

Classic Text 32 – Philosophy of Mind: Personal Identity and the Self

Both the ideas of personal identity and the self are rightly or wrongly so bound up with each other that it is difficult to discuss the one without reference to the other. There are two classic texts for this study unit. The first is an extract of chapters 10 - 13 from part 3 of Derek Parfit's *Reasons and Persons* (1984, 1987). The second is a chapter by Daniel Dennett – *The Self as a Center of Narrative Gravity* from the book *Self and Consciousness: Multiple Perspectives* (1992) that first appeared in Danish in 1986. Both can be downloaded for free [here](#) and [here](#). Note that under South African copyright law individual chapters and articles may be reproduced for educational purposes.



A Young Derek Parfit (1942 - 2017), One of the Most Influential Philosophers of the Late 20th and Early 21st Century with a Special Interest in Personal Identity, Rationality and Ethics

Reasons and Persons

Parfit's *Reasons and persons* received widespread praise for being "widely viewed as an outstanding contribution to a cluster of questions in metaphysics and ethics" (Philip Kitcher); "brilliantly clever and imaginative" (Bernard Williams) and according to P.F. Strawson, "Very few works in the subject can compare with Parfit's in scope, fertility, imaginative resource, and cogency of reasoning". David Chalmers, meanwhile, said in an interview that it gave him a "sense of how powerful analytic philosophy can be when done clearly and accessibly." (Wikipedia: *Reasons and Persons*) We trust that you will come to appreciate why.

Most popular accounts of personal identity assume that our existence is some deep and significant fact about the world. Using a number of thought experiments as well as actual, well documented cases, Parfit endeavours to show that this view is wrong and that there are no further facts about the world that make one person and the person at a later (or earlier) time one and the same. Instead, Parfit argues for a deflationary view of personal identity, according to which, what matters is simply a "Relation R", psychological connectedness, including memory, personality *etc.* Parfit's conclusion is similar Hume's "bundle theory" and to the view of the self in Buddhism's Skandha (Sanskrit for "heaps, aggregates, collections, groupings"). (Wikipedia: *Reasons and Persons*)

Parfit's *Reasons and persons* is divided into four parts: Self-defeating ethical theories, Rationality and Time, Personal identity and Responsibility towards future generations. Each part comprises of several chapters, within which there are several clusters of argument which are separately numbered. Our extract of part three begins with chapter 10.

10 What We Believe Ourselves to Be

At our present state of technology we are able to teleport individual atoms from one place to another without traversing the distance in between, but not organisms – not even very simple ones, let alone humans. Teleportation of humans, at this stage, is logically possible, in other words or not logically impossible, but not technologically feasible.

The motivation behind using imaginary or science fictional scenarios in philosophy or even physics is that they serve as analogies to real world cases that we may be trying to understand. If the scenario is a good analogy it will allow for the transfer of intuitions and insights from the imagined to the real-world case. Indeed, in our very first Classical Text, we considered two such analogies, separated by some 2 300 years – Plato's *Allegory of the Cave* and the film *The Matrix*. Parfit begins his science fictional scenario as follows:

I enter the Teletransporter. I have been to Mars before, but only by the old method, a space-ship journey taking several weeks. This machine will send me at the speed of light. I merely have to press the green button. Like others, I am nervous. Will it work? I remind myself what I have been told to expect. 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. Though I believe that this is what will happen, I still hesitate. But then I remember seeing my wife grin when, at breakfast today, I revealed my nervousness. As she reminded me, she has been often teletransported, and there is nothing wrong with *her*. I press the button. As predicted, I lose and seem at once to regain consciousness, but in a different cubicle. Examining my new body, I find no change at all. Even the cut on my upper lip, from this morning's shave, is still there. Several years pass, during which I am often Teletransported. I am now back in the cubicle, ready for another trip to Mars. But this time, when I press the green button, I do not lose consciousness. There is a whirring sound, then silence. I leave the cubicle, and say to the attendant: 'It's not working. What did I do wrong?' 'It's working', he replies, handing me a printed card. This reads: 'The New Scanner records your blueprint without destroying your brain and body. We hope that you will welcome the opportunities which this technical advance offers.' The attendant tells me that I am one of the first people to use the New Scanner. He adds that, if I stay for an hour, I can use the Intercom to see and talk to myself on Mars. 'Wait a minute', I reply, 'If I'm here I can't *also* be on Mars'.

Someone politely coughs, a white-coated man who asks to speak to me in private. We go to his office, where he tells me to sit down, and pauses. Then he says: 'I'm afraid that we're having problems with the New Scanner. It records your blueprint just as accurately, as you will see when you talk to yourself on Mars. But it seems to be damaging the cardiac systems which it scans. Judging from the results so far, though you will be quite healthy on Mars, here on Earth you must expect cardiac failure within the next few days.' The attendant later calls me to the Intercom. On the screen I see myself just as I do in the mirror every morning. But there are two differences. On the screen I am not left-right reversed. And, while I stand here speechless, I can see and hear myself, in the studio on Mars, starting to speak. (p. 199 - 200)

Parfit's scenario arouses some strong beliefs within us, not so much about the words he uses, such as personal pronouns, but about ourselves and our continued existence, as well as the nature of our

personal identity over time. According to Parfit, in what follows, some of these beliefs will turn out to be false, in a way that he suggests matters.

75. Simple Teletransportation and the Branch-Line Case

In the simplest case, the Scanner destroys my body and transmits a blueprint to Mars, where another machine assembles an exact *Replica* of me out of atoms there. My Replica will resemble me in every physical and psychological respect. Indeed he will think that he is me, and will have all my memories up to the moment I pressed the green button. If he returned to earth, everyone I know will believe he was me. Simple Teleportation of this sort is a common device in science fiction. While some readers of this fiction would believe that my Replica *would be me*, others would dispute this, believing that I died when I pressed the green button and that my Replica is, in fact, *someone else*, reconstructed to be exactly like me.

The second part of Parfit's fiction seems to support the latter view. If the New Scanner successfully transmits my blueprint to Mars but does not destroy my body, merely damaging my heart instead, I would exit the cubicle as if nothing happened since I pressed the green button, except to learn that in a few days I shall die... Later I talk via a two way video link to my Replica on Mars.

Since my Replica knows that I am about to die, he tries to console me with the same thoughts with which I recently tried to console a dying friend. It is sad to learn, on the receiving end, how unconsoling these thoughts are. My Replica then assures me that he will take up my life where I leave off. He loves my wife, and together they will care for my children. And he will finish the book that I am writing. Besides having all of my drafts, he has all of my intentions. I must admit that he can finish my book as well as I could. All these facts console me a little. Dying when I know that I shall have a Replica is not quite as bad as, simply, dying. Even so, I shall soon lose consciousness, forever. (p. 201)

With Simple Teletransportation, where I am destroyed before being replicated, we are inclined to believe that this is a mode of speed of light transport, where my Replica *is* me. In the second part of the story, the end of my life and the rest of my Replica's life overlap. Parfit calls this the *Branch-Line Case*, in which I believe that my Replica *cannot* be me, since I am talking to him. Even though he is exactly like me, I feel that he is one person and I am another. If I pinch myself, he will not know. When I have my fatal heart attack, he will feel nothing. And when I am dead, he will enjoy another forty years of life on the *Main Line* of our briefly overlapping lives.

If we agree that my Replica is not me, we may feel that my prospects on the Branch Line are pretty grim – almost as bad as death. Parfit however disagrees: being destroyed and replicated is about as good as ordinary survival. (p. 201)

76. Qualitative and Numerical Identity

According to Parfit, there are two kinds of identity, where we only considered one kind in Critical Reasoning 14. My Replica and I are *quantitatively identical*, or exactly alike. But we may not be *numerically identical*, *i.e.* one and the same person. For example, two billiard balls may be qualitatively but not numerically identical. If I paint one of these balls red, it ceases to be

qualitatively identical with itself as it was; however the now red ball and the white ball as it was remain numerically identical. They are one and the same ball. (p. 201)

We might say of someone, after a stroke, that 'he is no longer the same person'. According to Parfit, this is a claim about both sorts of identity. We claim that *he* is the same person, but is not *now* the same person. This is not a contradiction. What we mean is that his character has changed. He is numerically the identical person but is now qualitatively different.

When we are concerned about our future, we are chiefly concerned about our numerical identity. I might believe that after my marriage, I shall no longer be the same person; however marriage is not death. No matter how I might change me, I shall still be the same living person who will *be* me. On the other hand, psychological changes also matter. Certain neurodegenerative diseases can cause such extreme changes that they can rob persons of their identity. A man with advanced Alzheimer's disease may remember nothing of his past, nor even be able to recognise his wife of 40 years or their children. In such a case, it is no exaggeration to say that the person he was has ceased to exist. (p. 202)

77. The Physical Criterion of Personal Identity

When we think about the nature of persons and personal identity over time, we can distinguish several questions.

- 1) What is the nature of a person?
- 2) What makes a person at two different times the same person. *I.e.* what is necessarily involved in the continuity of existence of each person over time.

The answer to 2) might take the form of 'X today is one and the same person as Y at some time in the past *if and only if* ...' such that '...' states the *necessary and sufficient conditions* for personal identity over time.

Providing an answer to 2) partly answers 1). The necessary conditions for our continued existence depend on our nature. In addition, the simplest answer to 1) is that to be a person requires that we be self conscious, and aware of our own identity and its continued existence over time. We might also ask:

- 3) What is, in fact, involved in the continued existence of each person over time?

Many of the features that we ordinarily assume are involved in the continued existence of person are not necessarily so; therefore an answer to 2) might only be a partial answer to 3). *E.g.* having the same heart or character are usually associated with our continued existence but they are not necessary. According to Parfit, some writers use the ambiguous phrase 'the criterion of identity over time' which may mean 'our way of telling whether some present object is identical with some past object'; however Parfit intends to mean *what this identity necessarily involves, or consists in*. (p. 202)

For most physical objects, what Parfit calls the *standard view*, is that the criterion of identity over time is the spatio-temporal physical continuity of the object. In the simplest cases of physical continuity, such as the Great Pyramids, apparently static objects continue to exist. In other simple

cases, objects such as the Moon, move in regular ways. Other objects move in less regular ways, but they still trace out physically continuous spatio-temporal paths. Suppose that the billiard ball that I painted red was the same ball with which I made a winning shot last year. On the standard view, this is true only if the ball traced a continuous path between then and now. Thus, it is necessary that:

- 1) there is a line in time and space between where the ball rested before making my winning shot and where it is now that it is red, and that
- 2) at every point on this line there was a billiard ball, and that
- 3) the existence of a ball at each point on the line was caused, in part, by the existence of a billiard ball at a point immediately preceding it. (p. 203)

Some things however continue to exist even though they involve great changes throughout their physical continuity – an egg becomes a caterpillar, which becomes a chrysalis, which becomes a butterfly. Other objects may have gaps in their continued existence. Consider Parfit's gold watch which he was given for his birthday as a boy. Even though it lay in pieces at the watchmaker's repair shop for a month and thus did not have a history of full physical continuity, we are inclined to say that it is the same watch, perhaps because the parts did have a full history of physical continuity. (p. 203)

Alternatively, consider a wooden ship that has been repaired from time to time while afloat in the harbour. After fifty years none of the bits of wood out of which it was originally made remain. Yet again, we are inclined to say that it is one and the same ship because, as a ship, it has displayed the same physical continuity throughout its fifty years of existence. The same is true of most animal bodies. The cells of different parts of the body are replaced several times over a lifetime. Among humans, the cells lining the gut are replaced every three days, while red blood cells are replaced every 120 days, however some neurons are never replaced. (p. 203 - 204)

The standard view of physical continuity, of what makes an object one and the same over time, suggests that what makes me the same person over time is that I have the same body (including my brain) over time. Moreover, I will continue to exist if and only if this particular body continues to exist as the body of a living person. According to an improved version of *The Physical Criterion*:

- 1) What is necessary is not the continued existence of the whole body, but the continued existence of *enough* of the brain to be the brain of a living person. Thus person X today is one and the same as person Y some time past, if and only if, 2) enough of Y's brain continued to exist and is now X's brain, and 3) this physical continuity has not taken a "branching" form. 4) Personal identity over time just consists in the holding of facts like 2) and 3).

This is true of certain real cases in which some people continue to exist even though they may lose, or lose use of much of their body, including parts of their brain. (p. 204)

According to the Physical Criterion, Teletransportation would not be a form of travel but a kind of death. Reincarnation would also be impossible on the Physical Criterion, as would resurrection of the identical, physically continuous body. The Greek and Trojan heroes believed that if they died and were burned on a funeral pyre, and their ashes scattered, not even God could bring them to life again; although he could recreate a Replica of someone else who was exactly like them. Some

Christians however believe that God could resurrect *them* if He chose to reassemble their body out of the matter that constituted them when alive. This would be analogous to the case of the gold watch. (p. 204) Other Christians believe that they will be resurrected into a new heavenly body which will be discontinuous from their earthly body.

78. The Psychological Criterion

Some people believe in a kind of psychological continuity analogous to physical continuity. This involves the continued existence of a mental entity or purely spiritual substance. To call it a *soul* would be misleading because there are accounts of the soul, such as Aristotle's hylomorphic theory that are not dualistic. (See Classic Text 13.) Parfit later returns to this view but first considers the kind of psychological continuity involving facts with which we are familiar.

The continuity of memory has perhaps been most often discussed because it is memory that makes us aware of our continued existence over time; however there is more than one kind of memory and more than one kind of forgetting.

Memory is the capacity by which information is encoded, stored, and retrieved when needed.

Short-term memory a.k.a. **working memory** allows for the recall of a few items over a period of a minute or less without rehearsal, such as saying a phone number over and again.

Long-term memory can store vast quantities of information for potentially a lifetime.

- **Declarative**, or **explicit memory** refers to the conscious storage and recollection of information. This includes **episodic memory** which refers to memory of previous experiences including their spatio-temporal and emotional context. **Semantic memory** meanwhile, refers to memory encoded with a specific meaning.
- **Non-declarative**, or **implicit memory** refers to the unconscious storage and retrieval of information. This includes **procedural memory**, which refers to the slow and unconscious learning of skills such as learning to ride a bicycle. **Priming** meanwhile, is the process of subliminally arousing specific responses from memory following exposure to a certain stimulus. *E.g.* someone who was previously exposed to the word "yellow" will evoke a faster response to the related word "banana" than the unrelated word "television".

Amnesia is a deficit in memory caused by chemical and/or physical damage or disease to the brain. Amnesia may be caused at the stage of encoding, storage, and/or or retrieval.

- **Retrograde amnesia** is the inability to retrieve information that was acquired before the date of a specific brain operation or head trauma.
- **Anterograde amnesia** is the inability to create new memories following some chemical or physical assault to the brain. Long-term memories created prior to the event(s) are usually spared.

Amnesiacs are the exception to the idea of psychological continuity because they have typically lost one or more kind of memory, usually episodic or "experience memories" as Parfit calls them. This

includes memories about one's lived experiences, including some facts about one's past life. However, other impersonal facts are remembered including procedural memory, such as how to speak or swim. (p. 205)

According to Locke (Ch. 27, §16), episodic memory provides the criterion of personal identity; however Parfit believes it can only be part of the answer. Locke, for example, believed that someone cannot have committed a crime unless he remembers doing so, but if taken to be a view about what is involved personal existence, then this is clearly false. If it were true, then forgetting about past actions and experiences would be impossible, when clearly it *is* possible. Parfit for example, claims that he could not remember putting on his shirt on the morning of his writing. (p. 205)

It is however possible to revise Locke's view of a direct memory connection by one of *continuity of memory* between a person now and the same person, say, 20 years ago. Since most people remember some of their memories of the previous day there will be an overlapping chain of direct memories over the interval. On the revised version of Locke's view, person X today is the same as some past person Y if there is a continuity of memory between them. However even this revised view requires further revision so that it appeals to other facts. Besides direct memories, there are other kinds of direct psychological connections. There is clearly a connection between an intention formed at one time and an act performed by the same person at some later time in which the intention is carried out. There are also clearly direct connections between a person who holds a belief, desire, or any other psychological feature and the same person at a later time who continues to hold the same. (p. 205)

Clinical Case 1

Henry Molaison also known as H.M. suffered from severe epilepsy and had surgery to remove the medial temporal lobes of both of his cerebral hemispheres including the hippocampi, in order to limit the extent of his seizures. The surgery was a partial success in that it controlled his seizures but left him with profound anterograde amnesia, unable to create new memories; although his working memory and procedural memory remained intact. H.M.'s sense of personal identity and agency were preserved; however he had to be reintroduced anew to the medical staff caring for him at every subsequent meeting.

Parfit defines two general psychological relations:

- *Psychological connectedness* is the holding of particular direct psychological connections, and
- *Psychological continuity* is the holding of overlapping chains of strong connectedness.

According to Parfit, of these two, connectedness is more important, both in theory and in practice. Connectedness may obtain to any degree. Between person X today and person Y yesterday there may be thousands of direct psychological connections, or only one. If there were only one, then X and Y would not be regarded as the same person on the revised Lockean View. If however there were *enough* direct psychological connections persons X and Y would be regarded as the same person. But since connectedness is a matter of degree, it is not possible to precisely define what counts as enough. Parfit claims that there is enough connectedness if the number of direct connections, over any day, is at least half the number that obtain over every day in the lives of almost all actual people. When there are enough such connections, Parfit refers to this as *strong* connectedness. (p. 206)

Recall from Critical Reasoning 14 that a relationship is transitive if x has a relation R to y and y has the same relation R to z , then x has the same relation R to z . Personal identity is a transitive relationship. Parfit offers the following example: "If Bertie was one and the same person as the philosopher Russell, and Russell was one and the same person as the author of *Why I Am Not a Christian*, this author and Bertie must be one and the same person." Strong connectedness however, is *not* a transitive relationship. If someone is strongly connected to himself yesterday and the same person was strongly connected to himself two days ago, it does not follow that that the person today is strongly connected to himself twenty years ago. Between today and twenty years ago there are fewer than the number of direct psychological connections that obtain over any day in the lives of almost all actual people. *E.g.* while most adults have many memories of experiences that they had the previous day, they have few memories of experiences that they had on any day twenty years ago. (p. 206)

By 'the criterion of personal identity over time' Parfit means *what this identity necessarily involves or consists in*. Because the identity relationship is *inter alia* transitive, any criterion of identity must also be transitive. But since strong connectedness is not transitive, it cannot be the criterion of identity. Though a defender of Locke's view cannot appeal to psychological connectedness, he can appeal to psychological continuity, which *is* transitive. According to Parfit, *The Psychological Criterion* involves:

- 1) *psychological continuity* if and only if there are overlapping chains of strong connectedness. X today is one and the same person as Y at some past time if and only if,
- 2) X is psychologically continuous with Y,
- 3) this continuity has the right kind of cause, and
- 4) it has not taken a "branching" form.
- 5) And personal identity over time just consists in facts like 2) to 4) obtaining.

Criterion 4) will be explained later. (p. 206 - 207)

There are three versions of the Psychological Criterion depending on the interpretation of the *right kind of cause*. According to the *Narrow* version, this must be the *normal* cause. On the *Wide* version, this could be *any reliable* cause. On the *Widest* version, this could be *any* cause. The *Narrow Psychological Criterion* is defined in terms of words in their ordinary sense. Thus, I remember having an experience only if,

- 1) I seem to remember having an experience,
- 2) I did have this experience, and
- 3) my apparent memory is causally dependent on this past experience.

That 3) above is required is suggested by the following example by Parfit:

Suppose that I am knocked unconscious in a climbing accident. After I recover, my fellow-climber tells me what he shouted just before I fell. In some later year, when my memories are less clear, I might seem to remember the experience of hearing my companion shout just before I fell. And it might be true that I did have just such an experience. (p. 207)

Although conditions 1) and 2) above are met, we should be sceptical that I am remembering my past experience. According to Parfit, it is a well-established fact that people can never remember their last few experiences before they were knocked unconscious. Thus my apparent memory of hearing

my companion shout out is not a real memory of that past experience because it is not causally dependent in the right way on that past experience. In all likelihood, I have reconstructed this apparent memory based on what my companion later told that he shouted out.

Similar considerations apply to other kinds of continuity, such as continuity of character. On the Narrow Psychological Criterion, even if someone's character changes radically, there must have been a continuity of character if these changes had one of several normal causes. Sometimes changes of character are brought about deliberately. At other times they may occur in response to certain experiences or aging. However there would not be continuity of character if unwanted changes were produced by abnormal interference, such as directly tampering with the brain. (p. 207)

Although memory makes us aware of our own continued existence over time, there are other continuities that are important. We may even believe (*pace* Locke) that they are significant enough to provide for personal identity in the complete absence of memory.

If we appeal to the Narrow version of the Psychological Criterion, which requires a normal cause, then, in most cases, this coincides with the Physical Criterion. The normal causes of memory depend on the continued existence of most of the brain. Similarly, all of our psychological features depend on brain states or events. Thus, the continued existence of most of a person's brain is, at least, part of the normal cause of psychological continuity. On the Physical Criterion, a person continues to exist over time if and only if the following necessary and sufficient conditions are met:

- a) *enough* of person's brain continues to exist so that it remains the brain of a living person,
- b) the physical continuity has not taken a "branching" form.

On the Narrow Psychological Criterion, a) is necessary, but not sufficient. So, on the Narrow Psychological Criterion, a person continues to exist if and only if,

- c) there is psychological continuity,
- d) this continuity has its normal cause, and
- e) it has not taken a "branching" form.

Thus, a) is required as part of the normal cause of psychological continuity. (p. 208)

Returning to the case of Teletransportation, where my body is destroyed and replicated: The Scanner and Replicator would produce a person whose body is exactly the same, and who is psychologically continuous with me from the time I pressed the green button. The cause of this continuity would not be the normal cause, but it would nonetheless be a reliable cause. On both the Physical Criterion and the Narrow Psychological Criterion, my Replica would *not* be me, however on both the Wide and Widest Criteria, he *would* be me.

According to Parfit, we need not decide between the three versions of the Psychological Criterion. Consider the following partial analogy: Some people go blind due to damage to their eyes. Suppose that scientists develop a sufficiently advanced camera and microprocessor that sends electrical impulses down the optic nerve just like those sent through this nerve by an intact retina. A blind person fitted with such a device would be able to have visual experiences just like those he used to

have before going blind.¹ His visual experiences would be causally dependent, in a new but reliable way, on the light waves reflected from the objects before him. (p. 208 - 209)

Would such a person be *seeing* these objects? If we insist that seeing must involve the normal cause, we would have to say, “No”. But even if this person cannot see *in sensu stricto*, what he has is *just as good as* seeing, both in knowing at a distance what is in sight and as a source of visual pleasure. If we accept the Psychological Criterion, we could make an analogous claim: If psychological continuity does not have its normal cause, it may not provide personal identity. But even if we insist that this is so, we might still claim that what it provides is *as good as* personal identity. (p. 209)

79. The Other Views

Parfit’s central question has been “what is the criterion of personal identity over time – what does this identity involve, or consist in?” Having described the spatio-temporal physical continuity that, on the standard view, is the criterion of identity of physical objects; he then sets out two views about personal identity, namely the Physical and Psychological Criteria.

Many contemporary philosophers assume that *Materialism*, or *Physicalism* underlies these views. According to one version of Physicalism, every mental event is just a physical event in some particular nervous system. Those who are not Physicalists are either *Dualists* or *Idealists*. Dualists, recall, believe that mental events are *not* physical events, even if all mental events are causally dependent on physical events in a brain. Idealists however, believe that all states and events, whether external or internal, are, when understood correctly, purely mental. Given these distinctions, it may be assumed that Physicalists must accept the Physical Criterion of personal identity, but this is not so. (p 209)

Physicalists could accept the Psychological Criterion, including the version that allows for any reliable cause, or even any cause. Without modifying their metaphysical belief about the nature of the mental as physically instantiated they could accept that, in the case of Simple Teletransportation, my Replica would be me. *I.e.* they could reject the Physical Criterion of personal identity, without rejecting Physicalism. These criteria however do not exhaust all the views concerning personal identity. (p. 209)

Recall that on the Physical Criterion, personal identity over time just involves the physically continuous existence of enough of a brain so that it remains the brain of a living person. On the other hand, according to the Psychological Criterion, personal identity over time just involves the various kinds of psychological continuity, with the right kind of cause. Note that both of these views are *Reductionist* because they claim that:

- 1) the fact of a person’s identity over time just consists in certain more particular facts obtaining.

They may also claim that:

¹ Cochlear implants already serve a similar function in restoring hearing to children who are born deaf, as well as some adults with severe to profound hearing loss.

- 2) these facts can be described without either presupposing the identity of this person, or explicitly claiming that the experiences in this person's life are had by this person, or even explicitly claiming that this person exists. These facts can be described in an *impersonal* way.

But it may seem that 2) could not be true. A description of the psychological continuity that unifies a person's mental life must mention the person, and many other people too, in describing the *content* of a vast number of thoughts, desires, intentions, and other mental states. But according to Parfit, mentioning this person in this way does not involve either asserting that these mental states are had by this person, or asserting that this person exists. Supporting arguments are given later. (p. 209 - 210)

If we reject both of the Reductionist claims then our position is *Non-Reductionist*. For many Non-Reductionists *we are separately existing entities*. On this view, personal identity over time does not just consist in physical and/or psychological continuity. It involves a further fact, according to which a person is an entity distinct from his body, and experiences. On the Cartesian view, a person is a *purely mental* entity, or spiritual substance. Alternatively, we might believe that that a person is a separately existing *physical* entity of a kind not yet understood by contemporary physics.

There is another Non-Reductionist View, which denies that we are separately existing entities apart from our bodies and our experience, but that our personal identity *is* some further fact, which does not just consist in physical and/or psychological continuity. Parfit calls this the *Further Fact View*. (p. 210)

Both the Physical and Psychological Criteria, of which there are different versions, are themselves versions of the Reductionist View. However, according to Parfit, what is necessarily involved in a person's continued existence is less than what is in fact involved. Thus, while adherents to different criteria disagree about imaginary cases such as Teletransportation, they do agree about what is in fact involved in the continued existence of most actual people. (p. 210 - 211)

On the Reductionist View, each person's existence just involves the existence of a body, carrying out certain actions, entertaining certain thoughts, having certain experiences, and so on. In order to simplify the description of the Reductionist View, Parfit uses the term 'event' to cover both *boring* events such as the continued existence of a belief, or a desire and terms with potentially misleading implications such as 'mental state' or even just 'state'. A 'state', for example, must be the state *of* some entity, whereas this is not implied of an event. In terms of events then, all Reductionists would accept that:

- 3) A person's existence just consists in the existence of a body, and the occurrence of a series of interrelated physical and mental events.

Some Reductionists make the stronger claim that:

- 4) A person *just is* a particular body, and such a series of interrelated events.

Other Reductionists claim that:

- 5) A person is an entity that is *distinct* from the body, and such a series of events.

On this version of the Reductionist View, persons are not only composite objects, they are entities that *have* a body, and *have* particular thoughts, desires, and so on. Although 5) is true, so is 3). A person is not a *separately existing* entity. (p. 211)

This version of Reductionism may seem self-contradictory because 3) and 5) seem to be inconsistent; however consider Hume's analogy: "I cannot compare the soul more properly to anything than to a republic, or commonwealth." (Hume, 1739–40, Part IV, § 6) Most people, including Hume, are reductionists about nations. We accept that nations exist; however Ruritania does not exist, but France does. Though nations exist, they do not exist separately from their citizens and their territory. We would accept that:

6. A nation's existence just involves the existence of its citizens, living together in certain ways, on its territory.

Some Reductionists make the stronger claim that:

7. A nation just is these citizens and this territory.

Other Reductionists claim is that:

8. A nation is an entity that is distinct from its citizens and its territory.

If we believe that claims 6) and 8) are not inconsistent, as Parfit does, then for analogous reasons, we may accept that there is no inconsistency between 3) and 5). (p. 211 - 212)

For the sake of argument, we can ignore the difference between these two versions in most of what follows. Besides claiming 1) and 2) Reductionists might also claim:

- 9) Though persons exist, we could give a *complete* description of reality *without* claiming that persons exist.

Parfit calls this the view *that a complete description could be impersonal*. This view may seem to be self-contradictory. If a person exists, and a full description of what exists fails to mention persons, how can such a description be complete? Recall, from Classic Text 17, the example of *Hesperus* and *Phosphorus* or the Evening and Morning Star, both of which are the planet Venus. A complete description of what exists could claim that Venus exists without mentioning that the Evening Star exists. We do not need to make a separate claim for the Morning Star because using another name we have already claimed that this object exists.

A similar claim can be made when some fact can be described in two or more ways. If, as in 4), a Reductionist claims that a person *just is* a particular body, and such a series of interrelated events, we can describe this fact by claiming either:

- 10) that there exists a particular body, and a particular series of interrelated physical and mental events, or
- 11) that a particular person exists.

If 10) and 11) are alternative ways of describing the same fact, a complete description need not make *both* claims, since this fact has already been mentioned in 10). (p. 212 - 213)

Other Reductionists claim, as in 5), that a person is an entity that is *distinct* from the body, and a series of events, such as the person's acts, thoughts, and other physical and mental events. On this version of Reductionism, 10) does not describe the very same fact that 11) does, although 10) may *imply* 11). According to Parfit,

More cautiously, given our understanding of the concept of a person, if we know that 10) is true, we shall know that 11) is true. These Reductionists can say that, if our description of reality either states or implies, or enables us to know about, the existence of everything that exists, our description is complete. This claim is not as clearly true as the claim that a complete description need not give two descriptions of the same fact. But this claim seems plausible. If it is justified, and the Reductionist View is true, these Reductionists can completely describe reality without claiming that persons exist.

My claims about Reductionism draw distinctions that, in this abstract form, are hard to grasp. But there are other ways of discovering whether we are Reductionists in our view about some kind of thing. If we accept a Reductionist View, we shall believe that the identity of such a thing may be, in a quite unpuzzling way, *indeterminate*. If we do *not* believe this, we are probably Non-Reductionists about this kind of thing. (p. 213)

Consider clubs, by way of example. Suppose that a club that holds regular meetings exists for several years and then comes to an end. Some years later, some of the members of the former club form a club with the same name and the same rules. We may ask 'Have these people reconvened the *very same* club? Or have they merely started *another* club, which is exactly similar?' There might be no answer to such a question. The original club may have had a rule about how, after a period of non-existence, it might be reconvened. Or it may have had a rule preventing this. But what if there were no such rule and no facts either way supporting either answer to our question? If moreover, the people involved refused to give an answer, then the claim, 'This is the same club' would be *neither true nor false*.

Although there may be no answer to our question, there may be nothing we do not already know about the situation because a club is not a separate entity from its members acting together in certain ways. The continued existence of a club just involves members having meetings, and so on according to the club's rules. If we know all the facts about how such people hold meetings and about the club's rules, then we know all there is to know about the situation. We would not be puzzled by the question 'Is this the very same club?' even without being able to give a definitive answer. Parfit call questions of this sort *empty*. (p. 213)

When asking an empty question, there is only one fact or outcome under consideration, where different answers to such a question may merely be different descriptions of the same. Even without answering such a question, we may know all there is to know about the situation. There do not have to be two different possible answers, only one of which must be true. Since an empty question has no answer, we may decide to *give* it an answer. We could decide to call the later club the same as the original. Or we could decide to call it another club, that is exactly similar. Since we already know what happened, our decision is based on a choice between two different descriptions of the very same events. (p. 214)

When applying this Reductionist claim to ourselves, it may be hard to believe. In the imagined case of Teletransportation, we are inclined to believe that the question, 'Am I about to die?' must have an answer, 'Yes' or 'No'. Any future person must be either me, or someone else. Parfit calls this belief the view that *our identity must be determinate*.

Consider the following two explanatory claims, the first of which answers a new question. What unites the different experiences that are had by a single person at the same time? While I type this paragraph, I am aware of the movements of my fingers, the sunlight on my desk and the ruffling of leaves outside. What unites these experiences? Some claim that they are all *my* experiences. They are the experiences that are being had, at this time by a particular person, or *subject of experiences*. A similar question could be asked of my whole life: What unites the different experiences that, together, constitute this life? Again, some claim that what unites these experiences is they are all *my* experiences. Parfit calls these answers the view that *psychological unity is explained by ownership*. (p. 214)

These views are about the nature of personal identity; however Parfit introduces another pair of views that are not about the nature of this identity, but about its importance. Consider an ordinary case where, on any Reductionist View, there are two possible outcomes. On one, I am about to die. On the other I shall live for many years. If these years would be years worth living, obviously the second outcome would be preferable. The difference between these outcomes would be judged as important on most theories of rationality and morality. What is judged to be important here is whether, during these years, there will be someone living who will be *me*. And this is a question about personal identity. The default view, in such a case, is that this is always what is important. Parfit calls this the view that *personal identity is what matters*. However he claims the rival view that *personal identity is not what matters*.

What matters is Relation R – i.e. psychological connectedness and/or continuity, with the right kind of cause.

More controversially, he adds as a separate claim:

In an account of what matters, the right kind of cause could be any cause. (p. 215)

Certain imaginary cases make it easier to decide whether what matters is Relation R or personal identity. Consider again the Branch-Line case, where my life briefly overlaps with that of my Replica. Suppose that we believe that my Replica and I are two different people, and that I am about to die but my Replica will live for another forty years. If personal identity is what matters, then I should regard my prospect here as nearly as bad as ordinary death. However, if what matters is Relation R, with any cause, I should regard my imminent death as about as good as ordinary survival.

The divergence between these views is not confined to imaginary cases; however the distinction is less sharp because on both views, all or nearly all real lives include the relation that matters. On all of the plausible views about the nature of personal identity, both personal identity and psychological continuity coincide, and both roughly coincide with psychological connectedness. Later Parfit argues that it makes a great deal of difference which of these we believe to be what matters. If we cease to believe that our identity is what matters, this may make a difference to some of our emotions, such

as our attitude towards aging and death. We may also be led to modify our views about rationality and morality. (p. 215)

According to Parfit, some of these views are related so that many of them stand or fall together, making it easier to judge which are true. When we see how these views are related, we shall find that there are only two alternatives. Parfit gives a preview of his arguments in the following section as to how some of these views are related.

- If we do not believe that we are separately existing entities, we cannot defensibly believe that personal identity is what matters.
- If we do not believe that we are separately existing entities, we cannot defensibly believe that personal identity does not just consist in physical and psychological continuity, but is a further fact.
- If we believe that our identity must be determinate, it does not follow that we must believe that we are separately existing entities. If we believe that we are not separately existing entities, then it is indefensible to believe that to any question about personal identity there must always be an answer, Yes or No. Only if we are separately existing entities can it be true that our identity must be determinate.
- It is possible to believe that we are separately existing entities and yet deny that our identity must be determinate; however there are few people who would combine both of these claims.
- If we believe that psychological unity is explained by ownership, then we believe that that the unity of a person's consciousness at any time is explained by the fact that their different experiences are all being had by this person. If we also we believe that the unity of a person's whole life is explained by the fact that all of the experiences in this life are had by this person, then we cannot explain this if we reject the claim that we are separately existing entities. (p. 216)

In what follows, Parfit argues for the following conclusions:

- 1) We are not separately existing entities, apart from our bodies, and various interrelated physical and mental events. Our existence just involves the existence of our bodies, our deeds, the thinking of our thoughts, and the occurrence of certain other physical and mental events. Our identity over time just involves,
 - a) Relation R *i.e.* psychological connectedness and/or psychological continuity – with the right kind of cause, provided that
 - b) this relation does not take a “branching” form, that obtains between one person and two different future people.
- 2) Our identity is not always determinate. Although we can always ask questions like ‘Am I about to die?’, it is not true that in every case such a question must have an answer, Yes or No. There would be some cases in which this would be an empty question.
- 3) There are two unities to be explained: the unity of consciousness at any time, and the unity of a whole life. These two unities cannot be explained by claiming that different experiences are had by the same person. Instead, they must be explained describing the relations between these experiences, and their relations to a person's body. Moreover, we can refer

to these experiences, and fully describe the relations between them, without claiming that these experiences are had by a particular person.

- 4) Personal identity is not what matters. What fundamentally matters is Relation R, with any cause. This relation is what matters even when, one person is R-related to two other people; although Relation R is not a criterion of personal identity. However, the relations of physical continuity, and physical similarity may be of some minor importance. (p. 216 - 217)

Parfit's strategy is as follows: first to answer some objections to his claim that we could describe our lives in an *impersonal* way. Next he attempts to show that, even if we are not aware of this, we are naturally inclined to believe, and strongly so, that our identity *must* always be determinate. Then Parfit argues that this natural belief cannot be true unless we are separately existing entities. In support of conclusion 1) he argues that we are not separately existing entities, from which the remaining three conclusions follow. (p. 216 - 217)

Although most of us would want to preserve some claims that Parfit denies, he argues that most of us have a false view about ourselves, and about our actual lives. If we are persuaded that our natural inclination about this view is false, this may make a difference to our lives. (p. 217)

11 How We Are not What We Believe

Different views about personal identity make different claims about actual people, and ordinary lives. However these differences are thrown into sharper contrast when we consider certain imaginary cases. Most of Parfit's arguments appeal, in part, to such cases. Those that contravene the laws of nature he calls *deeply* impossible; others are *merely technically* impossible, whatever progress may be made in science and technology. Whether it matters that some case may be impossible depends on the question that the case is intended to elucidate or what we are trying to show. Even in physics, it is worth considering deeply impossible cases, such as Einstein's thought experiment of running down a beam of light. We do not always need to restrict our philosophic or scientific imagination to cases which are possible. However we should bear in mind that, depending on our question, impossibility may make some thought experiments irrelevant. (p. 219)

Parfit begins with an objection to the Psychological Criterion.

80. Does Psychological Continuity Presuppose Personal Identity?

Parfit recalls, as a child, trying to remain standing among the crashing waves of the Atlantic Ocean. Was the later Parfit the same person as the child who had that experience? According to Locke, it is memory, or 'consciousness' of that experience that makes one the same person. Bishop Butler (1736) however, thought this a "wonderful mistake". It is, he wrote, "self-evident, that consciousness of personal identity presupposes, and therefore cannot constitute personal identity, any more than knowledge in any other case, can constitute truth, which it presupposes". Recall that in §78 Parfit already proposed a revised version of Locke's view. Thus, the Psychological Criterion does not apply to individual episodic memories, but to the continuity of memory, and more, broadly to Relation R, which also encompasses other kinds of psychological continuity. But this does not address Butler's objection (p. 219)

One interpretation of Butler's objection might be as follows: The concept of memory presupposes that we can remember only *our own* experiences. Therefore the continuity of memory presupposes personal identity. The same is true of Relation R. If we claim that personal identity just consists in Relation R obtaining, then this must be false if Relation R itself presupposes personal identity. In order to answer this objection Parfit introduces a wider concept which he calls *quasi-memory*. Thus, I have an accurate quasi-memory of some past experience if

- 1) I seem to remember having an experience,
- 2) *someone* did have this experience, and
- 3) my apparent memory is causally dependent, in the right kind of way, on that past experience.

According to this definition, ordinary memories are a species of quasi-memories because they are quasi-memories of our own past experiences. (p. 220)

We do not quasi-remember other people's past experiences, but perhaps one day we shall. Presently, we do not know the causes of long-term memory however they are hypothesised to reside in memory traces, also known as engrams. The actual method of storage in the brain, whether by biophysical, or biochemical means, is still being debated. (alleydog.com) It was once thought that memories were localised in only a few brain cells, however experiments by Karl S. Lashley in rats demonstrated that memory is diffusely distributed in the brain (Darryl, 2001; Sa *et al.*, 2015)

Suppose however that neurosurgeons one day develop a way to create a copy of a memory-trace from one brain in another brain. This would allow us to quasi-remember other people's past experiences. Consider Parfit's imaginary vignette:

Venetian Memories. Jane has agreed to have copied in her brain some of Paul's memory-traces. After she recovers consciousness in the post-surgery room, she has a new set of vivid apparent memories. She seems to remember walking on the marble paving of a square, hearing the flapping of flying pigeons and the cries of gulls, and seeing light sparkling on green water. One apparent memory is very clear. She seems to remember looking across the water to an island, where a white Palladian church stood out brilliantly against a dark thundercloud. (p. 220)

Jane knows that she has received copies of some of Paul's memory-traces, so what should she believe about these apparent memories? She knows that she has never been to Italy, while Paul has been to Venice often. She also knows about the Church of San Giorgio Maggiore in Venice which she has seen in photographs. She would probably, justifiably believe that she is quasi-remembering some of Paul's experiences in Venice. Parfit adds the following detail to his vignette:

Jane seems to remember seeing something extraordinary: a flash of lightning coming from the dark cloud, which forked and struck both the bell-tower of San Giorgio and the red funnel of a tug-boat passing by. She asks Paul whether he remembers seeing such an extraordinary event. He does, and he has kept the issue of the *Gazzettino* where it is reported.

Jane would, almost certainly, not dismiss her apparent memory as a delusion. She would conclude that she has an accurate quasi-memory of how the flash of lightning looked to Paul. (p. 220 - 221)

According to Parfit, for Jane's quasi-memories to give her knowledge about Paul's experiences, she must know roughly how they have been caused, although this is not required in the case of ordinary memories. Apart from this difference, quasi-memories would provide a similar kind of knowledge of what other people's lives were like, *from the inside*, as it were. When Jane seems to remember walking about the Piazza, hearing the gulls, and seeing the white church, she now knows part of what it was like for Paul, on that day in Venice. Of course, Jane's apparent memories are mistaken in one respect. She may seem to remember *seeing* the lightning as if she *herself* had seen it. Thus, her apparent memory may tell her accurately what Paul's experience was like, but it falsely tells her that it was *she* who had this experience. (p. 221)

Jane's apparent memories come to here in what Peacocke (1983) calls *the first-person mode of presentation*. Thus, when she seems to remember walking across the Piazza, she might also seem to remember a child running *towards her*. But even if these apparent memories are presented in the first-person mode, Jane need not assume that they are delusions or that they are her *own* experiences. Even if she seems to remember seeing the forked lightning, she could justifiably conclude that she is quasi-remembering one of Paul's experiences. Perhaps Jane remembers shaving "her" beard, while seeing Paul's face in the mirror, in which case it would be clear to her that this was not one of her own experiences. At other times she might have to work out whether it was she or Paul who had some past experience, and sometimes this might be impossible. She might, for example say, 'I do vividly seem to remember that tune, but I don't know whether it was Paul or I who heard it.'

We do not have such apparent memories, but then we do not have quasi-memories of other people's past experiences. Although our memories come to us in the first person they also come with the belief that, unless they are delusions, they are about our own experiences. In the case of experience-based memories, this is a sensible belief. If, like Jane however, we were used to having quasi-memories of other people's past experiences, we would cease to automatically assume this belief with the recall of each memory. (p. 221 - 222)

Returning to Butler's objection to the Psychological Criterion of personal identity: accordingly, the continuity of memory cannot be, even in part, what makes a series of experiences those of a single person, since memory presupposes a person's continued identity. On Parfit's earlier interpretation, memory presupposes identity because, on our concept of memory, we can remember only our own past experiences. Butler's objection can now be met using the wider concept of quasi-memory.

According to Parfit's revised Psychological Criterion, we cannot claim that just because I have an accurate quasi-memory of some past experience, that I am the one who had this experience. As in the case of Jane and Paul, one person's mental life might include some quasi-memories of other people's lives. Parfit's revised Psychological Criterion must be modified to include these quasi-memory connections. Instead we must appeal to overlapping chains of many such connections. Mental life includes countless quasi-memories of earlier experiences, the connections between which and earlier experiences overlap like strands of a rope. According to Parfit, there is *a strong connectedness* of quasi-memory if, over each day, the number of direct quasi-memory connections is at least half the number in most actual lives. Therefore, overlapping strands of strong

connectedness provide *continuity of quasi-memory*. Revising Locke's earlier criterion: we may claim that the unity of each person's life is, in part, created by this continuity. Since the continuity of quasi-memory does not presuppose personal identity, it may yet be part of what constitutes personal identity. There may however; be other kinds of psychological continuity besides. (p. 222)

Returning again to Butler's objection to the Psychological Criterion of personal identity: perhaps he meant some thing different. According to Parfit, he may have meant:

In memory we are directly aware of our own identity through time, and aware that this is a separate, further fact, which cannot just consist in physical and psychological continuity. We are aware that each of us is a persisting subject of experiences, a separately existing entity that is not our brain or body. And we are aware that our own continued existence is, simply, the continued existence of this subject of experiences.

Is this really so? Are we directly aware of the existence of this separate entity which is the subject of experiences, not just in memory? (p. 223)

81. The Subject of Experience

Scottish philosopher, Thomas Reid, in his *Essays on the Intellectual Powers of Man* (1785) wrote:

my personal identity... implies the continued existence of that indivisible thing that I call myself. Whatever this self may be, it is something which thinks, and deliberates, and resolves, and acts, and suffers. I am not thought, I am not action, I am not feeling; I am something that thinks, and acts, and suffers.

Taken one way, Reid's observation is clearly true. Even Reductionists admit that people exist. And on our everyday concept of a person, we are not just thoughts and acts. People are thinkers and agents. Nor are we just a series of experiences, but people who *have* experiences. A Reductionist can also admit to this, in a sense, that in ordinary parlance, a person is *what has* experiences, or the *subject of experiences*. However, what the same Reductionist would deny is that that the subject of experiences is a *separately existing entity*, distinct from the body, and a series of physical and mental events. (p. 223)

Parfit asks whether it is true, in memory, that we are directly aware of what the Reductionist denies? Could it be that each of us is aware that we are aware that we are persisting subjects of experiences, an entity separate from our body – a Cartesian Ego? According to Parfit, the question cannot be settled by argumentation, except to say what he believes to be the case and that he is not the exception. In Classic Text 02 and 06 we have argued that the Cartesian view is just plain wrong; even if the majority of people are dualists, if not Cartesian dualists. We can therefore gloss over the remainder of this section, and the next, and proceed to section 83. (p. 223)

83. Williams's Argument Against the Psychological Criterion

Parfit has defended the Psychological Criterion in two ways:

- 1) Psychological continuity can be described in such a way as not to presuppose personal identity, and
- 2) the harbinger of this continuity is not an entity that exists separately from a person's body.

Bernard Williams (1970) advanced another objection to the Psychological Criterion. According to Williams, if some person's brain continues to exist, and to support consciousness, this person will continue to exist, however great the breaks are in the psychological continuity of this person's mental life. Consider:

Williams' Example. I am the prisoner of some callous neuro-surgeon, who intends to disrupt my psychological continuity by tampering with my brain. I shall be conscious while he operates, and in pain². I therefore dread what is coming.

The surgeon tells me that, while I am in pain, he will do several things. He will first activate some neurones that will give me amnesia. I shall suddenly lose all of my memories of my life up to the start of my pain. Does this give me less reason to dread what is coming? Can I assume that, when the surgeon flips this switch, my pain will suddenly cease? Surely not. The pain might so occupy my mind that I would even fail to notice the loss of all these memories. The surgeon next tells me that, while I am still in pain, he will later flip another switch, that will cause me to believe that I am Napoleon, and will give me apparent memories of Napoleon's life. Can I assume that this will cause my pain to cease? The natural answer is again No. To support this answer, we can again suppose that my pain will prevent me from noticing anything. I shall not notice my coming to believe that I am Napoleon, and my acquiring a whole new set of apparent memories. When the surgeon flips this second switch, there will be no change at all in what I am conscious of. The changes will be purely dispositional. It will only become true that, if my pain ceased, so that I could think, I would answer the question 'Who are you?' with the name 'Napoleon'. Similarly, if my pain ceased, I would then start to have delusory apparent memories, such as those of reviewing the Imperial Guard, or of weeping with frustration at the catastrophe of 1812. If it is only such changes in my dispositions that would be brought about by the flipping of the second switch, I would have no reason to expect this to cause my pain to cease. The surgeon then tells me that, during my ordeal, he will later flip a third switch, that will change my character so that it becomes just like Napoleon's. Once again, I seem to have no reason to expect the flipping of this switch to end my pain. It might at most bring some relief, if Napoleon's character, compared with mine, involved more fortitude. (p. 229 - 230)

In the imagined case above, there is nothing that I am told that gives me reason to expect that, during my ordeal, I shall cease to exist. I have as much reason to dread all of the pain that is in store for me, and this reason is not removed by all the other things I have to dread such as losing my memories, going mad, becoming like and ultimately being deluded that I am Napoleon. According to Williams, this example shows that I have reason to fear future pain, no matter what psychological changes precede it. Even after all of these changes, it will be I who feels this pain. If so, the

² Brain surgery is, in fact, not painful. The scalp is numbed but the brain lacks pain receptors. Patients undergoing brain surgery are maintained in a conscious state so that they can report on various sensations. The neuro-surgeon also takes care to probe and map out areas that are responsible for important functions such as language, motor and sensory functions so that they can be preserved.

Psychological Criterion of personal identity is mistaken. Indeed, in this case, between now and after my ordeal, there will be no continuity of memory, character *etc.* Therefore what is involved my continued existence, cannot be such continuity. (p. 230)

It may be objected that, if I remain conscious throughout my ordeal, there will be at least one kind of psychological continuity. Though I would lose all memory of my past life, I would still have a chain of overlapping short-term memories of my ordeal – sometimes called the *specious present* or the duration in time in which one's perceptions are considered to be in the present. (Wikipedia: Specious present)

To plug this gap, we could add one feature to Williams' example: having lost all my other memories, I would be rendered unconscious and then reawakened, with *no* memories whatsoever. As my ordeal continues, I would have new memories but there would be no continuity of memory over my interval of unconsciousness. (p. 230)

According to Parfit, it may be further objected that he described Williams' example in question-begging terms. Recall that Parfit suggested that when he is made to lose his memory, he might, because of the pain, fail to notice the change. Such a description assumes that, after the loss of memories, the person in pain would still be the same person. Perhaps, on the contrary, at this point, the person will cease to exist and a new person come into being within the same body. Perhaps Williams would reply that even though Parfit's description assumes that he would continue to exist, this is the overwhelmingly plausible assumption. But it is the defender of the Psychological Criterion who must show that this assumption is not justified. According to Parfit, "... this would be hard to show. It is hard to believe that, if I was made to lose my memories while I was in agony, this would cause me to cease to exist half-way through the agony. And it is hard to believe that the change in my character would have this effect." (p. 230)

Instead, Williams' argument seems to refute the Psychological Criterion and show that the Physical Criterion is the correct one. According to this view, so long as a person's body continues to exist and support consciousness, the person will continue to exist, no matter what psychological discontinuities there may be in the person's mental life. (p. 230)

84. The Psychological Spectrum

Williams, above, discusses a single case in which, after a few changes, there is no psychological continuity. In what follows, Parfit revises Williams' argument so that there is a *spectrum*, or range of cases, each very similar to its neighbours. These cases, which make up what Parfit calls the *Psychological Spectrum*, involve all possible degrees of psychological connectedness.

In the case at the furthest end of the spectrum, the surgeon would flip all switches simultaneously so that there would be no psychological connection between me and the resulting person, who would be wholly like Napoleon. In the cases at the near end of the spectrum, the surgeon would flip only a few switches. If he flipped only the first switch, I would lose only a few memories and would have only a few apparent memories of Napoleon. If he flipped the first two switches, I would lose a few more memories and would have a few more apparent memories of Napoleon. Something similar would be true of my character. Flipping any particular switch would lead to a small change. Flipping

two switches may lead to my character being slightly more like Napoleon. *E.g.* I may become more bad-tempered and unperturbed by the sight of people being killed. (p. 231)

This revised version of Williams' argument involves many different cases, according to which we must decide which are those in which I would survive. At the closest end of the spectrum, the surgeon does nothing and I survive wholly in tact. In the second case, I would lose a few memories, acquire a few delusions and become slightly more bad-tempered; though I would survive. In the third case, the changes would be slightly greater. The same is true of any two neighbouring cases on this side of the spectrum. However, it is hard to believe that, for any two adjacent cases, I would survive on the one and cease to exist on the next. My continued existence cannot plausibly depend on whether I would lose just a few more genuine memories, acquire a few more delusory memories and undergo some further small change in character. If no such small changes could cause me to cease to exist, I would presumably continue to exist in all such cases, even at the far end of the spectrum, where between me now and the resulting person, there would be *no* psychological connections. (p. 231)

It may however be objected, that the form of this argument resembles that of the *Sorites Problem*, or the *Paradox of the Heap*. (See Critical Reasoning 04.) According to Parfit,

Suppose we claim that the removal of a single grain cannot change a heap of sand into something that is not a heap. Someone starts with a heap of sand, which he removes grain by grain. Our claim forces us to admit that, after every change, we still have a heap, even when the number of grains becomes three, two, and one. But we know that we have reached a false conclusion. One grain is not a heap. In [my] appeal to the Psychological Spectrum, [I] claim that no small change could cause [me] to cease to exist. By making enough small changes, the surgeon could cause the resulting person to be in no way psychologically connected with [me]. The argument forced [me] to conclude that the resulting person would be [me]. This conclusion may be just as false as the conclusion about the grain of sand. (p. 232)

But Parfit need not solve the Sorites Problem in order to defend this objection, when the following remarks may suffice:

When considering heaps, we realise that there are borderline cases. We may not know whether two, four, eight or sixteen grains of sand constitute a heap but this is not a result of ignorance. The concept of a heap is vague, with vague borderlines. Therefore when the Sorites Argument is applied to heaps we are happy to solve the problem by *stipulation*. We may make an arbitrary stipulation that the word 'heap' applies only to an assemblage of ten or more grains. By doing so we deprive the argument of one of its premises. Therefore according to our precise stipulation, the removal of the tenth last grain will result in a heap becoming something other than a heap. Nevertheless, this dismissal seems less plausible when applied to concepts such as phenomenal colour or personal identity. Most of us would agree that our continued existence is quite unlike the continuous existence of a heap. (p. 232)

Consider the range of cases involved in the Psychological Spectrum that are used to provide an argument against the Psychological Criterion. A Reductionist might claim:

The argument assumes that, in each of these cases, the resulting person either would or would not be me. This is not so. The resulting person would be me in the first few cases. In the last case he would not be me. In many of the intervening cases, neither answer would be true. I can always ask, 'Am I about to die? Will there be some person living who will be me?' But, in the cases in the middle of this Spectrum, there is no answer to this question. Though there is no answer to this question, I could know exactly what will happen. This question is, here, *empty*. In each of these cases I could know to what degree I would be psychologically connected with the resulting person. And I could know which particular connections would or would not hold. If I knew these facts, I would know everything. I can still ask whether the resulting person would be *me*, or would merely *be someone else* who is partly like me. In some cases, these are two different possibilities, one of which must be true. But, in *these* cases, these are not two different possibilities. They are merely two descriptions of the very same course of events. (p. 232 - 233)

The claims above are analogous to those we would accept about heaps. We are not committed to the belief that any assemblage of grains must either be a heap or not. There are borderline cases where there is no obvious answer to the question 'Is this still a heap?' Nor do we believe that there must *be* a Yes or No answer. In such cases, this is an empty question. We already know all that there is to know about such cases without answering the question. When applied to our own existence however we may not all be convinced. If I were to undergo an operation somewhere in the middle of spectrum, I could be sure that the resulting person would be in agony, but I would not know if it would be I who would be in agony or even if I shall still be alive.

Most of us, however believe that we are not like heaps, so it is very hard to dismiss such questions as empty. Most of us believe that, somehow, our identity must be determinate. Even in such "borderline cases" the question 'Am I about to die?' must have a definite answer, Yes or No. If someone will be alive and in agony, either this person will be me or it will not. It is difficult to make sense of any middle ground, such as the person in agony will be *partly* me. We can imagine someone in agony drifting in and out of consciousness, but that person, when fully conscious, cannot be only partly me, most of us believe. (p. 233)

According to Parfit, the Reductionist View provides an answer to Williams' argument, one that Williams rejects. Instead he concludes that, if my brain continues to exist, and to be the brain of a living person, I shall be that person. And this would be so even if there were *no* psychological connections between myself now and myself later. Although, Williams does concede that his conclusion may "perhaps" be wrong, in which case "... we need to be shown what is wrong with it". (Williams, 1973 p. 63)

85. The Physical Spectrum

One objection is that a similar argument applies to physical continuity. Consider again a range of possible cases along a *Physical Spectrum*, that involve different degrees of physical continuity. At the near end of the physical spectrum there would be a later person, fully continuous, both physically and psychologically, with me as I am now, just as in the case of normal continued existence. At the far end of the spectrum however, there would be a later person psychologically, but not physically

continuous with me as I am now. Cases such as Teletransportation lie at the far end of the physical spectrum. (p. 234)

Imagine the following cases along the physical spectrum: a case close to the near end of the spectrum in which scientists replace 1% of the cells in my body, including my brain, with exact clones. Somewhere in the middle of the spectrum they would replace 50%, and close to the far end they would replace 99%. At the furthest end of the spectrum, my body would be entirely destroyed and replaced by an exact Replica of me comprising of new organic matter.

The first few cases at the close end of the physical spectrum are already technically feasible. Portions of brain-tissue from one part of a mammal's brain have already been successfully transplanted to the same part of another mammal's brain of the same species. Similar technology could enable surgeons to provide functional replacements for parts of human brains that have been damaged due to disease or traumatic brain injury. Such brain tissue transplants have proved easier than transplants of more familiar organs, such as hearts and kidneys, because the brain's immune system does not reject them in the way that transplanted organs in the rest of the body are rejected. (See Moawad, 2020 on "What to Expect From a Brain Cell Transplant".) Although the cases at the near end of the physical spectrum can now be realised, most of the cases much further along the spectrum are now not possible and will probably remain so. However, their impossibility is merely "technical", to use Parfit's term. Since we are merely considering such cases to discover what we believe, their technical impossibility does not matter. (p. 234)

Suppose we believe that at the furthest end of the physical spectrum, my Replica would not be me, rather someone else who was exactly like me. At the nearest end of the spectrum, there would be no replacement of tissue so that the resulting person would be me. What about the intermediary cases? If 1% of my biological material were replaced, I would surely continue to exist because I do not require all of my body, including all of my brain, to exist. Indeed people regularly lose parts of their bodies in accidents or lose billions of neurons at a time due to concussion or substance abuse. However, what about 10%, or 30%, or 60%, or even 90% replacement? (p. 234 - 235)

According to Parfit, this range of cases challenges the Physical Criterion, which is one version of the Reductionist View. If you were about to undergo such an operation, you might believe this version of Reductionism. You might say to yourself:

In any central case in this range, the question 'Am I about to die?' has no answer. But I know just what will happen. A certain percentage of my brain and body will be replaced with exact duplicates of the existing cells. The resulting person will be psychologically continuous with me as I am now. This is all there is to know. I do not know whether the resulting person will be me, or will be someone else who is merely exactly like me. But this is not, here, a real question, which must have an answer. It does not describe two different possibilities, one of which must be true. It is here an empty question. There is not a real difference here between the resulting person's being *me*, and his being *someone else*. This is why, even though I do not know whether I am about to die, I know everything [about the case]. (p. 235)

For those who accept the Physical Criterion, this is the correct reaction to the range of cases above; however, most people would not accept such claims. For someone who insists that my Replica would not be me, they would have to conclude that there must be some critical percentage which is

such that by replacing less than such a percentage, it will be *me* who wakes up from the operation. However, by replacing more than such a percentage will result in *some other person*, who is merely like me. Alternatively, suppose that there is some crucial part of my brain, such that if it is not replaced, the resulting person would still be me, but if it were replaced the resulting person would be someone else. This is not a separate conclusion because we could ask, what if different percentages of this crucial part were replaced? Presumably, we would again be forced to conclude that there must be some critical percentage. (p. 235)

The above view is not incoherent, but it is hard to believe consistently. What makes it even harder to believe is this: we could not *discover* what the critical percentage is by carrying out sample cases along our imaginary physical spectrum. Suppose I say, 'Let's replace 50% of my cells and I will tell you what happens when I wake up from the operation.' We know in advance that in every case the resulting person will be inclined to believe that he is me, but that does not prove that he *is* me. Such an experiment could not yield an answer to our question. (p. 235)

Such considerations assume that all of a person's psychological features depend the state of the cells in his body, especially the nervous system. Therefore we can assume that an organic Replica of me would be psychologically exactly like me. If however we reject this assumption then we could respond to this range of cases in a different way, which Parfit discusses in the next section. If, on the other hand, our assumption is correct and all of these people would be exactly like me, we might believe one of three alternatives.

- 1) We could accept the Reductionist response above.
- 2) We could believe that there *is* a sharp boundary between adjacent cases such that if the surgeon replaced only certain cells, the resulting person would be me. If instead the surgeon replaced a few more cells, the resulting person would not be me, although he would be exactly like me. Even if there were such a sharp boundary somewhere along this range of cases, we could never discover just where it lies.
- 3) We could believe that in all of the cases, the resulting person would be me. (p. 236)

Most people would be disinclined to believe 3); however if we accept it we believe that psychological continuity bestows personal identity. We would believe this to be so even when this continuity does not have its normal cause, *i.e.* the continued existence of a particular body. Williams' argument however seemed to show that psychological continuity is not necessary for personal identity, when physical continuity would be sufficient. When we consider the Physical Spectrum, a similar argument seems to show that physical continuity is not necessary for personal identity, when psychological continuity would be sufficient.

According to Parfit, we could accept both of these conclusions, to wit that either continuity bestows personal identity. Although such a view would be coherent, it would invite serious objections. One objection arises if we combine, not both conclusions, but both arguments. (p. 236)

86. The Combined Spectrum

Now consider a range of possible cases that involve all possible variations in the degrees of *both* physical *and* psychological connectedness. Parfit calls this the *Combined Spectrum*. At the near end

of this spectrum is the normal case in which a future person would be fully continuous with me as I am now, both physically and psychologically. This person would be me in the same way that, in my actual life, it would be I who wakes up tomorrow. At the far end of this spectrum the resulting person would have no continuity with me as I am now, neither physically nor psychologically. A case at this end of the Combined Spectrum would involve the complete destruction of my body which is then replicated out of new organic matter. Suppose this person to be not Napoleon, but Greta Garbo. Suppose further that, when Garbo was 30, a group of scientists recorded the states of all of her cells in her body. (p. 236 - 237)

At the closest end of this spectrum nothing would be done. In the second case, just along, a few of my cells would be replaced with cells that are not *exact* duplicates of the originals. Therefore there would be somewhat less psychological connectedness between the person who wakes up from this operation and me. This person would also not share *all* of my memories, and his character would be slightly different to mine. He would also have some of Garbo's memories and one or two of Garbo's characteristics. He would enjoy acting, which I don't, and his eyes would be more like Garbo's. Further along the Combined Spectrum, a greater proportion of my cells would be replaced with dissimilar cells. The resulting person would be psychologically connected with me in fewer ways and in more ways connected with Garbo as she was at age 30. There would be similar changes in the resulting person's body. At the far end of the spectrum, most of my cells would be replaced with dissimilar cells so that the person who wakes up from the operation would have only a few of my original cells. This person would have only a few psychological connections with me. She would also have a few apparent memories that correspond to my past, but in every other way, she would be just like Garbo, both physically and psychologically. (p. 237)

According to Parfit, these cases provide a strong argument for the Reductionist View. Again the argument assumes that our psychological features depend on the states of our nervous system. In Classical Text 06 we dismissed the view of the Cartesian Ego, for *inter alia* not being able to explain the causal links between non-physical mental states and physical bodily states. Therefore the idea of a Cartesian Ego casts no light on the range of cases along the Combined Spectrum, which relies on a causal connection between mind and body.

Except for cases close to the near end of the Combined Spectrum, most of the rest of the cases are likely to remain technically impossible; therefore we shall not be able to directly discover whether the results would be as Parfit describes them. Instead, we must try to decide what we now believe about such cases. Recall that in the case of the first two Spectra, we had three alternatives: accepting the Reductionist position; believing that there must be some sharp borderline between different identities; and believing that the resulting person would be me in every case. Of these, the last seemed most implausible. (p. 238)

Considering the cases along the Combined Spectrum, we realise we cannot accept this last conclusion. At the furthest end of the Combined Spectrum surgeons destroy my body entirely and make a Replica of Garbo out of new organic matter. There would be no connection whatsoever between this new person and me. Indeed, the resulting person could *not* be me; therefore we are forced to choose between the other two alternatives above. (p. 238)

If we continue to insist that that identity must be determinate, we would believe that there must always be an answer to the question, 'Would the resulting person be me?' Yes or No. We would then be forced to accept the following claims:

Somewhere in this Spectrum, there is a sharp borderline. There must be some critical set of the cells replaced, and some critical degree of psychological change, which would make all the difference. If the surgeons replace slightly fewer than these cells, and produce one fewer psychological change, it will be me who wakes up. If they replace the few extra cells, and produce one more psychological change, I shall cease to exist, and the person waking up will be someone else. There must be such a pair of cases somewhere in this Spectrum, *even though there could never be any evidence where these cases are.* (p. 238 - 239)

These claims are hard to believe. Specifically, it is hard to believe 1) that the difference between life and death could consist of such minor discrepancies between adjacent cases. Most of us would be inclined to believe that there would *always* be a difference between some future person being me and his being someone else. And these differences would be profound, not trivial, as above. It is also hard to believe 2) that there must be a sharp borderline somewhere along the Spectrum, that we could never have evidence to discover. And if there could never be such evidence, it makes no sense to claim that there must be such a borderline. According to Parfit, even if 2) is true, 1) and 2) taken together are extremely implausible – so implausible that the Reductionist View is the only alternative. On this view of the cases in the middle of the Combined Spectrum, it is an empty question whether the resulting person would be me. (p. 239)

There are others who insist that our identity must be determinate, but that we are not separately existing entities, distinct from our bodies and our experiences. Parfit believes that this view is indefensible. How do we what explain this allegedly determinate personal identity? The answer must be that the true criterion of personal identity must cover every case. Whatever this criterion is, it must again draw a sharp borderline somewhere along the Combined Spectrum. But if we are not separately existing entities, how could there be such a borderline? We could stipulate that in one case, the resulting person would be me, and in the next he would not be me, but in what would the difference consist? (p. 239)

There are yet others who believe that, even though we are not separately existing entities, personal identity is some further fact, other than different kinds of physical and psychological continuity. But in what could this further fact consist? What could make this fact obtain or not obtain in the various cases in this range? (p. 239 - 240)

According to Parfit, the Combined Spectrum shows that certain views must be held together. We cannot believe that our identity involves some further fact, unless we also believe that we are separately existing entities, distinct from our bodies. Neither can we believe that that our identity must be determinate, unless we believe that these separate entities have an all-or-nothing existence. (p. 240)

Finally there are those who believe that the identity of *everything* must be determinate. Parfit calls this the strict form of the doctrine *no entity without identity*. According to this doctrine, we cannot name or even refer to a particular object, unless our criterion of identity for this object yields a definite answer in every conceivable case. On this view, we might mistakenly believe that we are

referring to some object, when there is no criterion of identity for such an object. *E.g.* we might mistakenly believe that the name 'France' refers to a nation, when nations cannot be referred to, because there is no strict criterion of identity for nations in every conceivable case. Therefore, nations must not exist. Those who accept this doctrine may believe that it could not be similarly true that persons do not exist. If this is so, and persons do exist, then the criterion of personal identity must provide a definite answer in every conceivable case.

This view does not imply that persons are separately existing entities, although it could make this view more plausible. However if we hold this view and agree that persons are not separately existing entities, then the criterion of personal identity must draw a sharp borderline, quite unwittingly, somewhere in the Combined Spectrum. And if, as Parfit has claimed, personal identity does not involve some further fact, then this view is even less plausible than the Reductionist view. (p. 240)

According to Parfit, there is another way in which some writers claim that our identity must be determinate. If, on this view, there are cases where we cannot answer a question about the identity of some object, we shall have inconsistent beliefs about its identity. There are indeed such cases, such as the status of a club, that may or not be the same club as the original club, or a collection of grains of sand, that may or may not be a heap. Their ontological status is simply indeterminate and any statement about their status would be neither true nor false; though it does not follow that such a claim must be incoherent. But suppose, for the sake of argument, that it were. This implies that when we find cases that are not covered by what we believe to be some strict criterion of identity, we should revise our beliefs about extending this criterion. When applied to cases somewhere in the middle of the Combined Spectrum, we do not believe that there must be some sharp borderline. Rather, we should *draw* such a line, in order to avoid incoherence. (p. 240)

This view hardly differs from the Reductionist View. If we decide to draw such a line, we should be mindful that is neither intrinsically nor morally significant – our choice will be arbitrary. Although we must draw this line somewhere between two adjacent cases, the difference between them would be trivial and should not affect our attitude towards these two cases. It would be irrational to regard the one case as being as good as ordinary survival and the next as bad as ordinary death. When we consider the range of cases, we may still wonder, 'Will the resulting person be me?' By drawing such a line, we choose to *give* an answer to this question. But since our choice is arbitrary, we cannot use our answer to justify any claim about what matters. If this is how we answer the question about my identity, then we have made it true *a fiat* that in this range of cases, personal identity is *not* what matters. And this, according to Parfit, is the most important claim in the Reductionist View. While Reductionists claim that, in some cases, questions about personal identity are indeterminate, Parfit recommends that we ought to give answers to such questions, even if we have to do so in way that is arbitrary and devoid of significance. While this "tidy-minded" version eliminates indeterminacy by uninteresting stipulation, the difference between the two versions is so slight that Parfit decides to ignore it. (p. 241)

According to the simplest version of Physicalism, every mental event is a physical event. Recall that Physicalists could also accept the *Psychological Criterion* of personal identity. Nor are Reductionists necessarily Physicalists, although almost all are. Those who are not Physicalists could be either Dualists, who believe that mental events are separate from physical events, or Idealists, who believe that all events are purely mental. If we believe that we are Cartesian Egos, then we believe in one

form of Dualism, but Dualists can also be Reductionists about personal identity. It is possible to believe that mental events are distinct from physical events, and that the unity of a person's life just consists of various connections that obtain between all the mental and physical events which, together, comprise of the person's life. This is the, almost never encountered, Dualistic version of the Reductionist View. (p. 241)

In the following chapter Parfit argues that if we are Reductionists, we should not try to decide between the different criteria of personal identity, because, *inter alia*, personal identity is not what matters. Before that however, he explores Reductionist claims further.

Reductionists agree that there is a difference between numerical identity and exact similarity. In some but not all cases, there would be a real difference between a person being me and that person being someone else who just happens to be exactly like me. Two clubs, for example, may exist at the same time, and apart from their members, be exactly alike. If I am a member of one of these clubs and you also claim to be a member, I might ask, 'Are you a member of the very same club of which I am a member? Or are you merely a member of the other club, that is exactly alike?' This is not an empty question because there are two possible answers. However, there may not be two different possibilities in the case that we are discussing the relationship between some present club and a past club. Recall from Section 79, that there was nothing that could justify either the claim that we have the very same club, or that we have a new club that happens to be exactly alike. In that case there would *not* be two possibilities. We could come up with a similar example for nations. (p. 241 - 242)

In the same way, there are some cases where there is a genuine difference between a person being me and that person being someone else who just happens to be exactly like me. Reconsider the Branch-Line Case of Teletransportation where the scanner does not destroy my body. In that case, my life on Earth overlaps with the life of my Replica on Mars. Given that there are two bodies whose lives overlap, we may conclude that we are qualitatively but not numerically identical. If I am the person on Earth and my Replica now exists on Mars, it makes a genuine difference whether I will feel pain, or whether it will instead be felt by my Replica. (p. 242)

If we return to the case of Simple Teletransportation, where there is no overlap between my life and that of my Replica, we would have a different scenario. We could say that my Replica will be me or we could instead say that he will be someone else exactly like me. But these are not mutually exclusive hypotheses, unless my continued existence involved some *further fact*. If my continued existence merely involves physical and psychological continuity, there will be some future person who will be physically exactly like me and who will be fully psychologically continuous with me. Because of the transmission of my blueprint, our psychological continuity will have a reliable cause, but it won't have its normal cause, because this future person will not be physically continuous with me. This is a full description of the relevant facts. There are no further relevant facts about which we are ignorant. If personal identity does not involve a further fact, we should not believe that there are two mutually exclusive differences. If there were, in what could these differences consist? (p. 242)

According to Parfit, some non-Reductionists would agree that in the above case, there are not two possibilities. These non-Reductionists believe that in the case of Teletransportation, my Replica would not be me. If we were wrong to say that my Replica is me, we could apply Parfit's reasoning to cases in the middle of the Physical Spectrum instead. We may decide that my Replica has a quarter

as many identical cells as me, or a half or three quarters. In such cases there are not two different possibilities: that my Replica is me, or that he is someone else who is like me. These would merely be different descriptions of the same outcome. If however, we believe that there is always a real difference between someone being me and his being someone else, we must believe that this difference occurs somewhere along this range of cases. Somewhere there would have to be a sharp borderline, though we could never tell where. But as Parfit has claimed, this belief is even more implausible than the Reductionist View. (p. 243)

In the case of clubs, there is sometimes, but not always, a difference between numerical identity and exact similarity. Sometimes, but not always, the question 'Is it the same club, or is it merely exactly similar?' is empty. This could be true of people too, either at the end or in the middle of the Physical Spectrum. But this is hard to believe. If I imagine myself about to press the green button, it is hard to believe that there is not a real question 'Am I about to die, or shall I instead wake up on Mars?'. But, as Parfit has argued, this belief cannot be justified unless personal identity involves a further relevant fact. And there is no such fact unless I am a separately existing entity, apart from my body. A Cartesian Ego would qualify as such an entity, but there is no empirical evidence for this and plenty of philosophical arguments against it. (p. 243)

Parfit is concerned that many, if not most, readers will not be convinced by his claims supported by his consideration of the Combined Spectrum. In the next section therefore, he advances other arguments for the Reductionist View.

12 Why Our Identity is not What matters

87. Divided Minds

Clinical Cases 2 - "Split-Brains"

In the 1950's and beyond Roger Sperry made pioneering discoveries in animals and later humans concerning the functional specialization of the cerebral hemispheres, for which he received the Nobel Prize in Physiology or Medicine in 1981. The early work on animals was published as [Sperry, 1964](#). In the 1960's Sperry was joined by Michael Gazzaniga and their joint work on split-brains in humans was summarised in [Gazzaniga, 1967](#). At the time only ten human patients had undergone surgery to sever their corpus callosum in order to treat intractable epilepsy. Of the ten, four volunteered to take part in Sperry and Gazzaniga's research. After the surgery the patient's personality, intelligence and emotions were unaffected; however the tests conducted by the researchers revealed that the patients demonstrated unusual mental abilities. The tests involved visual, tactile and auditory stimulation.

When a stimulus was presented to only one sensory field it was registered solely by the opposite cerebral hemisphere, with the hemisphere on the same side of the stimulus seemingly unaware of the stimulus. This is because, anatomically, motor and sensory nerves cross-over to the opposite side of the brain. If the hemisphere receiving the stimulus was also on the same side as the language processing areas of the patient's brain, the stimulus could be verbally described; otherwise it could be identified, but not described. Split-brain patients therefore appear to have two centres of consciousness, not one.

The most obvious features of the human brain are the massive, paired cerebral hemispheres that support a variety of higher level cognitive functions. The two hemispheres are connected by several bundles of fibre, the largest of which is the corpus callosum. Some patients with intractable epilepsy are treated by surgeons by severing these fibres, in order to reduce the severity of epileptic seizures, by confining their causes to a single hemisphere. These operations are generally successful but have the unintended consequence of creating “two separate spheres of consciousness”. (Sperry, 1966 p. 299)

This effect is revealed by various neuropsychological assessments. Due to the “crossing over” of nerve tracts in the brain, the right arm is controlled by the left hemisphere and *vice versa*; similarly the right halves of the visual field are represented in the left hemisphere and *vice versa*. When the corpus callosum is severed, information from the two hemispheres is no longer able to be shared with the other hemisphere directly. Thus, when such patients are presented with different information to different visual fields, they produce different answers to questions depending on whether they are written by their left or right hand. Parfit explains what a simplified assessment may entail:

One of these people is shown a wide screen, whose left half is red and right half is blue. On each half in a darker shade are the words, ‘How many colours can you see?’ With both hands the person writes, ‘Only one’. The words are now changed to read, ‘Which is the only colour that you can see?’ With one of his hands the person writes ‘Red’, with the other he writes ‘Blue’. (p. 245)

If such a person responds in this way, there is no reason to doubt that he is having visual experiences, and that he is seeing both red and blue. However, in seeing red he is not aware of seeing blue and *vice versa* – hence the term ‘two spheres of consciousness’. Such a person literally has two centres of consciousness that cannot directly communicate with one another because of the severed connection.

Language, including speech and comprehension, is lateralised, mostly to the left cerebral hemisphere. If a patient who has undergone a commissurotomy (severing of the corpus callosum) is given an unseen object in his right hand he will be able to identify it and name it because the right hand is controlled by the left hemisphere where the faculty of language resides. If however another unseen object is placed in his left hand, controlled by the right hemisphere, the tactile sensations of the object are correctly perceived but the object cannot be named, nor the sensations verbalised. Commissurotomy patients who have their language regions on the right cerebral hemisphere display the same, but reversed pattern. (Wikipedia: Split-brain)

After a certain amount of time each hemisphere can sometimes regain control of both hands. Thomas Nagel (1971 p. 153) describes the kind of conflict that can arise:

A pipe is placed out of sight in the patient’s left hand, and he is then asked to write with his left hand what he was holding. Very laboriously and heavily, the left hand writes the letters P and I. Then suddenly the writing speeds up and becomes lighter, the I is converted to an E, and the word is completed as PENCIL. Evidently the left hemisphere has made a guess based on the appearance of the first two letters, and has interfered... But then the right

hemisphere takes over control of the hand again, heavily crosses out the letters ENCIL, and draws a crude picture of a pipe.

Sometimes more sinister conflicts may arise. One commissurotomy patient complained that sometimes, when he embraced his wife, his left hand pushed her away. (p. 246)

There is another complication in actual cases. While the left hemisphere typically supports linguistic and mathematical abilities of an adult, the right hemisphere typically 'has' these abilities at the level of a young child. However the right hemisphere is more advanced in other respects such as musicality and pattern recognition. After the age of three or four, it is thought that the two hemispheres follow a model of 'division of labour', with each developing certain abilities. The lesser linguistic abilities on the right are not intrinsic or permanent. People who have had trauma to their left hemisphere often regress to the linguistic abilities of a child but can re-learn adult speech. Furthermore, a minority of people evince no evidence of laterality in the abilities of the two hemispheres. (p. 246)

Parfit imagines that he is one of this minority with exactly similar cerebral hemispheres. Suppose that he is equipped with a device that can block communication between his hemispheres and that this device is consciously controlled by him raising or lowering an eyebrow. By raising an eyebrow he can "divide" his mind by blocking communication between his hemispheres and "reunite" it by lowering the eyebrow. This ability would have many practical applications. *E.g.*

My Physics Exam. I am taking an exam, and have only fifteen minutes left in which to answer the last question. It occurs to me that there are two ways of tackling this question. I am unsure which is more likely to succeed. I therefore decide to divide my mind for ten minutes, to work in each half of my mind on one of the two calculations, and then to reunite my mind to write a fair copy of the best result. What shall I experience? When I disconnect my hemispheres, my stream of consciousness divides. But this division is not something that I experience. Each of my two streams of consciousness seems to have been straightforwardly continuous with my one stream of consciousness up to the moment of division. The only changes in each stream are the disappearance of half my visual field and the loss of sensation in, and control over, one of my arms. Consider my experiences in my 'right-handed' stream. I remember deciding that I would use my right hand to do the longer calculation. This I now begin. In working at this calculation I can see, from the movements of my left hand, that I am also working at the other. But I am not aware of working at the other. I might, in my right-handed stream, wonder how, in my left-handed stream, I am getting on. I could look and see. This would be just like looking to see how well my neighbour is doing, at the next desk. In my right-handed stream I would be equally unaware both of what my neighbour is now thinking and of what I am now thinking in my left-handed stream. Similar remarks apply to my experiences in my left-handed stream. My work is now over. I am about to reunite my mind. What should I, in each stream, expect? Simply that I shall suddenly seem to remember just having worked at two calculations, in working at each of which I was not aware of working at the other. This, I suggest, we can imagine. And, if my mind had been divided, my apparent memories would be correct. (p. 246 - 247)

In describing this case, Parfit assumes that there would be two separate streams of thoughts and sensations. Indeed, if his two hands visibly wrote out two calculations that he later remembered,

this is what we should assume. Moreover it would be implausible to assume that either of these calculations could have been worked out unconsciously because physics problems are paradigms of conscious problem solving.

It might be objected that Parfit's example above ignores the "the necessary unity of consciousness". But he has not ignored it, he has denied it. It is a fact that people with disconnected cerebral hemispheres have two streams of consciousness and two series of thoughts and experiences which the contralateral hemispheres are unaware of. Each of these streams of consciousness displays its own unity of consciousness. People's mental history need not always be like a canal, with only one channel, but sometimes more like a river, with occasionally separate streams. Therefore Parfit's example of what it would be like to separate and then reunify our minds is both coherent and imaginable. (p. 247)

It might also be objected that in his imagined case, Parfit does not have a divided mind, rather than two minds. But this objection does not raise a real question, since they are two ways of describing one and the same outcome. Similarly, it may be objected that the imagined result is not one person with a divided mind, but two people sharing one body, each solely controlling one arm and sensing one half of the visual field. But this too does not raise a real question, since they are again two ways of describing one and the same outcome. At any rate, this is how a Reductionists would understand it. (p. 247 - 248)

However not all people are Reductionists, so there may be some who believe it is a real question as to whether this case involves more than one person. Perhaps we could be persuaded so if the division were permanent. But this belief is very hard to accept in the imagined case of the Physics Exam. Recall that in that case there were two streams of consciousness for only 10 minutes and that later I remembered doing both calculations during that period when both of my hands could be seen writing out alternative solutions. Given that the period of disunity was so brief and modest, it is hard to believe that the case involved more than one person. It would be even harder to believe that during those 10 minutes, I ceased to exist and two new people came into existence, each of whom works on only one solution. On this interpretation, the whole episode involved three people, two of whom existed for only 10 minutes. Moreover each of these two fleeting people would have believed that they were me, together with apparent memories of my past. After these 10 minutes I would have acquired apparent memories of each of these two people, except that I would have mistakenly believed that I had all of the thoughts and sensations that these two people had. It stretches credulity to think that any person could be so mistaken, and that the episode did indeed involve three quite different people. (p. 248)

According to Parfit, it is equally hard to believe that the episode involves two different people, with me doing one calculation and some other person doing the other. When I first divide my mind, in doing the one calculation, I may believe that the other calculation is being done by someone else. But in doing the other calculation I may have the same belief. When I reunite my mind, I would then seem to remember believing that while doing the one calculation, the other calculation was being done by someone else. When I seem to remember these beliefs, there would be no reason to suppose that the one was true and the other false. After several rounds of division and reunification of my mind, I would cease to have such beliefs. In each of my streams of consciousness, I would

believe that I was also present in my other stream of consciousness, having thoughts and sensations of which in my present stream, I was now unaware. (p. 248)

References

BUTLER, J. (1736) *The Analogy of Religion*, first Appendix

DARRYL, B. (2001) Fifty Years Since Lashley's In Search of the Engram: Refutations and Conjectures. *Journal of the History of the Neurosciences* **10** (3): 308-318. doi:10.1076/jhin.10.3.308.9086

DENNETT, D. (1992) The Self as a Center of Narrative Gravity. In F. Kessel, P. Cole and D. Johnson, eds, *Self and Consciousness: Multiple Perspectives*. Erlbaum: Hillsdale, NJ. Danish translation, "Selvet som fortællingens tyngdepunkt" (1986) *Philosophia* **15**, 275-88

GAZZANIGA, M. (1967) The Split Brain in Man. *Scientific American* **217** (2):24-9

HUME, D. (1739–40) *A Treatise of Human Nature*. Clarendon Press: Oxford

LOCKE, J. (1695) *Essay Concerning Human Understanding*.

MOAWAD, H. (2020) *What to Expect From a Brain Cell Transplant*.

<https://www.verywellhealth.com/brain-transplant-4780507>

NAGEL, T. (1971) Brain Bisection and the Unity of Consciousness. *Synthese* **22**, reprinted in Nagel, 1979: *Mortal Questions*. p. 152 ff. Cambridge University Press

PARFIT, D (1984, 1987) *Reasons and Persons*. Ch. 10-13 p. 199-302. Clarendon Press, Oxford

PEACOCKE, C. (1983) *Sense and Content*. Clarendon Press, Oxford

REID T. (1785) *Essays on the Intellectual Powers of Man*.

SA, J. et al. (2015) Finding the Engram. *Nature reviews. Neuroscience* **16** (9):521-34

SPERRY, R. (1964) The Great Cerebral Commissure. *Scientific American* **210** (1):42-52

SPERRY, R. (1966) In J. C. Eccles, ed., *Brain and Conscious Experience*. Springer Verlag: Berlin

WILLIAMS, B. (1970) The Self and the Future, *Philosophical Review* **79**, (2): 161-180

WILLIAMS, B. (1973) *Problems of the Self*. Cambridge University Press

Note: An update to this study unit will be published in the coming weeks.