither permit the original subject to be identical with both psychological individuals that commissurotomies allegedly yield nor with only one of them (cf. e.g. Nagel 1986, 43 f. and Parfit 1971). However, if the role of the indivisible ARAS in the generation of consciousness is appreciated, this objection shows itself to be unsubstantiated.

ies are offered as last-resort treatments for medically refractory epilepsy. Instead of only disconnecting the cerebral hemispheres along the midline, a whole hemisphere, or a large part of it, is removed.

In most adults, language is lateralised. To mitigate the risk of aphasia, hemispherectomies are therefore usually only performed on the hemisphere that is non-dominant for language.[407] If the operation is carried out before the patient has reached the age of five, however, the brain's neuroplasticity usually enables the latter to reorganise and, via contralateral function transfer, to achieve the restauration of this ability.[408] In this case, even the language-dominant hemisphere can be removed without the patient losing her linguistic abilities. Thus, while the extent of post-surgery cognitive and memory impairment varies depending on age, lateralisation of brain functions, and many other factors, this demonstrates that each hemisphere is capable of supporting consciousness on its own. Normally, hemispherectomy patients can therefore 'expect their cognition to be relatively unchanged after surgery'.[409] Many of them remain employable and lead independent lives.[410] The physical basis for a psychological subject to persist must consequently be significantly smaller than the whole brain.

One might object that in epilepsy patients the hemisphere to be removed is usually ridden with extensive tissue damage, which is what causes the seizures in the first place.[411] Hence, there is the possibility that the contralateral side of the brain has already taken over most functions, which might mean that non-epileptic people in whom such a reorganisation does not occur could not be conscious with just one hemisphere.

The Wada test brings clarity here. Introduced in 1949 by Juhn Wada, it serves to determine the lateralisation of language and memory in individual patients.[412] Before the advent of f-MRI, the test used to be an invaluable tool for the surgical planning of cortical resections. It works as follows. While the patient is awake, a barbiturate is injected into either the left or the right internal carotid artery via a catheter. Since these vessels, which arise from the common carotid arteries, each supply only one hemisphere, the anaesthetic also just reaches one side of the brain.[413] Each cerebral hemisphere receives sensory information from the respective contralateral side of the body, which is also the side whose motor functions it controls. Therefore, when the patient holds up both arms, the arm contralateral to the targeted hemisphere drops shortly

407 McGovern et al. 2019, 2416.
408 Silva et al. 2020, 1276.
409 McGovern et al. 2019, 2426.
410 McClelland / Maxwell 2007, 374.
411 Silva et al. 2020, 1279.
412 Quigg 2019. The Wada test features in two philosophical papers whose authors argue that the procedure, when employed over longer periods, could lead to the formation of two independent persons who are associated with the same organism. This scenario is then used as a counter-argument to biological views of personal identity (Reid 2016; Puccetti 1969, 75 f.).
413 Kundu / Rolston / Grandhi 2019, 2. For an overview of the arteries that supply the brain, see section 4.4.3.1.

after the drug has taken effect, indicating hemiplegia. As mentioned, one hemisphere is dominant for language in the vast majority of the population.[414] When this hemisphere is injected, the patient experiences impaired speech production or even loses her language abilities entirely. This permits doctors to determine language lateralisation and to establish which hemisphere should be spared during surgery. Memory functional dominance is assessed in the same manner. When the effect of the barbiturate has dissipated, doctors inject the contralateral vessel and repeat the neuropsychiatric testing.[415]

In a sense, the transient hemispheric anaesthetisation of the Wada test can be conceived of as a reversible hemispherectomy, and it confirms the conclusion that we drew from the latter. The two procedures clearly show that in addition to the brainstem only one cerebral hemisphere is required for consciousness to manifest. The actual amount of cerebral tissue that is needed for the generation of consciousness may, of course, be even smaller – the exact neural correlates of consciousness have not been found.[416]

Moreover, even if one did possess all relevant empirical data, it could still be the case that one would then be faced with a situation analogous to the one described in section 3.5, where we were inquiring which particular discontinuity marks the death of an organism: on the one hand, our concept of consciousness may be too imprecise to specify an even more precise threshold for it being present or absent (linguistic vagueness); on the other hand, neurophysiological matters may be indeterminate in and of themselves (ontic vagueness). For the purpose of this thesis, it will suffice to be able to define the minimal material basis of the capacity for consciousness, and consequently the amount of physical continuity of the substance brain required for the persistence of a psychological subject, as consisting of one cerebral hemisphere with the subcortical structures and the brainstem – and, in the light of what we have established in section 4.4.3.2, likely also a sensory organ that provides environmental stimuli.

5.3.4 How a Psychological Subject Ceases to Exist

Finding a solution to the problem with which we began required many detours – some of them simply to be able to pose the right questions. Now we are hopefully in a position finally to specify what conditions must obtain for the existence of the kind of entity that we are to come to an end.

We have determined that each of us persists as long as his or her brain retains the capacity for generating consciousness, that is, as long as the anatomical structures that are responsible for this ability remain functional; and that he or she ceases to exist when the brain irreversibly loses this capacity. This makes the retention of the ca-

414 Silva et al. 2020, 1276.
415 Kundu / Rolston / Grandhi 2019, 2.
416 Lagercrantz / Changeux 2009, 255; Zeman 2001, 1275.

pacity for consciousness both a necessary and a sufficient condition for us to continue to exist; and the loss of this capacity both a necessary and a sufficient condition for us to cease to exist.

We can now apply this conclusion about the question of personal identity to the neurophysiological findings from the previous chapters to establish what type of brain damage indicates that a psychological subject has ceased to exist. Put differently: we are moving from the level of the *definition* of death back to the level of the bodily *criteria* at which we began.

Since the stream of consciousness is a product of its two factors wakefulness and awareness, the preservation of only one of these elements is insufficient for retaining the capacity for being conscious and therefore for us to continue to exist. As discussed in section 5.2.2.1, wakefulness and awareness are provided by two anatomically distinct brain regions. Wakefulness is a function predominantly of the ARAS that originates from the reticular formation in the brainstem, whereas awareness is a function predominantly of the cerebral hemispheres. Hence, the destruction of either of these parts of the brain is sufficient to cause the irreversible loss of the capacity for consciousness and, consequently, also constitutes a sufficient condition for us to cease to exist.

When the reticular formation in the brainstem becomes dysfunctional, we cease to exist because we cannot *awake*. The result of this type of damage is an irreversible coma. Regardless of whether the cerebrum would be able to supply the awareness component of consciousness, these structural prerequisites of awareness remain dormant when the patient cannot be roused. In section 5.2.2.4, we explored the possibility of a dysfunctional ARAS being replaceable by an artificial device in the very distant future. If this could be achieved, the integrity of the original brain structure would lose its status as a necessary condition of our persistence. Consequently, its destruction would then also lose the status as a sufficient condition of our ceasing to exist. From today's perspective, the chances of such a development appear very low, however.

When, conversely, both cerebral hemispheres become dysfunctional, one may still awake, but *awareness* will be absent. This is the situation of patients in a persistent vegetative state and of anencephalic infants. Their functional brainstems bring forth periodical wakefulness, but in the absence of at least one intact cerebral hemisphere this arousal fails to be associated with any mental content. It remains empty and without a target. No stream of consciousness can manifest.[417]

In clinical reality, not all instances of brain damage are assignable either to the category of localised brainstem defects or to that of total decerebration, of course. Matters are more often than not much less unambiguous. Persistent vegetative states can result from only partial destruction of the cerebrum,[418] and ARAS defects can as-

417 Meier 2020a, 101.
418 Zeman 2002b, 217.

sume different degrees of severity.[419] Even more difficult to interpret are conditions like the minimally conscious state in which the cerebrum *and* the brainstem are affected diffusely and to varying degrees. The concepts developed in this thesis can therefore only serve as poles on the wide spectrum of configurations that impact on the persistence of the capacity for consciousness. Deviations from the relatively coarse categorisations described here must be assessed individually.

5.4 Conclusion

The aim of this chapter was to uncover the persistence conditions of psychological subjects. We began by analysing the classical Lockean account and its modern neo-Lockean successors. According to these views, persons persist diachronically by virtue of being psychologically continuous with their former selves. Psychological continuity between two points in time is established through the possession of a sufficient number of standing mental states like intentions, beliefs, or attitudes. Hence, personal identity is taken to consist in the retention of long-term memories.

To determine whether psychological continuity is a convincing basis of our transtemporal existence, we considered two pathological conditions in which this relation is partly disrupted, namely, retrograde amnesia and Alzheimer's disease. In severe forms of amnesia, only a small fraction of standing mental states is retained. It therefore seemed as if the criterion of personal identity that Lockean views endorse dictates that these patients must have ceased to exist. This would be a highly implausible entailment.

We established that advocates of memory-based accounts would likely have to concede that the low *number* of psychological connections that amnesia spares indeed suggests this conclusion, but that they could maintain that different *types* of these connections carry different weights in a person's persistence. One reason for granting certain types of connections a privileged status is their uniqueness. We invoked the distinction between episodic, semantic, and procedural memory to test this possible reply. Since semantic memories are usually partly spared in amnesia and procedural memories remain even entirely unaffected, some amnesiacs, notably those who possess specialist factual knowledge or exceptional motor skills, will therefore still be endowed with a very unique set of mental states despite having lost a large portion of their episodic memories. Thus, although it is normally this latter type of memory that is considered to be most intimately linked to personal identity as it permits an individual to create an autobiographical narrative, Lockeans might escape objections based on amnesia by assigning greater weight to certain unique psychological connections of the semantic and procedural types.

Alzheimer's was the second memory-affecting condition that we analysed. In the final stages of this disease, the loss of psychological connections is even more ex-

419 Edlow et al. 2013; Moll et al. 2009.

tensive than in the case of amnesia. Patients suffering from this condition experience a rapid decline not only in episodic and semantic, but also in procedural memory. The overwhelming majority of standing mental states eventually falls pray to the disease. Nonetheless, however, there is a period during which consciousness in the basic sense of being awake and aware is retained. That a subject who has perceptions and experiences emotions should have ceased to exist as memory-based views of personal identity inevitably imply is very implausible. If, therefore, demented individuals survive the loss of their long-term memories, the preservation of the latter cannot be a necessary condition of our diachronic persistence.

To ensure that this conclusion is accurate, we investigated whether there is also a real-life condition that exemplifies the opposite permutation, that is, a condition in which long-term memories are preserved while all other mental capacities are irreversibly absent. We found that this constellation occurs in the relatively rare case of an isolated lesion in the ascending reticular activating system. The ARAS is responsible for the generation of arousal, which is why patients who have sustained damage to the brainstem area from which this system originates cannot awake and remain forever comatose. Conversely, the neural correlates of standing mental states, and hence psychological continuity, remain unaffected by this kind of damage as their persistence is conditional only on the *cerebrum* being oxygenated and supplied with glucose. ARAS defects therefore establish a division between the capacity for consciousness and long-term memory, in which the former is extinguished whereas the latter is retained. Hence, the Lockean condition of diachronic existence continues to be fulfilled although the person, who is never going to awake again, must have ceased to exist.

We considered three objections that proponents of memory-based accounts might raise to this conclusion. They might insist that the ARAS is a constitutive part of the realisation base of a mental state, so that they are under no obligation to recognise whatever physical traces remain in the oxygenated cerebral tissues after the irreversible loss of wakefulness as proper neural correlates of mental states. Psychological continuity would be disrupted and the Lockean criterion of diachronic persistence would no longer apply. We replied that if everyday phenomenona like dreamless sleep and transient coma are not to present unsurmountable obstacles to their account, Lockeans must regard as sufficient the possession of standing mental states and cannot demand that occurrent states be retained, too. The former, however, persist independently of cortical activation and thus of ARAS activity. As dreamless sleep, transient coma, and permanent ARAS defects are therefore on a par in this regard, defenders of memory-based views must recognise the specific unconscious microstructural configurations of cerebral tissue as proper correlates of standing mental states either in all three cases or in none of them.

Advocates of memory-based accounts may accept that a lesion in the ARAS does not extinguish the neural correlates of mental states, but instead maintain that it changes the states' content. As psychological continuity only obtains between qualit-

atively identical mental states, this relation would then be disrupted. According to Eysenck's influential model of personality, dissimilarities in cortical arousal correlate with behavioural differences, which, if true, would mean that certain character traits are as much a consequence of ARAS activity as they are of microstructural dispositions in the cerebrum. Against this objection we contended that even if Eysenck's model were physiologically accurate, the charge would only pertain to a subcategory of the many types of mental states that Lockeans generally acknowledge. In all other cases, the influence that the ARAS exerts on mental states is confined to a global regulation of their accessibility to consciousness by generating wakefulness. The respective content of the states remains unchanged.

Lastly, proponents of memory-based views could assert that a coma resulting from the destruction of the ARAS is not an irreversible condition since the relevant brain structures could at a future time be repaired or replaced. We put forward a technical and a metaphysical objection to this supposition. The ARAS is a highly complex network that provides widespread innervation to various areas of the cerebrum. Clinical studies showed some success in stimulating a partly defective ARAS, but only when the principal neural structures were largely intact. That functional prostheses will ever be available seems therefore improbable. Moreover, it is metaphysically unclear whether an individual who awoke with such a prosthesis would still be the same subject, given that one of the two elements of the capacity for consciousness would be radically modified. It is therefore reasonable to assume that following the destruction of the ARAS, the capacity for consciousness is not contingently but necessarily absent.

At least two of the three conditions that we considered in this chapter – Alzheimer's disease and a dysfunctional ARAS – demonstrated that the possession of long-term memory cannot be wherein our diachronic persistence consists. Given that we had rejected most other views of personal identity and their respective criteria in the previous chapters, the sole remaining candidate was consciousness itself. We began to explore whether consciousness could be the essence of our existence by reflecting on an obvious problem: unlike standing mental states, the stream of consciousness disappears whenever we are asleep, anaesthetised, or become unconscious for other reasons. If consciousness is wherein our our persistence consists, it must be attached to, or exist in conjunction with, an entity that does not vanish during these intervals. The sole promising contender for fulfilling this role is the brain.

By determining when our persistence conditions diverge from those of our brains, we then sought to establish what relation we bear to the former. Considering all possibly relevant events in the sequence of processes that ultimately lead to the destruction of the brain, we found that our existence usually ends later than the termination of the stream of consciousness but earlier than the physical disintegration of the brain. We concluded that the event at which our persistence conditions and those of our brains diverge can only be the irreversible loss of the *capacity* for consciousness, which occurs as the result of microstructural transformations in the cerebral tissues,

induced by anoxia that persists for longer than approximately three minutes. The devastation of the cerebral tissues progresses inversely proportional to the possession of the capacity for consciousness until, finally, the material prerequisites for consciousness and, hence, the physical basis of our existence as psychological subjects have entirely dissipated.

Although giving a detailed description of the relation that the capacity for consciousness bears to the brain will not be feasible until one manages to establish how exactly neural tissue brings forth mental phenomena, we tentatively speculated about how the most popular ontological positions would conceptualise this relation. For purely practical reasons, we chose the traditional Aristotelian framework over four-dimensionalism and the constitution view. Analogous to the relation that a statue's form bears to the bronze of which it is made, one may understand the capacity for consciousness as a mode of the substance brain. The substance can continue to persist without this accident, as it does in an irreversible coma, but the accident cannot exist in the absence of its underlying substrate. The ontological dependencies are consequently asymmetrical. If this is correct, each of us is a psychological subject whose diachronic existence consists in the persistence of the capacity for consciousness realised in the microstructure of the brain of which this capacity is a mode.

In certain situations, the brain is subject to drastic changes of its macrostructure, which prompted us to investigate how much of its constituent matter must be retained for the capacity for consciousness to persist and the psychological subject to continue to exist. We began by analysing the results of commissurotomies, but found the available evidence to be inconclusive. Hemispherectomies and the Wada test delivered a clearer picture that licensed the conclusion that a single hemisphere in conjunction with a functional brainstem is sufficient for retaining the capacity for consciousness.

Now we were finally able to specify what modifications must occur in a brain for a psychological subject to cease to be. Since we exist as long as our brains retain the capacity for generating both wakefulness and awareness, the microstructural integrity of the brainstem, which ensures the former, and of at least one hemisphere, which underlies the latter, are required for us to persist. Consequently, the irreversible loss of only one of these functions is a sufficient condition for our existence to come to an end. The status of long-term memories or of the organism as a whole is not decisive.[420]

420 Modified parts of this chapter appeared in the *Journal of Medical Ethics* under the title *Are the Irreversibly Comatose Still Here? The Destruction of Brains and the Persistence of Persons* (Meier 2020a).

He who lives more lives than one, more deaths than one must die.

(Oscar Wilde)

6 Conclusion

Let us now summarise our findings. We began with a methodological observation. Most philosophers who work on questions of personal identity rely on thought experiments to test their hypotheses and to decide between rivalling views. Often, however, one and the same hypothetical situation elicits contradictory intuitions instead of resolving the conflict and settling the debate. Asking why this is so, we found that since establishing our persistence conditions requires test scenarios in which physical and psychological continuities can be teased apart, the imagined situations are usually much more distant from the actual world than is the case in most other areas of philosophy. And the more fantastical the thought experiment, the less likely it becomes that it conforms to the standards of proper experimental design that guide ordinary physical experimentation. We reviewed influential examples from the literature, in which we detected problems with objectivity, reliability, and validity. In addition to these methodological shortcomings, many thought experiments that have been employed in the debate about personal identity also make unjustified assumptions about physiological details of the human body. This prompted us to try a novel, predominantly empirical approach.

Fifty years have passed since brain death was first implemented as a criterion of death. Its advocates believe that with the destruction of the brain, integrated functioning among the different organs ceases irreversibly, somatic unity dissolves, and the organism turns into a corpse. To test whether this assertion is correct, we compared the physiological profile of brain-dead bodies to that of other pathological conditions in which integrated functioning is also diminished to a certain degree. Each of these conditions exemplified a different aspect of the functional deficits that occur in brain death. This enabled us to show that, given adequate external support, all vital functions can continue in a brain-dead body even if all means by which the brain could exercise control, neural and endocrine, are lost. The destruction of the brain does therefore not disband somatic unity.

By introducing a classification of four different ways in which vital functions can cease to be performed, we then took these physiological considerations to a more abstract level. We demonstrated that it is highly problematical to base one's judgment as to whether a biological entity is dead or alive on the status of the neurological control mechanism of its functions, as the brain-death criterion does, rather than on the execution of the tasks themselves. The growing sophistication of external life support gave rise to a dangerous decoupling of the actual performance of a vital function from the retention of neurological control over it. We therefore concluded that half a cen-

tury after its introduction, the neurological criterion is facing the same fate as its cardiopulmonary predecessor: technological progress has made it obsolete.

When ventilators and heart transplantations became available in the 1960s, one could still solve the problem by changing the criteria. The locus of death was shifted from cardiopulmonary function to neurological function, and a different organ was endowed with the task of indicating our death. Given how sophisticated intensive-care medicine has become five decades later, and since there is no other organ that would be a promising candidate, one cannot resort to this solution a second time, however.

Being left without any criterion of death that is applicable in intensive-care settings leads to major practical problems, of which the most pressing concerns organ donation: if brain-dead bodies are still living organisms, explanting organs from these patients violates the dead-donor rule. Abandoning this rule, as some authors suggested,[421] would presumably even aggravate the difficulties since the admission that patients are in fact still alive when the explantation begins is likely to deter prospective donors and to cause a further decrease in the number of available organs. One needs a different way out of this predicament.

Rather than trying to adapt the *criterion* of death yet again, the solution that I have proposed in this thesis is changing its underlying *definition*. The dead-donor rule refers to whatever kind of entity we essentially are. Currently, it is taken for granted that this must be the organism. But are we really essentially organisms? If it turned out that our persistence conditions and those of our organisms are not congruent, and that the destruction of our brains end our existence despite not being tantamount to the ceasing to exist ouf our organisms, explanting organs from brain-dead donors would not be in breach of the dead-donor rule. This is where questions of personal identity became relevant.

Of the three main views of personal identity – immaterial substance accounts, biological accounts, and psychological accounts – we excluded the first one from our investigation. We attempted to decide between the remaining two views by establishing whether a brain could retain the attributes that psychological accounts of personal identity regard as the essence of our persistence in isolation from its original organism. To this end, we identified the major channels through which the brain and the rest of the body communicate and considered various pathological conditions and medical procedures in which certain aspects of this exchange of electrical signals and chemical messengers is disturbed. Thus, we gained an understanding of the effects that the disruption of a specific neural pathway or endocrine axis exerts on the mental capacities of the brain. We also identified the major hurdles to vascularly supplying a detached head and considered two ways in which the latter could be sustained in the absence of the original organism – connected to the circulation of another human body or attached to a sophisticated mechanical device – of which we only deemed the first to be

421 Miller / Truog 2016, 113–152; Jox 2014; Sade / Boan 2014; Collins 2010; Truog / Miller 2008. For criticism of this proposal, see Deutscher Ethikrat 2015, 104–113 and Bernat 2013.

close enough to medical reality for bearing sufficient weight in a philosophical argument. We concluded that an individual's mental capacities could indeed persist in the absence of his or her original organism and that if this is so, it is overwhelmingly plausible that we are essentially psychological subjects.

We then proceeded to establish the persistence conditions of these entities. The most influential type of psychological view is the memory-based account that John Locke pioneered. According to this view, we persist diachronically by virtue of being psychologically continuous with our former selves. Psychological continuity is established through the possession of a sufficient number of standing mental states that are carried forward through time.

To determine whether psychological continuity is convincing as the basis of our transtemporal existence, we analysed two pathological conditions – amnesia and Alzheimer's disease – in which this relation is partly disrupted since only a small fraction of standing mental states is retained. While advocates of memory-based accounts could argue that not only the sheer number but also the types of mental states that the subject still possesses are decisive for her persistence, we found this reply to be unsatisfactory at least in the case of Alzheimer's. Memory-based views inevitably entail that individuals in the final stages of this disease have ceased to exist. For a psychological account of personal identity, this is a highly implausible result.

We also investigated the opposite permutation, that is, a condition in which long-term memories are preserved while all other mental capacities are irreversibly absent. This constellation occurs when an isolated lesion disables the brainstem's ascending reticular activating system while the cerebrum stays intact. Since this system is responsible for the generation of wakefulness, which is a prerequisite of being conscious, subjects in whom the ARAS is dysfunctional are going to be forever comatose. However, the persistence of the neural correlates of a subject's standing mental states, and hence of psychological continuity, remains unaffected by this type of brain damage. The Lockean condition of diachronic existence is consequently still fulfilled although it is obvious that the person must have ceased to exist.

Having thus found both a real-life condition in which the disruption of psychological continuity *precedes* our actual ceasing to exist and one in which the retention of psychological continuity *succeeds* our de facto going out of existence, we concluded that memory-based views are incompatible with neurophysiological facts. Consequently, we turned to their consciousness-based rivals. Deeming consciousness the essence of our existence entails the problem that, unlike standing mental states, the stream of consciousness disappears periodically – even in non-pathological situations. If consciousness is indeed wherein our persistence consists, it must therefore be attached to, or exist in conjunction with, an entity that persists during these intervals. According to modern understanding, this entity can only be the brain. We tried to specify what kind of relation we might bear to our brains, and proposed that one may understand the capacity for generating consciousness as a mode of this substance. If this

is accurate, each of us is a psychological subject whose diachronic existence consists in the persistence of the capacity for consciousness, realised in the brain of which this capacity is a mode.

Since the brain is sometimes subject to drastic changes to its macrostructure, we analysed the results of neurosurgical operations to determine how much of this substance's constituent matter must be retained for the capacity for consciousness to persist and the psychological subject to continue to exist. We found that a single hemisphere in conjunction with a functional brainstem is sufficient for retaining the capacity for consciousness, presumably on the proviso that there is still a source of incoming sensory information.

Now we were finally in a position to identify the physiological discontinuity that indicates when a psychological subject has ceased to exist. We were thus changing the level of the investigation once again, moving from the *definition* of death back to its *criteria*, only that this time these criteria were not meant to indicate the functional status of organisms but the persistence or the absence of psychological entities. Since wakefulness and awareness are required for the retention of the capacity for consciousness, both the brainstem and at least one cerebral hemisphere must remain intact for us to persist. Consequently, the destruction of either of these brain areas is a sufficient condition for our existence to come to an end.

If this is correct, brain death is indeed our death – but for reasons entirely different from those cited in its official justification. The destruction of the brain marks the moment of our ceasing to exist because we are essentially psychological subjects whose persistence is dependent on the microstructural integrity of this material substrate. But the destruction of the brain is not the death of our organisms. Organisms and psychological subjects have different persistence conditions, which is why the death of an organism and the ceasing to be of a psychological subject need not coincide. It is therefore only logical also to demand two different criteria – one for the ceasing to exist of each entity.[422]

As the irreversible loss of function in the brain is not the moment at which the organism dies but the point at which the death of the organism would no longer *entail* a psychological subject's ceasing to be, this revised definition of death enables the explantation of organs from brain-dead patients while simultaneously preserving the dead-donor rule – on the assumption, of course, that it is permissible to withdraw life support from a severely ill organism in order to save one or more psychological subjects.[423] Thus, the definition solves the most pressing practical problem that resulted from the finding that brain death is not the cessation of integrated organismic functioning.

422 See also Lizza 2018, 3; Savulescu 2003, 129; McMahan 1995, 102 f. Culver and Gert (1982, 182 f.) disagree.

423 Whether this is permissible is not a metaphysical but a moral question that is beyond the scope of this thesis.

That our revised definition of death facilitates the practice of organ donation is, although highly desirable, only a welcome by-product of our strictly metaphysical considerations. As stated in the introduction, the prospect of practical advantages, no matter how great, must not contaminate the investigation into our persistence conditions. We therefore conducted our study without regard to the question of organ donation or any other clinical or social benefits that the respective solution might bring.

It has frequently been objected that, given the ever-increasing need for donor organs, adopting a psychological basis for the criterion of death would soon pave the way for a slippery slope.[424] Severely cognitively impaired individuals, so the argument goes, would become endangered by an eroding psychological definition of death that has no clearly delineated boundaries. David Lamb even speaks of 'euthanasia in the persistent vegetative state'.[425]

It seems to me that we are already entering a situation in which the exact opposite is the case. In the age of constantly improving effectiveness in the external provision of organismic functions, not psychological but biological definitions of death have become prone to yield vague criteria. The better life-support systems are at replacing bodily functions, the less clear it is whether the entity in the hospital bed is still a living organism or rather a machine comprising organic parts. We have reached a point at which technological progress has made obsolete not only the traditional cardiopulmonary criterion of organismic death but also its neurological successor. If this development continues, doctors will be forced to draw a line at a certain level of mechanical support. This, if anything, deserves the charge of leading to a slippery slope. Adopting a psychological definition of death, thereby transferring the determination of our ceasing to exist to the only realm in which any attempts at artificial substitution have been utterly unsuccessful, will therefore not introduce a slippery slope but rather help to avoid it. Our constant defeat in replacing the brain's mental capacities, paradoxical as it may seem, is the prerequisite of the irredeemable irreversibility that any definition of death must have at its core. The better machines become at doing what our organisms do, the less can our definition of death rely on these functions – and the more should it focus on what they cannot substitute: consciousness.

424 See, for example, Bernat 2001, 177; Pallis / Harley 1996, 3; Culver / Gert 1982, 183.
425 Lamb 1985, 112.

7 References

Academy of Medical Royal Colleges. 2008. *A Code of Practice for the Diagnosis and Confirmation of Death*. London: Academy of Medical Royal Colleges.

Ad Hoc Committee of the Harvard Medical School to Examine the Definition of Brain Death. 1968. 'A Definition of Irreversible Coma'. *Journal of the American Medical Association* 205 (6): 85–88.

Adrianov, O.S., I.N. Bogolepova, S.M. Blinkov, and L.A. Kukuev. 1993. 'The Study of V. I. Lenin's Brain [in Russian]'. *Uspekhi Fiziologicheskikh Nauk* 24 (3): 40–52.

Aggarwal, A., P. Aggarwal, M. Khatak, and S. Khatak. 2010. 'Cerebral Ischemic Stroke: Sequels of Cascade'. *International Journal of Pharma and Bio Sciences* 1 (3): 1–24.

Ali, S.A., T. Begum, and F. Reza. 2018. 'Hormonal Influences on Cognitive Function'. *Malaysian Journal of Medical Sciences* 25 (4): 31–41.

Alzheimer's Society in the United Kingdom. 2012. *The Later Stages of Dementia*. London: Alzheimer's Society.

Ardiel, E.L., and C.H. Rankin. 2010. 'The Importance of Touch in Development'. *Journal of Paediatrics and Child Health* 15 (3): 153–156.

Aristotle. 1995. *The Complete Works of Aristotle*. Edited by Barnes, J. Princeton: Princeton University Press.

Baker, L.R. 1999. 'What Am I?' *Philosophy and Phenomenological Research* 59 (1): 151–159.

———. 2000. *Persons and Bodies: A Constitution View*. Cambridge: Cambridge University Press.

———. 2007. 'Persons and Other Things'. *Journal of Consciousness Studies* 14 (5–6): 17–36.

Barker, J.H., J.M. Frank, and L. Leppik. 2015. 'Head Transplantation: Editorial Commentary'. *CNS Neuroscience & Therapeutics* 21 (8): 613–614.

Bartlett, G. 2018a. 'Occurrent States'. *Canadian Journal of Philosophy* 48 (1): 1–17.

———. 2018b. 'Functionalism and the Problem of Occurrent States'. *The Philosophical Quarterly* 68 (270): 1–20.

Bauer, G., F. Gerstenbrand, and E. Rumpl. 1979. 'Varieties of the Locked-in Syndrome'. *Journal of Neurology* 221: 77–91.

Bayne, T. 2008. 'The Unity of Consciousness and the Split-Brain Syndrome'. *The Journal of Philosophy* 105 (6): 277–300.

Beck, S. 2014. 'Transplant Thought-Experiments: Two Costly Mistakes in Discounting Them'. *South African Journal of Philosophy* 33 (2): 189–199.

Becker, L.C. 1975. 'Human Being: The Boundaries of the Concept'. *Philosophy & Public Affairs* 4 (4): 334–359.

Bernal, J. 2015. 'Thyroid Hormones in Brain Development and Function'. In *Endotext*, edited by Feingold, K.R., B. Anawalt, A. Boyce, G. Chrousos, K. Dungan, A. Grossman, J.M. Hershman, et al. South Dartmouth: MDText.com.

Bernat, J.L. 1998. 'A Defense of the Whole-Brain Concept of Death'. *The Hastings Center Report* 28 (2): 14–23.

———. 2001. 'Philosophical and Ethical Aspects of Brain Death'. In *Brain Death*, edited by Wijdicks, E.F.M., 171–187. Philadelphia: Lippincott Williams & Wilkins.

———. 2010. 'How the Distinction between "Irreversible" and "Permanent" Illuminates Circulatory-Respiratory Death Determination'. *Journal of Medicine and Philosophy* 35 (3): 242–255.

———. 2013. ‘Life or Death for the Dead-Donor Rule?’ *New England Journal of Medicine* 369 (14): 1289–1291.
———. 2018. ‘A Conceptual Justification for Brain Death’. *Hastings Center Report* 48 (6): S19–S21.
Bernat, J.L., C.M. Culver, and B. Gert. 1981. ‘On the Definition and Criterion of Death’. *Annals of Internal Medicine* 94 (3): 389–394.
Bertram, G.W. 2012. ‘Fiktionale Reflexionen von Begriffen: Zur Theorie philosophischer Gedankenexperimente’. In *Philosophische Gedankenexperimente: Ein Lese- und Studienbuch*, edited by Bertram, G.W., 24–72. Stuttgart: Reclam.
Birnbacher, D. 2017. *Tod*. Boston: de Gruyter.
Blatti, S. 2012. ‘A New Argument for Animalism’. *Analysis* 72 (4): 685–690.
———. 2019. ‘Animalism’. *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University.
Blatti, S., and P.F. Snowdon, eds. 2016. *Animalism: New Essays on Persons, Animals, and Identity*. Oxford: Oxford University Press.
Block, N.J. 1978. ‘Troubles with Functionalism’. *Minnesota Studies in the Philosophy of Science* 9: 261–325.
———. 1995. ‘On a Confusion about a Function of Consciousness’. *Behavioral and Brain Sciences* 18 (2): 227–287.
Block, N.J., and J.A. Fodor. 1972. ‘What Psychological States Are Not’. *The Philosophical Review* 81 (2): 159–181.
Braddon-Mitchell, D., and F. Jackson. 2007. *The Philosophy of Mind and Cognition*. 2nd ed. Malden: Blackwell.
Brett, M. 2001. ‘Motherhood and Memory: A Review’. *Psychoneuroendocrinology* 26 (4): 339–362.
Broad, C.D. 1945. ‘Some Reflections on Moral-Sense Theories in Ethics’. *Proceedings of the Aristotelian Society* 45 (1): 131–166.
Brukhonenko, S., and S. Tchetchuline. 1929. ‘Expériences avec la tête isolée du chien’. *Journal de Physiologie et de Pathologie Générale* 27: 31–45.
Buckwalter, W., D. Rose, and J. Turri. 2015. ‘Belief through Thick and Thin’. *Noûs* 49 (4): 748–775.
Bundesärztekammer. 2015. *Richtlinie für die Regeln zur Feststellung des Todes und die Verfahrensregeln zur Feststellung des endgültigen, nicht behebbaren Ausfalls der Gesamtfunktion des Großhirns, des Kleinhirns und des Hirnstamms*. Berlin: Bundesärztekammer.
Butler, J. 1927. *The Analogy of Religion*. London: Dent & Sons.
Campbell, S. 2006. ‘The Conception of a Person as a Series of Mental Events’. *Philosophy and Phenomenological Research* 73 (2): 339–358.
Campbell, T., and J. McMahan. 2010. ‘Animalism and the Varieties of Conjoined Twinning’. *Theoretical Medicine and Bioethics* 31 (4): 285–301.
Canavero, S. 2013. ‘HEAVEN: The Head Anastomosis Venture Project Outline for the First Human Head Transplantation with Spinal Linkage (GEMINI)’. *Surgical Neurology International* 4 (2): S335–S342.
Carruthers, P. 1995. *Introducing Persons: Theories and Arguments in the Philosophy of Mind*. London: Routledge.
Carter, W.R. 1988. ‘Our Bodies, Our Selves’. *Australasian Journal of Philosophy* 66 (3): 308–319.
Chalmers, D. 1995. ‘Facing up to the Problem of Consciousness’. *Journal of Consciousness Studies* 2 (3): 200–219.

———. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press.
Chau, K.H., T. Friedman, M. Tranquilli, and J.A. Elefteriades. 2013. ‘Deep Hypothermic Circulatory Arrest Effectively Preserves Neurocognitive Function’. *The Annals of Thoracic Surgery* 96 (5): 1553–1559.
Clark, C.G., J.G. Murray, I.M. Slessor, and J.H. Wyllie. 1964. ‘Complete Vagotomy and Its Consequences: Follow-up of 146 Patients’. *British Medical Journal* 2 (5414): 900–903.
Coenen, A.M.L. 2007. ‘Consciousness without a Cortex, but What Kind of Consciousness Is This?’ *Behavioral and Brain Sciences* 30 (1): 87–88.
Cohadon, F., and E. Richer. 1993. ‘Stimulation cérébrale profonde chez des patients en état végétatif post-traumatique: 25 observations’. *Neurochirurgie* 39 (5): 281–292.
Cohnitz, D. 2006. *Gedankenexperimente in der Philosophie*. Paderborn: Mentis.
Cole, D. 2020. ‘The Chinese Room Argument’. *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University.
Collerton, D., and E. Perry. 2007. ‘Do Multiple Cortical–Subcortical Interactions Support Different Aspects of Consciousness?’ *Behavioral and Brain Sciences* 30 (1): 88–89.
Collins, M. 2010. ‘Reevaluating the Dead Donor Rule’. *Journal of Medicine and Philosophy* 35 (2): 154–179.
Corazzol, M., G. Lio, A. Lefevre, G. Deiana, L. Tell, N. André-Obadia, P. Bourdillon, et al. 2017. ‘Restoring Consciousness with Vagus Nerve Stimulation’. *Current Biology* 27 (18): R994–R996.
Corcoran, C.D., P. Thomas, J. Phillips, and V. O’Keane. 2006. ‘Vagus Nerve Stimulation in Chronic Treatment-Resistant Depression’. *British Journal of Psychiatry* 189 (3): 282–283.
Crane, T. 2001. *Elements of Mind: An Introduction to the Philosophy of Mind*. Oxford: Oxford University Press.
Crossman, A.R., and D. Neary. 2015. *Neuroanatomy: An Illustrated Colour Text*. 5th ed. Edinburgh: Elsevier.
Culver, M.C., and B. Gert. 1982. *Philosophy in Medicine: Conceptual and Ethical Issues in Medicine and Psychiatry*. New York: Oxford University Press.
Cypel, M., T. Waddell, L.G. Singer, L. del Sorbo, E. Fan, M. Binnie, N.D. Ferguson, and S. Keshavjee. 2017. ‘Bilateral Pneumonectomy to Treat Uncontrolled Sepsis in a Patient Awaiting Lung Transplantation’. *The Journal of Thoracic and Cardiovascular Surgery* 153 (4): e67–e69.
Damasio, A.R. 1994. *Descartes’ Error: Emotion, Reason, and the Human Brain*. New York: Avon.
———. 2010. *Self Comes to Mind: Constructing the Conscious Brain*. New York: Pantheon.
Davidson, D. 1987. ‘Knowing One’s Own Mind’. *Proceedings and Addresses of the American Philosophical Association* 60 (3): 441–458.
DeGrazia, D. 2005. *Human Identity and Bioethics*. Cambridge: Cambridge University Press.
Deliyannakis, E., F. Ioannou, and A. Davaroukas. 1975. ‘Brain Stem Death with Persistence of Bioelectric Activity of the Cerebral Hemispheres’. *Clinical Electroencephalography* 6 (2): 75–79.

Dempsey, E.W., and R.S. Morison. 1941. ‘The Production of Rhythmically Recurrent Cortical Potentials after Localized Thalamic Stimulation’. *American Journal of Physiology* 135 (2): 293–300.
Dennett, D.C. 1998. ‘Where Am I?’ In *Brainstorms: Philosophical Essays on Mind and Psychology*, 310–323. Cambridge, MA: MIT Press.
Descartes, R. 2008. *Meditations on First Philosophy: With Selections from the Objections and Replies*. Translated by Moriarty, M. Oxford: Oxford University Press.
Deutscher Ethikrat. 2015. *Hirntod und Entscheidung zur Organspende: Stellungnahme*. Berlin: Deutscher Ethikrat.
Doesburg, S.M., and M.W. Lawrence. 2007. ‘Corticothalamic Necessity, Qualia, and Consciousness’. *Behavioral and Brain Sciences* 30 (1): 90–91.
Duhem, P. 1906. *La théorie physique: Son objet et sa structure*. Paris: Chevalier & Rivière.
Edlow, B.L., R.L. Haynes, E. Takahashi, J.P. Klein, P. Cummings, T. Benner, D.M. Greer, et al. 2013. ‘Disconnection of the Ascending Arousal System in Traumatic Coma’. *Journal of Neuropathology & Experimental Neurology* 72 (6): 505–523.
Edlow, B.L., E. Takahashi, O. Wu, T. Benner, G. Dai, L. Bu, P.E. Grant, et al. 2012. ‘Neuroanatomic Connectivity of the Human Ascending Arousal System Critical to Consciousness and Its Disorders’. *Journal of Neuropathology & Experimental Neurology* 71 (6): 531–546.
Ellis, M. 1999. ‘Life without Memory: The Case of Clive Wearing’. *The Mind. Second Edition [Video]*. Colorado State University.
Emery, S.F., and K.M. Robertson. 2001. ‘Organ Procurement and Preparation for Transplantation’. In *Brain Death*, edited by Wijdicks, E.F.M., 189–213. Philadelphia: Lippincott Williams & Wilkins.
Evans, B.M. 2003. ‘Sleep, Consciousness and the Spontaneous and Evoked Electrical Activity of the Brain: Is There a Cortical Integrating Mechanism?’ *Neurophysiologie Clinique* 33 (1): 1–10.
Eysenck, H.J. 2006. *The Biological Basis of Personality*. New Brunswick: Transaction.
Farkas, K. 2008. *The Subject’s Point of View*. Oxford: Oxford University Press.
FitzGerald, M.J.T., G. Gruener, and E. Mtui. 2012. *Clinical Neuroanatomy and Neuroscience*. 6th ed. Edinburgh: Elsevier.
Ford, L.C., and R.M. Cruz. 2004. ‘Bilateral Glossopharyngeal Nerve Paralysis after Tonsillectomy: Case Report and Anatomic Study’. *The Laryngoscope* 114 (12): 2196–2199.
Freud, S. 2016. *Das Unbewusste*. Edited by Bayer, L. and H.-M. Lohmann. Stuttgart: Reclam.
Frings, M., and R.J. Jox. 2015. *Gehirn und Moral: Ethische Fragen in Neurologie und Hirnforschung*. Stuttgart: Thieme.
Fuchs, T. 2017a. *Das Gehirn – ein Beziehungsorgan: Eine phänomenologisch-ökologische Konzeption*. 5th ed. Stuttgart: Kohlhammer.
———. 2017b. ‘Self across Time: The Diachronic Unity of Bodily Existence’. *Phenomenology and the Cognitive Sciences* 16 (2): 291–315.
———. 2018. *Ecology of the Brain: The Phenomenology and Biology of the Embodied Mind*. Oxford: Oxford University Press.
Gallie, I. 1936. ‘Is the Self a Substance?’ *Mind* 45 (177): 28–44.

Garrigues, H.J. 1889. *Der Scheintod: Vortrag vor dem deutschen gesellig-wissenschaftlichen Verein von New York*. Gemeinverständliche Vorträge 15. New York: Deutscher gesellig-wissenschaftlicher Verein von New York.
Gasbarri, A., M.C.H. Tavares, R.C. Rodrigues, C. Tomaz, and A. Pompili. 2012. 'Estrogen, Cognitive Functions and Emotion: An Overview on Humans, Non-Human Primates and Rodents in Reproductive Years'. *Reviews in the Neurosciences* 23 (5–6): 587–606.
Gendler, T.S. 2000. *Thought Experiment: On the Power and Limits of Imaginary Cases*. New York: Garland.
Geniole, S.N., B.M. Bird, J.S. McVittie, R.B. Purcell, J. Archer, and J.M. Carré. 2020. 'Is Testosterone Linked to Human Aggression? A Meta-Analytic Examination of the Relationship between Baseline, Dynamic, and Manipulated Testosterone on Human Aggression'. *Hormones and Behavior* 123 (104644): 1–11.
Gertler, B. 2007. 'Overextending the Mind?' In *Arguing about the Mind*, edited by Gertler, B. and L.A. Shapiro, 192–206. New York: Routledge.
Gervais, K.G. 1986. *Redefining Death*. New Haven: Yale University Press.
Ghaemmaghami, V., P. Gdalevitch, L. Gentilello, and H. Frankel. 2008. 'Vagal Nerve Transection after Gunshot Wound to Neck'. *The Journal of Trauma, Injury, Infection, and Critical Care* 64 (3): 846.
Giacino, J.T., D.I. Katz, N.D. Schiff, J. Whyte, E.J. Ashman, S. Ashwal, R. Barbano, et al. 2018. 'Comprehensive Systematic Review Update Summary: Disorders of Consciousness'. *Archives of Physical Medicine and Rehabilitation* 99 (9): 1710–1719.
Gibson, E.L. 2006. 'Emotional Influences on Food Choice: Sensory, Physiological and Psychological Pathways'. *Physiology & Behavior* 89 (1): 53–61.
Gillett, G. 1986. 'Brain Bisection and Personal Identity'. *Mind* 95 (378): 224–229.
Goldman, A.I. 1970. *A Theory of Human Action*. New Jersey: Prentice-Hall.
Gordon, J.K., and J. McKinlay. 2012. 'Physiological Changes after Brain Stem Death and Management of the Heart-Beating Donor'. *Continuing Education in Anaesthesia, Critical Care & Pain* 12 (5): 225–229.
Gramm, H.-J., H. Meinhold, U. Bickel, J. Zimmermann, B. von Hammerstein, F. Keller, R. Dennhardt, and K. Voigt. 1992. 'Acute Endocrine Failure after Brain Death?' *Transplantation* 54 (5): 851–857.
Gray, J.A. 1981. 'A Critique of Eysenck's Theory of Personality'. In *A Model for Personality*, edited by Eysenck, H.J., 246–276. Berlin: Springer.
Green, M.B., and D. Wikler. 1980. 'Brain Death and Personal Identity'. *Philosophy & Public Affairs* 9 (2): 105–133.
Grundy, D., and A. Swain. 1993. *ABC of Spinal Cord Injury*. 2nd ed. London: BMJ.
Hagemann, D., and E. Naumann. 2009. 'States vs. Traits: An Integrated Model for the Test of Eysenck's Arousal/Arousability Hypothesis'. *Journal of Individual Differences* 30 (2): 87–99.
Häggqvist, S. 1996. *Thought Experiments in Philosophy*. Acta Universitatis Stockholmiensis 18. Stockholm: Almqvist & Wiksell.
Han, S.-G., G.-M. Kim, K.H. Lee, C.-S. Chung, and K.-Y. Jung. 2006. 'Reflex Movements in Patients with Brain Death: A Prospective Study in a Tertiary Medical Center'. *Journal of Korean Medical Science* 21 (3): 588–590.
Hasker, W. 2004. 'The Constitution View of Persons'. *International Philosophical Quarterly* 44 (1): 23–34.

Hassler, R. 1971. ‘Regulation der psychischen Aktivität’. In *Hirnforschung und Psychiatrie: Stand und Zukunftsaspekte naturwissenschaftlicher Forschung in Deutschland*, edited by Kurzrock, R., 27–35. Berlin: Colloquium.
Hassler, R., G. Dalle Ore, G. Dieckmann, A. Bricolo, and G. Dolce. 1969. ‘Behavioural and EEG Arousal Induced by Stimulation of Unspecific Projection Systems in a Patient with Post-Traumatic Apallic Syndrome’. *Electroencephalography and Clinical Neurophysiology* 27 (3): 306–310.
Haugeland, J. 1998. *Having Thought: Essays in the Metaphysics of Mind*. Cambridge, MA: Harvard University Press.
Hawley, K. 2001. *How Things Persist*. Oxford: Oxford University Press.
———. 2014. ‘Persistence and Time’. In *The Cambridge Companion to Life and Death*, edited by Luper, S., 47–63. Cambridge: Cambridge University Press.
Hoffman, J.C. 1979. ‘Clarifying the Debate on Death’. *Soundings* 62 (4): 430–447.
Horne, M. 2009. ‘Are People in a Persistent Vegetative State Conscious?’ *Monash Bioethics Review* 28 (2): 12.1-12.12.
Hudson, H. 2007. ‘I Am Not an Animal!’ In *Persons: Human and Divine*, edited by Inwagen, P. van and D.W. Zimmerman, 216–234. Oxford: Clarendon Press.
Hume, D. 1896. *A Treatise of Human Nature*. Edited by Selby-Bigge, L.A. Oxford: Clarendon Press.
Hyde, D., and D. Raffman. 2018. ‘Sorites Paradox’. *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University.
Inwagen, P. van. 1995. *Material Beings*. Ithaca: Cornell University Press.
Jackson, F. 1982. ‘Epiphenomenal Qualia’. *The Philosophical Quarterly* 32 (127): 127–136.
James, W. 1890. *The Principles of Psychology*. Vol. 1. New York: Hold & Co.
Jang, S.H., and H.S. Kim. 2015. ‘Aneurysmal Subarachnoid Hemorrhage Causes Injury of the Ascending Reticular Activating System: Relation to Consciousness’. *American Journal of Neuroradiology* 36 (4): 667–671.
Jang, S.H., and H.G. Kwon. 2015. ‘The Direct Pathway from the Brainstem Reticular Formation to the Cerebral Cortex in the Ascending Reticular Activating System: A Diffusion Tensor Imaging Study’. *Neuroscience Letters* 606: 200–203.
Johnston, M. 1992. ‘Constitution Is Not Identity’. *Mind* 101 (401): 89–106.
———. 2016. ‘Remnant Persons: Animalism’s Undoing’. In *Animalism: New Essays on Persons, Animals, and Identity*, edited by Blatti, S. and P.F. Snowdon, 89–127. Oxford: Oxford University Press.
Jousset, N., J.P. Jacob, A. Gaudin, D. Mauillon, M. Penneau, and C. Rougé-Maillart. 2009. ‘Prélèvements d’organes sur donneurs décédés après arrêt cardiaque’. *La Presse Médicale* 38 (5): 740–744.
Jox, R.J. 2014. ‘Sketching the Alternative to Brain Death: Dying through Organ Donation’. *The American Journal of Bioethics* 14 (8): 37–39.
Karlsson, A.-K. 2006. ‘Autonomic Dysfunction in Spinal Cord Injury: Clinical Presentation of Symptoms and Signs’. In *Progress in Brain Research*, edited by Weaver, L.C. and C. Polosa, 1–8. Amsterdam: Elsevier.
Keller, M. 2011. ‘Tatort Klinik Madrid’. *GEO* 105: 104–114.
Koch, C. 2010. ‘Reviving Consciousness’. *Scientific American Mind* 20 (7): 20–21.
Kolak, D. 1993. ‘The Metaphysics and Metapsychology of Personal Identity: Why Thought Experiments Matter in Deciding Who We Are’. *American Philosophical Quarterly* 30 (1): 39–50.
Konstantinov, I.E. 2009. ‘At the Cutting Edge of the Impossible: A Tribute to Vladimir P. Demikhov’. *Texas Heart Institute Journal* 36 (5): 453–458.

Koo, K.C., G.S. Shim, H.H. Park, K.H. Rha, Y.D. Choi, B.H. Chung, S.J. Hong, and J.W. Lee. 2013. ‘Treatment Outcomes of Chemical Castration on Korean Sex Offenders’. *Journal of Forensic and Legal Medicine* 20 (6): 563–566.
Korein, J. 1978. ‘The Problem of Brain Death: Development and History’. *Annals of the New York Academy of Sciences* 315 (1): 19–38.
Koštál, V., P. Šimek, H. Zahradníčková, J. Cimlová, and T. Štětina. 2012. ‘Conversion of the Chill Susceptible Fruit Fly Larva (Drosophila Melanogaster) to a Freeze Tolerant Organism’. *Proceedings of the National Academy of Sciences* 109 (9): 3270–3274.
Kosteljanetz, M., J.K. Øhrstrøm, S. Skjødt, and P.S. Teglbjærg. 1988. ‘Clinical Brain Death with Preserved Cerebral Arterial Circulation’. *Acta Neurologica Scandinavica* 78 (5): 418–421.
Kühne, U. 2005. *Die Methode des Gedankenexperiments*. Frankfurt a. M.: Suhrkamp.
Kundu, B., J.D. Rolston, and R. Grandhi. 2019. ‘Mapping Language Dominance through the Lens of the Wada Test’. *Neurosurgical Focus* 47 (3): 1–7.
Kurthen, M., and D.B. Linke. 1995. ‘Vom Hirntod zum Teilhirntod’. In *Wann ist der Mensch tot? Organverpflanzung und ‘Hirntod’-Kriterium*, edited by Hoff, J. and J. in der Schmitten, 2nd ed., 82–94. Reinbek: Rowohlt.
Lagercrantz, H., and J.-P. Changeux. 2009. ‘The Emergence of Human Consciousness: From Fetal to Neonatal Life’. *Pediatric Research* 65 (3): 255–260.
Lamb, D. 1978. ‘Diagnosing Death’. *Philosophy & Public Affairs* 7 (2): 144–153.
———. 1985. *Death, Brain Death, and Ethics*. London: Croom Helm.
Lange, C. de. 2012. ‘The Man Who Saves Stephen Hawking’s Voice’. *New Scientist*, no. 2846.
Laureys, S. 2005. ‘The Neural Correlate of (Un)Awareness: Lessons from the Vegetative State’. *Trends in Cognitive Sciences* 9 (12): 556–559.
Law Reform Commission of Canada. 1981. *Report 15: Criteria for the Determination of Death*. Ottawa: Department of Justice Canada.
Le Gallois, C.J.-J. 1812. *Expériences sur le principe de la vie: Notamment sur celui des mouvemens du cœur, et sur le siège de ce principe*. Paris: d’Hautel.
Leibniz, G.W. 2014. *Leibniz’s Monadology: A New Translation and Guide*. Translated by Strickland, L. Edinburgh: Edinburgh University Press.
Levin, J. 2018. ‘Functionalism’. *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University.
Lewis, C.S. 2015. *Studies in Words*. 2nd ed. Cambridge: Cambridge University Press.
Lewis, D. 1976. ‘Survival and Identity’. In *The Identities of Persons*, edited by Rorty, A.O., 17–40. Berkeley: University of California Press.
Lienert, G.A., and U. Raatz. 1994. *Testaufbau und Testanalyse*. 5th ed. Weinheim: Psychologie Verlagsunion.
Liljegren, M., G. Naasan, J. Temlett, D.C. Perry, K.P. Rankin, J. Merrilees, L.T. Grinberg, W.W. Seeley, E. Englund, and B.L. Miller. 2015. ‘Criminal Behavior in Frontotemporal Dementia and Alzheimer Disease’. *JAMA Neurology* 72 (3): 295–300.
Lillehei, C.W. 1982. ‘A Personalized History of Extracorporeal Circulation’. *Transactions – American Society for Artificial Internal Organs* 28: 5–16.
Lilly, J.C. 1977. *The Deep Self: Profound Relaxation and the Tank Isolation Technique*. New York: Simon and Schuster.
Lilly, J.C., and J.T. Shurley. 1961. ‘Experiments in Solitude in Maximum Achievable Physical Isolation with Water Suspension of Intact Healthy Persons’. In

Psychophysiological Aspects of Space Flight, 238–247. New York: Columbia University Press.
Lizza, J.P. 2004. ‘The Conceptual Basis for Brain Death Revisited: Loss of Organic Integration or Loss of Consciousness?’ In *Brain Death and Disorders of Consciousness*, edited by Machado, C. and D.A. Shewmon, 51–59. Boston: Springer.
———. 2011. ‘Where’s Waldo? The “Decapitation Gambit” and the Definition of Death’. *Journal of Medical Ethics* 37 (12): 743–746.
———. 2018. ‘Defining Death: Beyond Biology’. *Diametros* 55: 1–19.
Lledo, P.-M., M. Alonso, and M.S. Grubb. 2006. ‘Adult Neurogenesis and Functional Plasticity in Neuronal Circuits’. *Nature Reviews Neuroscience* 7 (3): 179–193.
Lleó, A., and R. Blesa. 2017. ‘Clinical Course of Alzheimer’s Disease’. In *Alzheimer’s Disease*, edited by Waldemar, G. and A.S. Burns, 2nd ed., 27–33. Oxford: Oxford University Press.
Lo Mauro, A., and A. Aliverti. 2016. ‘Physiology of Respiratory Disturbances in Muscular Dystrophies’. *Breathe* 12 (4): 318–327.
Locke, J. 2008. *An Essay Concerning Human Understanding*. Edited by Phemister, P. Oxford: Oxford University Press.
Lorusso, R., S. Gelsomino, O. Parise, M. Di Mauro, F. Barili, G. Geskes, E. Vizzardi, et al. 2017. ‘Neurologic Injury in Adults Supported with Veno-Venous Extracorporeal Membrane Oxygenation for Respiratory Failure: Findings from the Extracorporeal Life Support Organization Database’. *Critical Care Medicine* 45 (8): 1389–1397.
Lowe, E.J. 1995. *Locke on Human Understanding*. London: Routledge.
———. 2006. ‘Non-Cartesian Substance Dualism and the Problem of Mental Causation’. *Erkenntnis* 65 (1): 5–23.
Lucretius Carus, T. 2014. *On the Nature of Things*. Translated by Rouse, W.H.D. Cambridge, MA: Harvard University Press.
Mach, E. 1926. *Erkenntnis und Irrtum: Skizzen zur Psychologie der Forschung*. 5th ed. Leipzig: Barth.
Machado, C. 1999. ‘Consciousness as a Definition of Death: Its Appeal and Complexity’. *Clinical Electroencephalography* 30 (4): 156–164.
Machado, S., M. Cunha, D. Minc, C.E. Portella, B. Velasques, L.F. Basile, M. Cagy, R. Piedade, and P. Ribeiro. 2009. ‘Alzheimer’s Disease and Implicit Memory’. *Arquivos de Neuro-Psiquiatria* 67 (2-A): 334–342.
Mackie, D. 1999. ‘Personal Identity and Dead People’. *Philosophical Studies* 95 (3): 219–242.
Madden, R. 2016. ‘Thinking Parts’. In *Animalism: New Essays on Persons, Animals, and Identity*, edited by Blatti, S. and P.F. Snowdon, 180–207. Oxford: Oxford University Press.
Mahmoud, R., S.R. Wainwright, and L.A.M. Galea. 2016. ‘Sex Hormones and Adult Hippocampal Neurogenesis: Regulation, Implications, and Potential Mechanisms’. *Frontiers in Neuroendocrinology* 41: 129–152.
Maiwall, R., J.S. Maras, S.L. Nayak, and S.K. Sarin. 2014. ‘Liver Dialysis in Acute-on-Chronic Liver Failure: Current and Future Perspectives’. *Hepatology International* 8 (S2): S505–S513.
Maki, P.M. 2005. ‘Estrogen Effects on the Hippocampus and Frontal Lobes’. *International Journal of Fertility and Women’s Medicine* 50 (2): 67–71.
Martin, M.B. 2015. ‘Truncal Vagotomy Without Drainage: Are There Long-Term Concerns?’ *Surgical Endoscopy* 29 (11): 3086–3089.

Martin, R. 1998. *Self-Concern: An Experiential Approach to What Matters in Survival*. Cambridge: Cambridge University Press.
Mauthner, L. 1890. ‘Zur Pathologie und Physiologie des Schlafes nebst Bemerkungen über die “Nona”’. *Wiener Medizinische Wochenschrift* 40 (23–25): 961–964; 1001–1004; 1049–1052; 1092–1095; 1144–1146; 1185–1188.
McClelland, S., and R.E. Maxwell. 2007. ‘Hemispherectomy for Intractable Epilepsy in Adults: The First Reported Series’. *Annals of Neurology* 61 (4): 372–376.
McGinn, C. 2008. ‘Consciousness as Knowingness’. *The Monist* 91 (2): 237–249.
McGovern, R.A., A.N.V. Moosa, L. Jehi, R. Busch, L. Ferguson, A. Gupta, J. Gonzalez-Martinez, E. Wyllie, I. Najm, and W.E. Bingaman. 2019. ‘Hemispherectomy in Adults and Adolescents: Seizure and Functional Outcomes in 47 Patients’. *Epilepsia* 60 (12): 2416–2427.
McMahan, J. 1995. ‘The Metaphysics of Brain Death’. *Bioethics* 9 (2): 91–126.
———. 2002. *The Ethics of Killing: Problems at the Margins of Life*. Oxford: Oxford University Press.
———. 2006. ‘An Alternative to Brain Death’. *Journal of Law, Medicine, and Ethics* 34: 44–48.
———. 2009. ‘Death, Brain Death, and Persistent Vegetative State’. In *A Companion to Bioethics*, edited by Kuhse, H. and P. Singer, 2nd ed., 286–298. Malden: Wiley-Blackwell.
Meaidi, A., P. Jennum, M. Ptito, and R. Kupers. 2014. ‘The Sensory Construction of Dreams and Nightmare Frequency in Congenitally Blind and Late Blind Individuals’. *Sleep Medicine* 15 (5): 586–595.
Meier, L.J. 2016. ‘Is Brain Death Death?’ BPhil Thesis, Oxford: University of Oxford.
———. 2019. ‘What Matters in the Mirror of Time: Why Lucretius’ Symmetry Argument Fails’. *Australasian Journal of Philosophy* 97 (4): 651–660.
———. 2020a. ‘Are the Irreversibly Comatose Still Here? The Destruction of Brains and the Persistence of Persons’. *Journal of Medical Ethics* 46 (2): 99–103.
———. 2020b. ‘Memories without Survival: Personal Identity and the Ascending Reticular Activating System’. *The Journal of Medicine and Philosophy* [forthcoming].
———. 2020c. ‘The Demise of Brain Death’. *The British Journal for the Philosophy of Science* [forthcoming].
Mele, A.R. 2003. *Motivation and Agency*. New York: Oxford University Press.
Mendelson, W.B. 1987. *Human Sleep: Research and Critical Care*. New York: Plenum.
Merker, B. 2007. ‘Consciousness without a Cerebral Cortex: A Challenge for Neuroscience and Medicine’. *Behavioral and Brain Sciences* 30 (1): 63–81.
Merkle, R.C. 1992. ‘The Technical Feasibility of Cryonics’. *Medical Hypotheses* 39 (1): 6–16.
Merricks, T. 1999. ‘Persistence, Parts, and Presentism’. *Nous* 33 (3): 421–438.
———. 2001. *Objects and Persons*. Oxford: Clarendon Press.
Miller, F.G., and R.D. Truog. 2009. ‘The Incoherence of Determining Death by Neurological Criteria: A Commentary on “Controversies in the Determination of Death”, A White Paper by the President’s Council on Bioethics’. *Kennedy Institute of Ethics Journal* 19 (2): 185–193.
———. 2016. *Death, Dying, and Organ Transplantation: Reconstructing Medical Ethics at the End of Life*. Oxford: Oxford University Press.
Moll, C.K.E., A. Sharott, W. Hamel, A. Münchau, C. Buhmann, U. Hidding, S. Zittel, M. Westphal, D. Müller, and A.K. Engel. 2009. ‘Waking Up the Brain: A Case Study of Stimulation-Induced Wakeful Unawareness during Anaesthesia’. In

Coma Science: Clinical and Ethical Implications, edited by Laureys, S., N.D. Schiff, and A.M. Owen, 125–145. Amsterdam: Elsevier.
Mollaret, P., and M. Goulon. 1959. ‘Le coma dépassé’. *Revue Neurologique* 101 (1): 3–15.
Montgomery, P.C., P.H.R. Evans, and P.J. Gullane, eds. 2009. *Principles and Practice of Head and Neck Surgery and Oncology*. 2nd ed. London: informa.
Monti, M.M., A. Vanhaudenhuyse, M.R. Coleman, M. Boly, J.D. Pickard, L. Tshibanda, A.M. Owen, and S. Laureys. 2010. ‘Willful Modulation of Brain Activity in Disorders of Consciousness’. *New England Journal of Medicine* 362 (7): 579–589.
Morin, A. 2007. ‘Consciousness Is More than Wakefulness’. *Behavioral and Brain Sciences* 30 (1): 99.
Moruzzi, G., and H.W. Magoun. 1949. ‘Brain Stem Reticular Formation and Activation of the EEG’. *Electroencephalography and Clinical Neurophysiology* 1 (4): 455–473.
Moskopp, D. 2015. *Hirntod: Konzept – Kommunikation – Verantwortung*. Stuttgart: Thieme.
Multi-Society Task Force on PVS. 1994. ‘Medical Aspects of the Persistent Vegetative State. First Part’. *The New England Journal of Medicine* 330 (21): 1499–1508.
Naci, L., and A.M. Owen. 2013. ‘Making Every Word Count for Nonresponsive Patients’. *JAMA Neurology* 70 (10): 1235–1241.
Nagel, J. 2013. ‘Knowledge as a Mental State’. In *Oxford Studies in Epistemology*, edited by Gendler, T.S. and J. Hawthorne, 4:273–308. Oxford: Oxford University Press.
Nagel, T. 1974. ‘What Is It Like to Be a Bat?’ *The Philosophical Review* 83 (4): 435–450.
———. 1985. ‘Brain Bisection and the Unity of Consciousness’. In *Mortal Questions*, 147–164. Cambridge: Cambridge University Press.
———. 1986. *The View from Nowhere*. New York: Oxford University Press.
Nair-Collins, M., and F.G. Miller. 2017. ‘Do the “Brain Dead” Merely Appear to Be Alive?’ *Journal of Medical Ethics* 43 (11): 747–753.
Niazi, S.A., and F.J. Lewis. 1958. ‘Profound Hypothermia in Man’. *Annals of Surgery* 147 (2): 264–266.
Nida-Rümelin, M., and D. O Conaill. 2019. ‘Qualia: The Knowledge Argument’. *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University.
Noonan, H.W. 2003. *Personal Identity*. 2nd ed. London: Routledge.
Olson, E.T. 1997. ‘Was I Ever a Fetus?’ *Philosophy and Phenomenological Research* 57 (1): 95–110.
———. 1999. *The Human Animal: Personal Identity without Psychology*. New York: Oxford University Press.
———. 2007. *What Are We?: A Study in Personal Ontology*. Oxford: Oxford University Press.
Oterdoom, L.H., S.M. ten Dam, S.D.W. de Groot, W. Arjaans, and A.A. van Bodegraven. 2014. ‘Limited Long-Term Survival after in-Hospital Intestinal Failure Requiring Total Parenteral Nutrition’. *American Journal of Clinical Nutrition* 100 (4): 1102–1107.
Ouaknine, G.E. 1978. ‘Cardiac and Metabolic Alterations in Brain Death: Discussion Paper’. *Annals of the New York Academy of Sciences* 315 (1): 252–264.

Oudenhove, L. van, S. McKie, D. Lassman, B. Uddin, P. Paine, S. Coen, L. Gregory, J. Tack, and Q. Aziz. 2011. ‘Fatty Acid–Induced Gut-Brain Signaling Attenuates Neural and Behavioral Effects of Sad Emotion in Humans’. *The Journal of Clinical Investigation* 121 (8): 3094–3099.
Owen, A.M., M.R. Coleman, M. Boly, M.H. Davis, S. Laureys, and J.D. Pickard. 2006. ‘Detecting Awareness in the Vegetative State’. *Science* 313 (5792): 1402.
Pallis, C., and D.H. Harley. 1996. *ABC of Brainstem Death*. 2nd ed. London: BMJ.
Parfit, D. 1971. ‘Personal Identity’. *The Philosophical Review* 80 (1): 3–27.
———. 1987. *Reasons and Persons*. Oxford: Clarendon Press.
———. 2012. ‘We Are Not Human Beings’. *Philosophy* 87 (1): 5–28.
Park, S.E., D.S. Choi, H.S. Shin, H.J. Baek, H.C. Choi, J.E. Kim, H.Y. Choi, and M.J. Park. 2017. ‘Splenial Lesions of the Corpus Callosum: Disease Spectrum and MRI Findings’. *Korean Journal of Radiology* 18 (4): 710–721.
Parvizi, J., and A.R. Damasio. 2003. ‘Neuroanatomical Correlates of Brainstem Coma’. *Brain* 126 (7): 1524–1536.
Perry, J. 1972. ‘Can the Self Divide?’ *The Journal of Philosophy* 69 (16): 463–488.
Pittsburgh Post-Gazette. 1934. ‘Atom Energy Hope Is Spiked by Einstein’, December 29.
Plato. 1997. *Complete Works*. Edited by Cooper, J.M. and D.S. Hutchinson. Indianapolis: Hackett.
Plum, F., and J.B. Posner. 1980. *The Diagnosis of Stupor and Coma*. 3rd ed. Philadelphia: Davis.
Posluszny, J., P.T. Rycus, R.H. Bartlett, M. Engoren, J.W. Haft, W.R. Lynch, P.K. Park, K. Raghavendran, and L.M. Napolitano. 2016. ‘Outcome of Adult Respiratory Failure Patients Receiving Prolonged (≥14 Days) ECMO’. *Annals of Surgery* 263 (3): 573–581.
Potts, M. 2001. ‘A Requiem for Whole Brain Death: A Response to D. Alan Shewmon’s “The Brain and Somatic Integration”’. *The Journal of Medicine and Philosophy* 26 (5): 479–491.
Powley, T.L. 2000. ‘Vagal Input to the Enteric Nervous System’. *Gut* 47 (Suppl. 4): iv30–iv32.
President’s Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research. 1981. *Defining Death: A Report on the Medical, Legal and Ethical Issues in the Determination of Death*. Washington: U.S. Government Printing Office.
Prezioso, G., C. Giannini, and F. Chiarelli. 2018. ‘Effect of Thyroid Hormones on Neurons and Neurodevelopment’. *Hormone Research in Paediatrics* 90 (2): 73–81.
Price, H.H. 1969. *Belief: The Gifford Lectures Delivered at the University of Aberdeen in 1960*. London: Allan & Unwin.
Puccetti, R. 1969. ‘Brain Transplantation and Personal Identity’. *Analysis* 29 (3): 65–77.
———. 1973. ‘Brain Bisection and Personal Identity’. *The British Journal for the Philosophy of Science* 24: 339–355.
———. 1981. ‘The Case for Mental Duality: Evidence from Split-Brain Data and Other Considerations’. *Behavioral and Brain Sciences* 4 (1): 93–99.
Putnam, H. 1967. ‘Psychological Predicates’. In *Art, Mind, and Religion: Proceedings of the 1965 Oberlin Colloquium in Philosophy*, edited by Capitain, W.H. and D.D. Merrill, 37–48. Pittsburgh: University of Pittsburgh Press.

———. 1999. ‘Brains in a Vat’. In *Knowledge: Readings in Contemporary Epistemology*, edited by Bernecker, S. and F.I. Dretske, 1–21. Oxford University Press.
Quante, M. 1995. ‘Wann ist ein Mensch tot? Zum Streit um den menschlichen Tod’. *Zeitschrift für philosophische Forschung* 49 (2): 167–193.
———. 2002. *Personales Leben und menschlicher Tod: Personale Identität als Prinzip der biomedizinischen Ethik*. Frankfurt a. M.: Suhrkamp.
Quigg, M. 2019. ‘The Wada Test: The Best Steam Engine on the Tracks’. *Epilepsy Research* 156 (106157): 1–2.
Rachels, J. 1986. *The End of Life: Euthanasia and Morality*. Oxford: Oxford University Press.
Reid, M.D. 2016. ‘A Case in Which Two Persons Exist in One Animal’. In *Animalism: New Essays on Persons, Animals, and Identity*, edited by Blatti, S. and P.F. Snowdon, 253–265. Oxford: Oxford University Press.
Reid, T. 1878. *Essays on the Intellectual Powers of Man*. Edited by Walker, J. 10th ed. Philadelphia: Butler.
Rescher, N. 2005. *What If?: Thought Experimentation in Philosophy*. New Brunswick: Transaction.
Robbins, T.W. 1997. ‘Arousal Systems and Attentional Processes’. *Biological Psychology* 45 (1–3): 57–71.
Roskies, A.L. 2016. ‘Neuroscience’. In *The Oxford Handbook of Philosophical Methodology*, edited by Cappelen, H., T.S. Gendler, and J. Hawthorne, 587–606. Oxford: Oxford University Press.
Sade, R.M., and A. Boan. 2014. ‘The Paradox of the Dead Donor Rule: Increasing Death on the Waiting List’. *The American Journal of Bioethics* 14 (8): 21–23.
Sadovnikoff, N., and D. Wikler. 2014. ‘Brain Dead Patients Are Still Whole Organisms’. *The American Journal of Bioethics* 14 (8): 39–40.
Sandberg, A., and N. Bostrom. 2008. *Whole Brain Emulation: A Roadmap*. Oxford: Future of Humanity Institute.
Savulescu, J. 2003. ‘Death, Us and Our Bodies: Personal Reflections’. *Journal of Medical Ethics* 29 (3): 127–130.
Savulescu, J., and I. Persson. 2016. ‘Conjoined Twins: Philosophical Problems and Ethical Challenges’. *Journal of Medicine and Philosophy* 41 (1): 41–55.
Schechter, E. 2015. ‘The Subject in Neuropsychology: Individuating Minds in the Split-Brain Case’. *Mind & Language* 30 (5): 501–525.
———. 2018. *Self-Consciousness and ‘Split’ Brains: The Minds’ I*. Oxford: Oxford University Press.
Schiff, N.D., J.T. Giacino, K. Kalmar, J.D. Victor, K. Baker, M. Gerber, B. Fritz, et al. 2007. ‘Behavioural Improvements with Thalamic Stimulation after Severe Traumatic Brain Injury’. *Nature* 448 (7153): 600–603.
Schlake, H.-P., and K. Roosen. 2001. *Der Hirntod als der Tod des Menschen*. 2nd ed. Neu-Isenburg: Deutsche Stiftung Organtransplantation.
Schneider, W., and P.A. Ornstein. 2019. ‘Determinants of Memory Development in Childhood and Adolescence’. *International Journal of Psychology* 54 (3): 307–315.
Schwager, R.L. 1978. ‘Life, Death, and the Irreversibly Comatose’. In *Ethical Issues in Death and Dying*, edited by Beauchamp, T.L. and S. Perlin, 38–50. Englewood Cliffs: Prentice-Hall.

Schweizerische Akademie der Medizinischen Wissenschaften. 2011. *Feststellung des Todes mit Bezug auf Organtransplantationen: Medizin-ethische Richtlinien*. Basel: vista point.
Searle, J.R. 1980. 'Minds, Brains, and Programs'. *The Behavioral and Brain Sciences* 3 (3): 417–424.
———. 1992. *The Rediscovery of the Mind*. Cambridge, MA: MIT Press.
Shewmon, D.A. 1997. 'Recovery from "Brain Death": A Neurologist's Apologia'. *The Linacre Quaterly* 64 (1): 30–96.
———. 1999. 'Spinal Shock and "Brain Death": Somatic Pathophysiological Equivalence and Implications for the Integrative-Unity Rationale'. *Spinal Cord* 37 (5): 313–324.
———. 2001. 'The Brain and Somatic Integration: Insights into the Standard Biological Rationale for Equating "Brain Death" with Death'. *Journal of Medicine and Philosophy* 26: 457–478.
———. 2007. 'Comments on the Transcript of the Discussion'. In *The Signs of Death: The Proceedings of the Working Group 11-12 September 2006*, edited by Sánchez Sorondo, M., 371–381. Scripta Varia 110. Vatican City: Pontificia Academia Scientiarum.
———. 2012. 'You Only Die Once: Why Brain Death Is Not the Death of a Human Being'. *Communio* 39: 422–494.
———. 2018. 'Brain Death: A Conclusion in Search of a Justification'. *Hastings Center Report* 48 (6): S22–S25.
Shewmon, D.A., G.L. Holmes, and P.A. Byrne. 1999. 'Consciousness in Congenitally Decorticate Children: Developmental Vegetative State as Self-Fulfilling Prophecy'. *Developmental Medicine & Child Neurology* 41 (6): 364–374.
Shoemaker, S. 1964. *Self-Knowledge and Self-Identity*. Ithaca: Cornell University Press.
———. 1970. 'Persons and Their Pasts'. *American Philosophical Quarterly* 7 (4): 269–285.
———. 1992. 'Unger's Psychological Continuity Theory'. *Philosophy and Phenomenological Research* 52 (1): 139–143.
———. 1997. 'Self and Substance'. *Philosophical Perspectives* 31 (11): 283–304.
———. 1999. 'Self, Body, and Coincidence'. *Aristotelian Society Supplementary Volume* 73 (1): 287–306.
———. 2008. 'Persons, Animals, and Identity'. *Synthese* 162 (3): 313–324.
Shoemaker, S., and R. Swinburne. 1984. *Personal Identity*. Oxford: Blackwell.
Sider, T. 2001. *Four-Dimensionalism: An Ontology of Persistence and Time*. Oxford: Clarendon Press.
Silbernagl, S., and A. Despopoulos. 2009. *Color Atlas of Physiology*. 6th ed. Stuttgart: Thieme.
Silva, J.R., A.C. Sakamoto, Ú. Thomé, S. Escorsi-Rosset, M.V. Santos, H.R. Machado, A.C. Santos, and A.P. Hamad. 2020. 'Left Hemispherectomy in Older Children and Adolescents: Outcome of Cognitive Abilities'. *Child's Nervous System* 36: 1275–1282.
Snowdon, P.F. 1990. 'Persons, Animals, and Ourselves'. In *The Person and the Human Mind: Issues in Ancient and Modern Philosophy*, edited by Gill, C., 83–107. Oxford: Clarendon Press.
———. 1991. 'Personal Identity and Brain Transplants'. *Royal Institute of Philosophy Supplement* 29: 109–126.
Sorensen, R.A. 1992. *Thought Experiments*. New York: Oxford University Press.

Sperry, R.W. 1984. ‘Consciousness, Personal Identity and the Divided Brain’. *Neuropsychologia* 22 (6): 661–673.
Sperry, R.W., and B. Preilowski. 1972. ‘Die beiden Gehirne des Menschen’. *Bild der Wissenschaft* 9: 921–927.
Squire, L.R., and P. Alvarez. 1995. ‘Retrograde Amnesia and Memory Consolidation: A Neurobiological Perspective’. *Current Opinion in Neurobiology* 5 (2): 169–177.
Squire, L.R., and E.R. Kandel. 1999. *Memory: From Mind to Molecules*. New York: Scientific American Library.
Squire, L.R., and S.M. Zola. 1998. ‘Episodic Memory, Semantic Memory, and Amnesia’. *Hippocampus* 8 (3): 205–211.
Stecker, M.M., A.T. Cheung, A. Pochettino, G.P. Kent, T. Patterson, S.J. Weiss, and J.E. Bavaria. 2001. ‘Deep Hypothermic Circulatory Arrest: I. Effects of Cooling on Electroencephalogram and Evoked Potentials’. *The Annals of Thoracic Surgery* 71 (1): 14–21.
Steigleder, K. 2015. ‘Sind hirntote Menschen wirklich tot?’ In *Anthropologie 2.0? Neuere Ansätze einer philosophischen Anthropologie im Zeitalter der Biowissenschaften*, edited by Baedke, J., C. Brandt, and H.-U. Lessing, 103–115. Münster: LIT.
Steinhart, E. 2001. ‘Persons versus Brains: Biological Intelligence in Human Organisms’. *Biology and Philosophy* 16 (1): 3–27.
Strawson, G. 2010. *Mental Reality*. 2nd ed. Cambridge, MA: MIT Press.
Sturm, V., A. Kühner, H.P. Schmitt, H. Assmus, and G. Stock. 1979. ‘Chronic Electrical Stimulation of the Thalamic Unspecific Activating System in a Patient with Coma Due to Midbrain and Upper Brain Stem Infarction’. *Acta Neurochirurgica* 47 (3–4): 235–244.
Tatomir, A., C. Micu, and C. Crivii. 2014. ‘The Impact of Stress and Glucocorticoids on Memory’. *Clujul Medical* 87 (1): 3–6.
The Miami Metropolis. 1911. ‘Interview with Thomas Edison’, June 20.
Thompson, E., and D. Cosmelli. 2011. ‘Brain in a Vat or Body in a World?: Brainbound versus Enactive Views of Experience’. *Philosophical Topics* 39 (1): 163–180.
Thomson, J.J. 1997. ‘People and Their Bodies’. In *Reading Parfit*, edited by Dancy, J., 202–229. Oxford: Blackwell.
Torregrossa, G., M. Morshuis, R. Varghese, L. Hosseinian, V. Vida, V. Tarzia, A. Loforte, et al. 2014. ‘Results with Syncardia Total Artificial Heart Beyond 1 Year’. *ASAIO Journal* 60 (6): 626–634.
Treays, J. 2005. ‘Real Life: The Man with the Seven Second Memory’. *Real Stories [Video]*. ITV Studios.
Truog, R.D., and F.G. Miller. 2008. ‘The Dead Donor Rule and Organ Transplantation’. *New England Journal of Medicine* 359 (7): 674–675.
Tye, M. 2003. *Consciousness and Persons: Unity and Identity*. Cambridge, MA: MIT Press.
Unger, P. 1990. *Identity, Consciousness and Value*. New York: Oxford University Press.
Veatch, R.M. 1993. ‘The Impending Collapse of the Whole-Brain Definition of Death’. *The Hastings Center Report* 23 (4): 18–24.
———. 2008. ‘Donating Hearts after Cardiac Death: Reversing the Irreversible’. *New England Journal of Medicine* 359 (7): 672–673.
———. 2010. ‘Transplanting Hearts after Death Measured by Cardiac Criteria: The Challenge to the Dead Donor Rule’. *Journal of Medicine and Philosophy* 35 (3): 313–329.

Veatch, R.M., M. DeVita, T. Tomlinson, D. Rodríguez-Arias Vailhen, M.J. Smith, R.A. DeWeese, P. Lee, R.P. George, and D. Marquis. 2011. ‘The Not-So-Tell-Tale Heart’. *The Hastings Center Report* 41 (2): 4–11.
Voytko, M.L., G.P. Tinkler, C. Browne, and J.R. Tobin. 2009. ‘Neuroprotective Effects of Estrogen Therapy for Cognitive and Neurobiological Profiles of Monkey Models of Menopause’. *American Journal of Primatology* 71 (9): 794–801.
Walker, A.E., R. Bickford, M. Aung, and D. Poskanzer. 1977. ‘An Appraisal of the Criteria of Cerebral Death: A Summary’. *Journal of the American Medical Association* 237 (10): 982–986.
Wearing, D. 2005. *Forever Today: A Memoir of Love and Amnesia*. London: Transworld.
Wenk, G.L. 2003. ‘Neuropathologic Changes in Alzheimer’s Disease’. *The Journal of Clinical Psychiatry* 64 (Suppl. 9): 7–10.
Wertheimer, P., M. Jouvet, and J. Descotes. 1959. ‘A propos du diagnostic de la mort du système nerveux dans les comas avec arrêt respiratoire traités par respiration artificielle’. *La Presse Médicale* 67 (3): 87–88.
Wheeler, M. 2018. ‘Talking about More than Heads: The Embodied, Embedded and Extended Creative Mind’. In *Creativity and Philosophy*, edited by Gaut, B.N. and M. Kieran, 230–250. New York: Routledge.
White, R.J., L.R. Wolin, L.C. Massopust, N. Taslitz, and J. Verdura. 1971. ‘Primate Cephalic Transplantation: Neurogenic Separation, Vascular Association’. *Transplantation Proceedings* 3 (1): 602–604.
Wiggins, D. 1967. *Identity and Spatio-Temporal Continuity*. Oxford: Blackwell.
———. 1980. *Sameness and Substance*. Oxford: Blackwell.
———. 2003. *Sameness and Substance Renewed*. Cambridge: Cambridge University Press.
Wijdicks, E.F.M., and J.L.D. Atkinson. 2001. ‘Pathophysiologic Responses to Brain Death’. In *Brain Death*, edited by Wijdicks, E.F.M., 29–43. Philadelphia: Lippincott Williams & Wilkins.
Wikler, D. 1984. ‘Brain Death’. *Journal of Medical Ethics* 10 (2): 101–102.
———. 1993. ‘Brain Death: A Durable Consensus?’ *Bioethics* 7 (2–3): 239–246.
Wilkes, K.V. 1999. *Real People: Personal Identity without Thought Experiments*. Oxford: Clarendon Press.
Williams, B. 1970. ‘The Self and the Future’. *The Philosophical Review* 79 (2): 161–180.
Windt, J.M., T. Nielsen, and E. Thompson. 2016. ‘Does Consciousness Disappear in Dreamless Sleep?’ *Trends in Cognitive Sciences* 20 (12): 871–882.
Winslow, J.-B. 1748. *The Uncertainty of the Signs of Death*. Dublin: Faulkner.
Wittgenstein, L. 1967. *Zettel*. Edited by Anscombe, G.E.M. and G.H. von Wright. Translated by Anscombe, G.E.M. Berkeley: University of California Press.
Wollheim, R. 1999. *On the Emotions: The Ernst Cassirer Lectures 1991*. New Haven: Yale University Press.
Yamamoto, T., and Y. Katayama. 2005. ‘Deep Brain Stimulation Therapy for the Vegetative State’. *Neuropsychological Rehabilitation* 15 (3–4): 406–413.
Youngner, S.J., and E.T. Bartlett. 1983. ‘Human Death and High Technology: The Failure of the Whole-Brain Formulations’. *Annals of Internal Medicine* 99 (2): 252–258.
Zagon, A. 2001. ‘Does the Vagus Nerve Mediate the Sixth Sense?’ *Trends in Neurosciences* 24 (11): 671–673.
Zaner, R.M., ed. 1988. *Death: Beyond Whole-Brain Criteria*. Dordrecht: Kluwer.

Zeman, A. 2001. ‘Consciousness’. *Brain* 124 (7): 1263–1289.
———. 2002a. *Consciousness: A User’s Guide*. New Haven: Yale University Press.
———. 2002b. ‘The Persistent Vegetative State: Conscious of Nothing?’ *Practical Neurology* 2 (4): 214–217.
———. 2006. ‘What Do We Mean by “Conscious” and “Aware”?’ *Neuropsychological Rehabilitation* 16 (4): 356–376.
Zwarts, M., F.H.M. Kornips, and O.M. Vogels. 2001. ‘Clinical Brainstem Death with Preserved Electroencephalographic Activity and Visual Evoked Response’. *Archives of Neurology* 58 (6): 1010.

Candidate’s declaration

I, Lukas Jost Meier, do hereby certify that this thesis, submitted for the degree of PhD, which is approximately 65,000 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for any degree.

I was admitted as a research student at the University of St Andrews in September 2016.

I received funding from an organisation or institution and have acknowledged the funder in the full text of my thesis.

Date 01.07.2020 Signature of candidate

Supervisor's declaration

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date 01.07.2020 Signature of supervisor

Permission for publication

In submitting this thesis to the University of St Andrews we understand that we are giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. We also understand, unless exempt by an award of an embargo as requested below, that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that this thesis will be electronically accessible for personal or research use and that the library has the right to migrate this thesis into new electronic forms as required to ensure continued access to the thesis.

I, Lukas Jost Meier, have obtained, or am in the process of obtaining, third-party copyright permissions that are required or have requested the appropriate embargo below.

The following is an agreed request by candidate and supervisor regarding the publication of this thesis:

Printed copy

No embargo on print copy.

Electronic copy

Embargo on all of electronic copy for a period of 3 years on the following ground(s):

- Publication would be in breach of law or ethics

Supporting statement for electronic embargo request

The thesis contains extensive extracts from three journal articles. I have granted the respective publishers an exclusive licence to publish.

Title and Abstract

- I agree to the title and abstract being published.

Date 01.07.2020 Signature of candidate

Date 01.07.2020 Signature of supervisor

Underpinning Research Data or Digital Outputs

Candidate's declaration

I, Lukas Jost Meier, hereby certify that no requirements to deposit original research data or digital outputs apply to this thesis and that, where appropriate, secondary data used have been referenced in the full text of my thesis.

Date 01.07.2020 Signature of candidate