

Classic Text 32 – 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 a discussion of chapters 10 - 13 from part III 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. The latter can be downloaded for free [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 extensive 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 spaceship 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 Parfit's body and transmits a blueprint of him to Mars, where another machine assembles an exact *replica* of him out of atoms there. His replica will resemble him in every physical and psychological detail. Indeed, his replica will think that he is Parfit, and will have all his memories up to the moment he pressed the green button. If he returned to Earth, everyone Parfit knew would have believed he was him. Simple teleportation of this sort is a common device in science fiction. While some readers of this fiction would believe that his replica *would be him*, others would dispute this, believing that Parfit would have died when he pressed the green button and that his replica is, in fact, *someone else*, reconstructed to be exactly like him.

The second part of Parfit's fiction seems to support the latter view. If the new scanner successfully transmits his blueprint to Mars but does not destroy his body, merely damaging my heart instead, he would exit the cubicle as if nothing happened since he pressed the green button, except to learn that in a few days he will die... Later Parfit talks via a two way video link to his 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 Parfit is destroyed before being replicated, we are inclined to believe that this is a mode of speed of light transport, where his replica *is* him. In the second part of the story, the end of his life and the rest of his replica's life overlap. Parfit calls this the *branch-line case*, in which he believes that his replica *cannot* be him, since he is talking to him. Even though his replica is exactly like him, he feels that he is one person and his replica another. If Parfit were to pinch himself, his replica will not know. When he has his fatal heart attack, his replica will feel nothing. And when he is dead, his replica will enjoy another forty years of life on the *main line* of their briefly overlapping lives.

If we agree that Parfit's replica is not him, we may feel that his 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. His replica and he are **quantitatively identical**, having all their properties in common. But they may not be **numerically identical**, having the relation that everything has to itself and to nothing else, *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. A person might believe that after his marriage, he shall no longer be the same person; however marriage is not death. No matter how it might change him, he will still be the same living person who will *be* him. 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 neces-

sary. 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 most neurons are never replaced, but some are. (p. 203 - 204)

The standard view of physical continuity, of what makes an object one and the same over time, suggests that what makes someone the same person over time is that he has the same body (including brain) over time. Moreover, he 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 of personal identity**:

- 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 an 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 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 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 change, 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 did 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 psychological 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 particular 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; with 4) to 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 the climber is remembering his past experience. According to Parfit, it is a well-established fact that people can never remember their last few experiences before they are knocked unconscious. Thus Parfit's apparent memory of hearing his 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, he reconstructed this apparent memory based on what his companion later told him 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 Parfit's 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 him from the time he 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, Parfit's replica would *not* be him; however on both the wide and widest criteria, he *would* be.

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 per-

son 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 consists 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. Most of 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 properly understood, 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, Par-

Metaphysical Idealism

Idealism is the metaphysical doctrine according to which reality is indistinguishable from human consciousness, perception and understanding. In contrast to realists who regard the external world as primary, idealists regarded the world of objects and beings as a mental construct closely connected to ideas. In the past century idealism has been regarded in the West as profoundly wrong: The world appears as it does *not* because it is a projection of the mind but because macroscopic reality is objective and minds depend for their existence on the objective existence of their nervous systems in the world. Minds and their ideas are therefore derivative, not the primary building blocks of reality. In every organism studied so far, the direction of flow of perceptual information is from the external environment across a membrane into or onto the organism. Unfortunately there is no arguing with an idealist because all contrary evidence is interpreted by them as further evidence of reality as a mental construct. In this respect they resemble flat-earthers, who when shown a photograph of Earth from space respond: "See it is a flat disc!"

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.

fit's replica would be him. *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:

- 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 de-

scription 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 thing. 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? According to Parfit, while typed this paragraph, he was aware of the movements of his fingers, the sunlight on his desk and the rustling of leaves outside. What unites these experiences? Some claim that they are all *Parfit's* experiences. They are the experiences that were being had, at the time by a particular person, or *subject of experiences*. A similar question could be asked about a whole life: What unites the different experiences that, together, constitute this life? Again, some claim that what unites these experiences is they are all personal 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, Parfit is about to die. On the other he will 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 *him*. 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 Parfit's life briefly overlaps with that of his rep-

lica. Suppose that we believe that his replica and he are two different people, and that he is about to die but his replica will live for another forty years. If personal identity is what matters, then Parfit should regard his prospect here as nearly as bad as ordinary death. However, if what matters is relation R, with any cause, he should regard his 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 very 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 were 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 false if relation R itself presupposes personal identity. In order to answer this objection Parfit introduces a wider concept which he call *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, or both, is still being debated. 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 her 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 either. 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: 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, I 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' 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

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.

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 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.

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.

We continue in the first person: 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 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 a 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 case 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 the 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 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 present 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 him 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 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 reconstituted 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 all 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 Parfit’s body. In that case, his life on Earth overlaps with the life of his replica on Mars. Given that there are two bodies whose lives overlap, we may conclude that they are qualitatively but not numerically identical. If I were 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 Parfit’s life and that of his replica, we would have a different scenario. He could say that his replica will be him,

or we could instead say that he will be someone else exactly like him. But these are not mutually exclusive hypotheses, unless his continued existence involved some *further fact*. If Parfit's continued existence merely involves physical and psychological continuity, there will be some future person who will be physically exactly like him and who will be fully psychologically continuous with him. Because of the transmission of his blueprint, their 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 him. 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, Parfit's replica would not be him. If we were wrong to say that his replica is him, we could apply Parfit's reasoning to cases in the middle of the physical spectrum instead. We may decide that his replica has a quarter as many identical cells as him, or a half or three quarters. In such cases there are not two different possibilities: that my replica is him, or that he is someone else who is like him. 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 in favour of 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 other 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](#) (*sic*). 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 nerve fibres (**commissures**) the largest of which is the **corpus callosum** (from the Latin for "tough body"). 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 (**commissurotomy**), 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 commissurotomy patient 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 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 Parfit remembered doing both calculations during that period, when both of his 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, Parfit ceased to exist and that 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 him, together with apparent memories of his past. After these 10 minutes Parfit would have acquired apparent memories of each of these two people, except that he would have mistakenly believed that he 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 him doing one calculation and some other person doing the other. When he first divided his mind, in doing the one calculation, he may believe that the other calculation is being done by someone else. But in doing the other calculation he may have the same belief. When Parfit reunites his mind, he would then seem to remember believing that while doing the one calculation, the other calculation was being done by someone else. When he seems 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 his mind, he would cease to have such beliefs. In each of his streams of consciousness, he would believe that he was also present in his other stream of consciousness, having thoughts and sensations of which, in his present stream, he was now unaware. (p. 248)

88. What Explains the Unity of Consciousness?

Those who are not reductionists may believe that there is a true answer to the question 'Who has each stream of consciousness?' Suppose that for the above reasons, we believe that this case involves only one person: Parfit, and that for ten minutes he worked on the problem with a divided mind. (p. 248 - 249)

Recall next the view that psychological unity is explained by ownership. On this view, the unity of consciousness is explained by ascribing different experiences to this person or "subject of experiences". What unites these different experiences is that they are being had by the same person. This view is held both by those who believe that a person is a separately existing entity, and by those who deny this. This view also applies to the unity of each life. (p. 249)

Can we continue to accept this view when we consider the imagined case of the Physics exam? We may believe that, while Parfit's mind was divided, he had two separate series of experiences, and in having one, he was unaware of the other and *vice versa*. But what unites these different experiences? On the above view, the answer is that the experiences are being had by Parfit at this time. But this cannot be correct because he is not just having these experiences at this time, he is also having several other experiences in his other stream of consciousness. We need to explain the unity of consciousness within each of his streams of consciousness in each half of his divided mind. We cannot simply claim that all of these experiences are being had by him at the same time, making two

unities one. This ignores the fact that in having each of these two sets of experiences, he is unaware of the other. (p. 249)

If we insist that this unity should be explained by ascribing different experiences to a single subject, we must then believe that this case involves at least two different subjects. *I.e.* what unites the experiences in his left-hand stream of consciousness is that they are all had by one subject, and that what unites the experiences in his right-hand stream of consciousness is that they are all had by another subject. Therefore we must abandon the claim that the “subject of experiences” is the person. While Parfit’s mind is divided in its task, there are two different subjects of experiences. And since these are not the same subject of experiences, they cannot both be him. Since it is unlikely that Parfit is only one of the two, given the similarity of my two streams of consciousness, we might conclude that he is neither of these two subjects. The whole episode would then involve three entities: two of which cannot be claimed to be that with which we are familiar, namely a person. Even if we assume that he *is* one of the two subjects, the other cannot be him. (p. 249)

It seems we must now be sceptical. If we have to believe in subjects of experiences that are not persons, we may doubt whether they really exist. There are, of course, subjects of experiences in the animal kingdom, such as rats, that are not persons. But other species are irrelevant to this imagined case. On the above view, we are required to believe that the life of a *person* could involve two subjects of experiences that are not persons. (p. 250)

Reconsider Parfit’s experiences in his right-hand stream of consciousness. Perhaps in this stream, at a certain time, he is aware of thinking about part of the calculation, feeling writer’s cramp and hearing the scratchy sound of his neighbour’s old-fashioned pen. Surely we cannot explain the unity of these experiences as being had by the same subject of experiences which is *not* him. And if this entity is not a person, what kind of thing is it? It cannot be a Cartesian Ego, if Parfit claims to be such an Ego, since it is not him, and this case involves only one person. Perhaps this subject of experiences is some kind of Cartesian sub-ego – a purely mental entity which is merely part of a person? If we cannot explain the existence of a Cartesian Ego, then explaining the existence of a purely mental entity, which is merely part of a person’s Cartesian Ego is explaining the obscure by the even more obscure (*obscurum per obscurius*). (p. 250)

For someone who believes that unity is explained by ownership, even though they deny that we are separately existing entities, what unites their experiences, is that they are had by this person. But this cannot be the case. While Parfit is having one set of experiences in his right-hand stream of consciousness, he is also having another set in his left-hand stream. Therefore he cannot explain the unity of either set of experiences by claiming that these are the experiences he is having at this time because that would conflate both of these sets of experiences, *i.e.* treating the two as if they were one.

On the reductionist view, what unites Parfit’s experiences in his right-handed stream, at any time, is that there is a single state of awareness of these various experiences, including thinking about part of the calculation, feeling writer’s cramp and hearing the scratchy sound of his neighbour’s pen. Simultaneously there is another state of awareness in his left-hand stream. Parfit’s mind is divided because there is no single state of awareness of both of these sets of experiences. (p. 250)

It might be objected that this does not explain the unity of consciousness in each stream, so much as simply redescribe it. Parfit concedes that, in a sense, this is true; however this unity does not require a profound explanation. It is simply a fact that several experiences can be *co-conscious* of the objects of a single state of awareness. Compare this with the fact that there is a short-term memory of experiences within the last few moments a.k.a. the “specious present”. Just as there can be a single memory of just having had several experiences, such as hearing a bell toll three times, there can be a single state of awareness both of hearing the bell toll a fourth time and seeing ravens flying over the bell-tower. According to the Reductionist account, nothing more is involved in the unity of consciousness at any moment in time. Since there can be one state of awareness of several experiences, we do not need to explain this unity by ascribing them to the same person, or subject of experiences. (p. 250 - 251)

Parfit is at pains to restate other parts of his reductionist view:

Because we ascribe thoughts to thinkers, it is true that thinkers exist. But thinkers are not separately existing entities. The existence of a thinker just involves the existence of his brain and body, the doing of his deeds, the thinking of his thoughts, and the occurrence of certain other physical and mental events. We could therefore redescribe any person’s life in impersonal terms. In explaining the unity of this life, we need not claim that it is the life of a particular person. We could describe what, at different times, was thought and felt and observed and done, and how these various events were interrelated. Persons would be mentioned here only in the descriptions of the content of many thoughts, desires, memories, and so on. Persons need not be claimed to be the thinkers of any of these thoughts. (p. 251)

Parfit’s claims are supported by the fictional case of dividing his mind, as well as by actual cases of commissurotomies. The unity of different experiences is not in need of explanation, by ascribing them to him or these patients respectively. Nor can the unity of experiences be explained away in this fashion. There are thus two alternatives: we might ascribe the experiences in each stream to a subject of experiences which is not him (or anybody else), and hence not a person, or, if we doubt that such entities exist, we could fall back on the reductionist explanation. In such cases, at any rate, the latter should be the preferred explanation. (p. 251)

We can pass over the further discussion of the Cartesian view in the remainder of this section because, by now we assume, none of our readers takes the Cartesian view seriously.

89.What Happens When I Divide?

As another extension of an actual case of divided minds, Parfit supposes he were one of a pair of identical twins and that both his body (excluding his brain) and the brain (excluding the rest of his body) of his twin had been fatally injured. If neuro-surgery were to make sufficient advances, it will not be inevitable that both twins will die. Between the twins, they would still have one health brain and one healthy body (excluding the brain). In the future, surgeons might be able to patch these together.

While the one twin’s body could be kept alive with a heart-lung machine, the brain of the other could be extracted from his otherwise dead body and transplanted into the cranium of the other

twin. Though it may be feasible to reconnect blood vessels, it may never be feasible to splice all nerve endings together. The resulting body would thus be paralysed, but not brain-dead. If the resulting person regained consciousness after the operation they could be taught to communicate with others via a neural feedback device, perhaps by monitoring the activity of a nerve that once controlled some motor activity and by stimulating another nerve that provided some sensory input. According to Parfit, the stock example is that of a great scientist whose main aim in life is to continue thinking about certain abstract problems. (p. 253)

Suppose, for the sake of argument, that surgeons are able splice together nerves from the one twin's brain with those of the other twin's body. The resulting person would have no paralysis and would otherwise be healthy. Who would this person be? Perhaps there would be some disagreement between the physical and psychological criteria. Though the resulting person would be psychologically continuous with the brain donor, he will not have most of the brain donor's body. However the physical criterion does not require the continued existence of the whole body. If the first twin's brain continues to exist and be the brain of a living person who is psychologically continuous with that twin, then he continues to exist. This is true, no matter what happens to the rest of the first twin's body. Today, when someone successfully receives a donor heart, his identity is that of the surviving recipient, not the dead donor. If my brain is successfully transplanted into someone else's cranium, it may seem that I am the dead donor, whereas I am really the survivor and recipient of a new body. Receiving a new cranium and the rest of the body is just the limiting case of receiving multiple organ transplants. (p. 253)

In the case of identical twins, the match between recipient and donor would be nearly perfect; however if the new body were quite unlike the old, this might affect what they can do, and may indirectly lead to changes in their character. However, there is no reason to suppose that their psychological continuity would be compromised. According to Parfit, it has been objected that 'the possession of some sorts of character trait requires the possession of an appropriate sort of body'. Anthony Quinton has addressed this objection:

It would be odd for a six-year old girl to display the character of Winston Churchill, odd indeed to the point of outrageousness, but it is not utterly inconceivable. At first, no doubt, the girl's display of dogged endurance, a world-historical comprehensiveness of outlook, and so forth, would strike one as distasteful and pretentious in so young a child. But if she kept it up the impression would wear off. (Quinton, 1962)

According to Quinton, this objection might show that it matters whether a brain is housed in a certain *kind* of body, but not whether it was housed in any *particular* body. (p. 253 - 254)

On all versions of the psychological criterion, the resulting person in Parfit's imaginary case would be him. Most adherents to the physical criterion could also be persuaded that this would be the case. As Parfit has claimed, the physical criterion requires only the continued existence of enough of one's brain to be the brain of a living person, providing no one else has enough of this brain. Thus it would be the first twin who wakes up from the operation, and if the second twin's body were sufficiently alike, the first might even fail to notice that they had a new body. Since one hemisphere is, in fact, sufficient for survival, although certain abilities such as speech or the coordination of both hands may have to be relearned, receiving both hemispheres, according to the example, should provide the full range of abilities, without having to relearn them. (p. 254)

Now consider the possibility that Parfit could survive with only half of his brain in tact, the other half having been destroyed. Given the facts, it seems he could indeed survive if only half of his brain were successfully transplanted, and the other half destroyed. But what if the other hemisphere were *not* destroyed? This case was described by David Wiggins, in which he considers what it would be like for a person to divide like an amoeba. (1967 p. 50) Parfit simplifies the case by imagining that he is one of three identical triplets:

My Division. My body is fatally injured, as are the brains of my two brothers. My brain is divided, and each half is successfully transplanted into the body of one of my brothers. Each of the resulting people believes that he is me, seems to remember living my life, has my character, and is in every other way psychologically continuous with me. And he has a body that is very like mine. (p. 254 - 255)

This case is likely to remain impossible. Though it might be claimed that, in certain people, the two separated hemispheres would have the same range of abilities, this claim, as far as we know, is factually false, because several faculties are lateralised. Even though the cerebral hemispheres have actually been divided, it is unlikely that the mid-brain and hind-brain could be similarly divided without impairing their function. Finally, there is the small matter of splicing the nerves of the brain with those of the donor body. But these are merely “technical” impossibilities, not logically impossible. What matters is that that the division of a person’s consciousness has actually been realised (multiple times) and that considering this imaginary case may help us to decide both what we believe ourselves to be, and what, in fact, we are. (p. 255)

Parfit states, in advance, what he believes this case to show, *i.e.* it provides yet another argument against the view that we are separately existing entities. However, the main conclusion to be drawn is that *personal identity is not what matters*, even when our default belief is that our identity is what matters. Reconsider the branch-line case, where Parfit talks with his replica on Mars before he is about to die. If we believe that Parfit and his replica are different people, then it is natural to assume that his prospect is almost as bad as ordinary death. In a few days, his heart will fail and there will be no one living who will be Parfit. It is natural to assume that *this* is what matters, or at least this is what he assumes in discussing his division.

According to Parfit, if his double hemisphere transplant is successful, both of the resulting people will be psychologically continuous with him, as he is now, but what happens to Parfit? There are only four possibilities: 1) he does not survive; 2) he survives as one of two people; 3) he survive as the other person; or 4) he survives as both. (p. 255 - 256)

Contrary to 1), people have, in fact, survived with only one cerebral hemisphere, in spite of either being born with only one or the other being damaged due to disease or trauma. Therefore if one hemisphere of my brain were successfully transplanted, and the other hemisphere destroyed, I would survive. So how could I fail to survive if the other hemisphere were successfully transplanted also? Parfit asks rhetorically, “How could a double success be a failure?”

Considering 2) and 3): Perhaps he will be one of two resulting people. But if both hemispheres of his brain are exactly similar, for the sake of argument, so to start with, would be each resulting person. How then would it be impossible to survive as only one of two people? “What can make me one of them rather than the other?”

These three possibilities are not incoherent; indeed we can understand them. If we decide that identity is what matters, then 1) is not plausible. Division would not be as bad as death. Nor are 2) and 3) plausible. The fourth possibility remains, *i.e.* surviving as both of the resulting people. (p. 256)

How might we describe this possibility? We might claim, 'What we have called "the two resulting people" are not two people. They are one person with two bodies and a divided mind'. This claim is counterintuitive but cannot be dismissed outright. Recall that Parfit argued that we ought to admit as possible that a person could have a divided mind. In the imagined case of the physics exam there was only one person involved, two features of which made the case plausible. Firstly, the divided mind was reunited after only ten minutes, and secondly there was only one body involved. If however a mind were permanently divided, and each of its halves went on to develop in divergent ways, we would be less inclined to claim that the case only involves one person. (p. 256)

The case of complete division, where there are two bodies, appears to be a long way over the borderline. Perhaps after the operation the two "products" of division go on to live at opposite ends of the Earth. Suppose too that they have poor memories, and that their appearance changes in different ways. Perhaps, after many years, they might fail to recognise each other. According to Parfit, if such a pair were innocently playing tennis, we might have to claim, 'What you see out there is a single person, playing tennis with himself. In each half of his mind he mistakenly believes that he is playing tennis with someone else'. If we are not reductionists, we would believe that there is one true answer to the question whether or not these two tennis-players are a single person. Given what we mean by 'person', the answer must be No. It cannot be that what one partner sees behind the net, is, in fact, another part of himself. (p. 256 - 257)

If we admit that the two "products" are two different people, could we still claim that Parfit survives as both? He might say, 'I survive the operation as two different people. They can be different people, and yet be me, in the way in which the Pope's three crowns together form one crown'. This claim is also coherent but distorts our concept of a person. The Pope's three crowns, when put together, do form a fourth crown, but it is hard to imagine two people together as being a third person. Suppose the two resultant people fight a duel. How many people are fighting? One on each side, and one on both? And suppose a shot is fired and kills one of the resultant people: are these two acts? One murder and one suicide? And how many people are left alive? One or two? The composite third person, if he existed, would have no separate mental life. It is hard to believe that there could be such a third person. It is more plausible to treat the resulting people as a pair, and describe their relation to Parfit in a simpler way. (p. 257)

Parfit suggests other claims that may be made, such as that two resulting people are now different people, but that, before his division, they were the same person. But this suggestion is ambiguous. Perhaps the claim is that before his division, *together* they were him. On this account however, there were three different people even before his division. But this claim is even less plausible than the one rejected above. Alternatively, perhaps the claim is that the resulting people did not exist, as separate people, before his division. But if they didn't exist then, it could not be true that together they were Parfit.

Perhaps instead it is suggested that before his division, each of these people *was* Parfit, and after his division, neither is him, since he does not now exist. But if each of these people *was* Parfit, then whatever happened to him must have happened to them. If he did not survive his division, then nei-

ther did they. Since there *are* two resulting people, the case now involves *five* people: three before and two after. But this conclusion is absurd. What if we deny the assumption that leads to this conclusion? Can we claim that each of the resulting people *was* Parfit, but what happened to him did not happen to them? Before Parfit's division, it would be the case that each of the resulting people *is* him, but if what happens to him does not also happen to some person, that person cannot be him. (p. 257)

Suppose we appeal to **tensed identities** (identities that obtain between two continuant persons *at a given time*). If we call one of the resultant people *Lefty* we could ask, "Are *Lefty* and *Derek Parfit* names of one and the same person?" For those who believe in tensed identities, this is not a proper question because it fails to specify a time or time-interval. At any rate, such claims do not solve Parfit's problem. (p. 257 - 258)

Recall that Parfit discussed several implausible views about what happens when he divides. These include a single person, two, three, even five, all of which are too greater distortions of the concept of a person. Instead of considering further and even more implausible alternatives, Parfit therefore rejects the fourth possibility above, namely that he survives as both of the resulting people. But he has already rejected the other three possibilities: that he will be *one*, or *the other*, or *neither* of these people. But, as before, there is no way of knowing what happens, even if neurosurgeons succeed in performing the operation. If he did survive as one of the resulting people, he would believe that he had. But then he would know that the other resulting person falsely believes that he is Parfit, and that *he* survived. Knowing this, the first resulting person could not trust his own belief, because it might be he who has the false belief. And since both resulting people would both claim to be him, others would have no reason to believe either of them. So, even if this operation were to be carried out, we would learn nothing. (p. 258)

If there are not here four possibilities, each of which might happen, though we could never know which, then the reductionist view must be the correct one. Perhaps, when we know that each resultant person has half Parfit's brain, and would be continuous with him, we know all that there is to know about the matter. What are we supposing if we suggest that one of the resulting people might be him? What would make such an answer true? According to Parfit, there cannot be different possibilities, each of which might be the truth of the matter unless we are separately existing entities, such as Cartesian Egos. But we have already dismissed separately existing entities and Cartesian Egos, therefore we shall pass over Parfit's further objections to them. (p. 258)

If we embrace the reductionist view, these problems vanish. The claims proposed do not describe different possibilities, any of which *might* be true, and one of which *must* be true. Instead, these claims are merely different descriptions of the same outcome. And we know what this outcome is: There will be two future people, each with one of Parfit's brothers' body and psychologically continuous with him because they would have half of his brain. Knowing this, we know everything relevant about the situation. Therefore, persisting with the question, 'But shall I be one of these two people, or the other, or neither?' is an empty question. Recall the empty question about the reconstituted club, any answer to which could be *neither true nor false*. (p. 259 - 230)

Parfit distinguishes two ways in which a question may be empty. Some questions are both empty and have no answers. In such cases we might decide to *give* these questions answers; although it may be that any possible answer would be arbitrary. If so any answer we give might be pointless and

misleading. Recall the question, 'Shall I survive?' in the central cases in the combined spectrum. The same would be the case in other central cases of other spectra, in which there would be no survival at the far end of such spectra. But there are other cases in which a question may be empty. In such a case the question has an answer, in a sense, but the question itself is empty because it does not describe different possibilities, any of which might be true, and one of which must be true. Such a question merely gives us different descriptions of the same outcome. But if we decide to *give* it an answer, one description may be better than the others. If so, we can claim this description as the answer. In the case of Parfit's division, the best description is that neither of the resulting people will be him. (p. 260)

According to Parfit, in this case there are not different possibilities, so the important question is not, 'Which is the best description?' but 'What ought to matter to me? How ought I to regard the prospect of division? Should I regard it as like death, or as like survival?' When we have answered these questions, we may then decide whether we have given the best description. (p. 260)

Before discussing what matters, Parfit fulfils an earlier promise to address one objection to the psychological criterion, to wit that psychological continuity presupposes personal identity. Recall that in the case of memory, he appealed to the wider concept of quasi-memory. Thus Jane quasi-remembered having someone else's past experiences. Similarly, since at least one of the two resulting people will not be him, he can still quasi-remember living someone else's life. (p. 260)

In the earlier discussion, Parfit did not show that in describing the other relations that are involved in psychological continuity, we need not presuppose personal identity. But having described his division, this can now be shown. There is a direct relation that obtains between an intention and a later action in which the intention is carried out. It is surely the case that we can intend to perform only our own actions. But we can introduce the concept of a *quasi-intention* in which a person could quasi-intend to perform another's action, but this relation does not presuppose personal identity. (p. 260 - 261)

Parfit's case of Division, shows what this involves. Before the division he could have quasi-intended that one of the resulting people should roam the Earth, and the other stay at home. What he quasi-intended would not be done by him, but by the two resulting people. Normally, if a person intends that someone else should do something, they cannot make it so simply by forming such an intention. But if the person were to divide, it would be enough simply to form quasi-intentions. Both of the resulting people would inherit these quasi-intentions, and would carry them out, *ceteris paribus*. Since they might change their minds, there is no guarantee that they would carry out the quasi-intentions of the person before his division. The same is true of our own lives, since we can each change our own mind, there is no guarantee that we will ultimately carry out what we each intend. Indefinite procrastinators or vacillators come to mind. However, we do have some ability to control our futures by forming resolute intentions. And if I were to divide, I would have just as much ability to control the futures of the two resulting people by forming firm forming quasi-intentions. (p. 261)

According to Parfit,

Similar remarks apply to all of the other direct psychological connections, such as those involved in the continuity of character. All such connections hold between me and each of the

resulting people. Since at least one of these people cannot be me, none of these connections presupposes personal identity. (p. 261)

90. What Matters When I Divide?

What should our attitude be towards division? According to Parfit, some people would regard division as bad as, or nearly as bad as death; however for Parfit, this is irrational. Instead, we ought to regard division as about as good as ordinary survival. From the discussion above, the “products” of such an operation would be two different people. The relation between the person who divides and the products of the division is not missing any vital element that is contained in ordinary survival. The same relation would obtain if the original person stood in relation to only one of the resulting people. We have already seen that a person can, in fact, survive with only half a brain hemisphere in tact. Likewise, if it were possible, a person would survive the successful transplant of his brain into his twin brother’s cranium. Therefore, it would be possible for a person to survive the destruction of one of his hemispheres, the other one of which is subsequently transplanted into another’s body. In the case of division, the relation between the person undergoing division and each of the resulting people would contain everything required to survive as that person. Nothing is missing from the account of survival. If anything were to go wrong, it could only be the duplication. (p. 261)

Suppose that someone accepts this but still regards division as being almost as bad as death. There is no arguing with such a person. They are like someone who is told that taking a drug could double his years of life, yet still insists that taking the drug would be as bad as death. In the case of division, the extra years would run concurrently in both of the products, but this does not mean that there are *no* years to run. If such a person were to say, “But I shall lose my identity”, well, then there is more than one way of doing so: death is one, division is another. However double survival is not death; it is even less like death. (p. 261 - 262)

The problem with double survival is that it does not fit our logic of identity. Like several other reductionists, Parfit claims that,

Relation R is what matters. R is psychological connectedness and/or psychological continuity, with the right kind of cause.

And he claims that,

In an account of what matters, the right kind of cause could be any cause.

Other reductionists, however might insist that R should have a reliable cause, or its normal cause. But consider only cases where R would have its normal cause. In such cases, reductionists would accept that a future person will be R related to me as I am now, and no other person will be R related to me. If there is no such other person, then the fact that the future person will be me consists in relation R holding between us, nothing more. In nearly all actual cases, R takes a one-to-one form. When it does, such as between one present person and the future person, we can use the ordinary language of identity and claim that this future person will be this present person. (p. 262)

In the case of division, R takes a branching form. But since personal identity cannot take a branching form, I and the resulting two people cannot be the same person. And since I cannot be identical to

two people, and it would be incorrect to identify with arbitrarily one of the resulting people. It would be better to say that neither of the resulting people is me. But which is the most important relation: personal identity or relation R? In ordinary cases we do not need to decide because both relations overlap; however in the case of division the relations do not overlap, therefore we must decide which of the two matters. (p. 262)

If we believe that persons are separately existing entities, we could plausibly claim that identity is what matters; however we have already rejected this view. If we are reductionists, we cannot plausibly claim that of the two relations, identity is what matters. Therefore we must accept that the fact that personal identity just consists in relation R obtaining when it takes a non-branching form. (p. 262 - 263)

But besides relation R, Parfit also argued above that personal identity has one further feature, namely that relation R should obtain *uniquely* between one present person *and only* one future person. So rather than fall back on the claim that personal identity is what matters, Parfit augments the formulation of personal identity (PI) as relation R (R) plus the holding of a relation uniquely or in a one-to-one form (U). Thus,

$$PI = R + U$$

Most people are convinced that PI matters, or at least has some value. Assuming that R also has some value, then there are four possibilities:

- 1) R without U is of no value,
- 2) U enhances the value of R, but R still has value even without U,
- 3) U makes no difference to the value of R, or
- 4) U reduces the value of R but not enough to eliminate the value of R, since $R + U = PI$ still has some value.

According to Parfit, the presence or absence of U cannot make any difference to R. If someone is R-related to some future person, the presence or absence of U makes no difference to the intrinsic nature of the relation to this future person, and this is what matters most. So R without U would still have most of its value; therefore PI mostly matters because of the presence of R. Perhaps U might make some small difference, but this would be much less than the intrinsic value of R. (p. 263)

On its own, it would be difficult to accept that PI is not what matters, but when we consider the case of division, this difficulty vanishes. For Parfit, when we understand *why* neither resultant person will be me, we shall see, on reflection, that this does not matter, or matters only a little. (p. 263)

The case of division supports the reductionist view that our identity is not what matters, but does not imply that our identity is indeterminate. If we abandon the view that identity is what matters, then, if I divide, I shall die and neither of the resultant people will be me. However, according to Parfit, this way of dying is about as good as ordinary survival; although there may be room for some disagreements. Even if R is fundamentally what matters, U might make some small practical difference for better or for worse. The details however, do not change Parfit's line of argument in any rational or moral way. (p. 264)

Parfit concludes this section with a brief summary of his argument so far, before turning to a discussion of other writers' claims in the next section.

91. Why There is No Criterion of Identity That Can Meet Two Plausible Requirements

Besides Williams' argument, discussed in Section 83, he advances another against the psychological criterion. According to Williams, the criterion of personal identity must meet two requirements; however as Parfit shows, *no* plausible criterion of identity can meet both requirements. On the reductionist view, by contrast, analogous requirements can be met, providing further grounds for accepting this view. Williams, however does not assume the reductionist view; therefore in discussing the argument Parfit briefly sets aside this view for later. (p. 266)

Williams' argument develops out of a remark by Reid against Locke's claim that whoever "has the consciousness of present and past actions is the same person to whom they belong". According to Reid (1785), this implies "that if the same consciousness can be transferred from one intelligent being to another... then two or twenty intelligent beings may be the same person". Williams however argues that identity is a logically one-to-one relation, such that it is impossible for one person to be identical to more than one person. But we have seen that psychological continuity is not necessarily a one-to-one relation. Two future people could both be psychologically continuous with the person from which they divided, but they cannot both be the same person. Therefore psychological continuity cannot be the criterion of identity. Williams then claims that an acceptable criterion of identity must itself be a logically one-to-one relation, *i.e.* it must be a relation that could not *possibly* obtain between one person and the two future people resulting from the division of the first. (p. 267)

According to Parfit, some reply that this criterion might appeal to *non-branching* psychological continuity, one version of the criterion already discussed. On what Parfit calls the psychological criterion, a future person would be him if he were R-related to him. And since this criterion is a logically one-to-one relation, it has been claimed it would meet Williams' objection; however Williams rejects this answer. Instead he insists on the following requirements:

1. Whether a future person will be me must depend only on the *intrinsic* features of the relation between us. It cannot depend on what happens to *other* people.
2. Since personal identity has great significance, whether identity holds cannot depend on a trivial fact. (Williams, 1973 p. 20)

These requirements are plausible, but neither is met by the non-branching account of psychological continuity; therefore Williams rejects this version of the psychological criterion. Parfit believes that such an objection may be too abstract to be convincing; however its force can be shown by reimagining the case of simple teletransportation. Once the Scanner transmits a person's blueprint to Mars and destroys their body, the replicator on Mars will generate a perfect organic copy of the person on that side. The replica will believe that he is the same person and will, in every way, be psychologically continuous with the original person. (p. 267)

According to Parfit, suppose we accept the psychological criterion which appeals to relation R in its one-to-one form and also accept the wide version which allows R to have any reliable cause. This implies that the replica on Mars will be the original person. Suppose furthermore that an additional

blueprint is beamed to Jupiter's moon Io but that it is ignored by the scientists stationed there. If however the scientists on Io do make another replica of the original person, some time later, when that replica wakes up, the person on Mars will cease to exist [surely not!] Though the people on Mars would not notice any change, a new person will come into existence in the replica's body [surely not!] But Williams would object that whether the Martian replica continues to exist cannot depend on what happens to someone else millions of kilometres away in the Solar System on Io. Therefore Parfit's claim would violate Williams' requirement 1 above. (p. 267 - 268)

Recall that Parfit argued that what fundamentally matters is whether he would be R-related to at least one future person. But it is a trivial matter whether he should also be R-related to some other person. So on this version of the psychological criterion, whether he might be identical to some future person depends on this relatively trivial fact. Therefore Parfit's claim would violate Williams' requirement 2 above. (p. 268)

Williams might add that since teletransportation could produce many replicas of the original person, all of whom would be different people, we should deny that the first replica to wake up on Mars is the original person, even if only a single replica were made. If another replica were made, they could not also be the original person. So, if they could not both be the original person, even though they were produced in just the same way, we ought to conclude that neither would be the original person. But the original person's relation to one of the replicas is intrinsically the same, whether or not another was made. However, Williams might reply that since identity must depend on the intrinsic features of a relation, the original person could not be either replica, even if the other one weren't made. (p. 268)

For Williams, this argument supports a non-reductionistic version of the physical criterion – non-reductionistic because it assumes that personal identity is a further fact that requires (rather than consists in) physical continuity. According to Parfit however, Williams admits that a similar argument could challenge this view. Nevertheless, Williams rejects the psychological criterion because it involves a relation that can take a branching form that obtains between one person and two or more future people; thus failing to meet his two requirements above. According to Williams, physical continuity *could* take a branching form because, for example, it is possible to imagine a man dividing like an amoeba into two simulacra of himself. (Williams, 1973 p. 23) (p. 268)

Williams provides two answers to this objection. If we are in any doubt about the physical continuity of, say, a child and the adult he becomes, we may be interested whether this physical continuity took an abnormal branching form. If we knew the full history of the physical continuity such a body from childhood to adulthood, this would "inevitably reveal" whether there had been a case of amoeba-like division. However, the comparable claim for psychological continuity would not be so. We might know the full history of the psychological continuity between a person on Earth and his replica on Mars, yet fail to know about the other replica on Io. According to Parfit, "Branching is a problem for both the physical and the psychological criterion. But the problem is less serious for the physical criterion, since it would be in principle easier to know when the problem arises". (p. 268)

Williams makes the further claim that when a physical object divides, it is an intrinsic feature of its spatio-temporal continuity. However, when two people are psychologically continuous with an earlier person, this fact is not an intrinsic feature of either of these relations. Therefore, the physical criterion meets his requirement 1, whereas the psychological criterion does not. (p. 268 - 269)

Parfit's earlier imagined division provides objections to the physical criterion. Recall that he revised this in two ways. First he considered the case where his brain is transplanted into the body of his identical twin. It was clear then that he would be the surviving recipient and not the dead donor. If his surviving brain were to be given a new body, this would just be the limiting case of receiving new organs. The physical criterion ought only to appeal to the continuity of his brain. Moreover, in real cases, some people have survived with one of their hemispheres destroyed; therefore the physical criterion ought to appeal to the continuity, not of the whole brain, but enough of the brain to support conscious life. (p. 269)

But such continuity is not logically a one-to-one relationship. In Parfit's imagined case of division, each of the two resulting people would have enough of his brain to support conscious life. And we could not dismiss this case as impossible because the division of consciousness is a reality for commissurotomy patients. With our current technology, it not possible to divide the lower brain, but this is a technical rather than logical impossibility, in the way that teletransportation may never be possible. Such impossibilities, however do not weaken Williams' argument against the wide psychological Criterion, but if he appeals to such cases in that argument he cannot dismiss Parfit's imagined case of complete division. (p. 269)

In the latter case, Williams' argument implies that if Parfit's brain were to be divided and each half transplanted into one of his identical twins, Parfit would cease to exist and that the two resulting people would be new people. Williams must therefore revise Parfit's physical criterion so that it takes a non-branching form, perhaps along the following lines:

The Physical Criterion: If there will be a future person with enough of my brain to be the brain of a living person, this person will be me, unless there will also be someone else with enough of my brain.

However Williams would reject this criterion because it violates both of his requirements, above (p. 269)

Parfit provides a further example. Suppose that his division proceeds as follows: Parfit has two fatally brain damaged twins, Jack and Bill. A surgeon removes Parfit's brain and divides it accordingly. Each half is taken to a different wing of the hospital and prepared for transplantation into one of his brothers. The transplant of one half into Jack's body is successful; however before the other half can be transplanted into Bill's body it is dropped onto the concrete floor. If so, Parfit would wake up in Jack's body, but if the other half *had* been successfully transplanted, he would wake up in neither body. (p. 269)

But these claims violate Williams' requirement 1. Whether Parfit is the person who will awake in Jack's body ought only to depend on the intrinsic features of the relation between Parfit and this person. It cannot depend on what happens in the other wing of the hospital, just as it would be irrelevant whether the scientists on Io had or had not created a further replica of him. Whatever happens to Bill, the other half of Parfit's brain will not affect the relation between Parfit before the surgery and the person that will wake up in Jack's body. But this claim is denied by the physical criterion. Compared with the importance of the fact that half of Parfit's brain will survive in Jack's body, what happens to the other half is, for Parfit, relatively trivial; therefore this criterion also violates Williams' requirement 2. (p. 269 - 270)

Perhaps Williams might suggest

The New Physical Criterion: A future person will be me if and only if this person is both living and has *more than half* my brain. (cf. Wiggins, 1967 p. 55)

Since it is impossible for two future people to have more than half of a divided brain, the intrinsic feature of this relation can only take a one-to-one form; therefore it meets Williams' requirement 1. However, it fails to meet requirement 2. Parfit could be fully psychologically continuous with some future person *both* when the latter person has half of his brain *and* when this same person has slightly more than half. If what matters is physical continuity, then the difference between these cases, involving just a few more cells, is trivial. Therefore, the new physical criterion violates Williams' requirement 2. (p. 270)

Parfit introduces another objection to this criterion. Suppose a person suffers severe injuries to just over half of his brain. He would probably be paralysed on one side of his body and may have to be placed on life support; though his mental life may be unaffected. We would presumably believe that such a person survived his injuries; however on the new physical criterion we would have to claim that such a person has ceased to exist – that he is a new person, merely just like him. This is hard to believe.

According to Parfit, the physical criterion, in all its forms, faces strenuous objections. There are similar objections to the psychological criterion. However Williams' requirements, which seem plausible at first blush, cannot both be met on any plausible criterion of identity. If we were separately existing entities, like Cartesian Egos, both of Williams' criteria could be met, but we have already rejected the idea of such entities. (p. 270)

Parfit now returns to the reductionist view. In the case where half of his brain is successfully transplanted into Jack's body, the relation between him and the person waking up in Jack's body is one of psychological continuity, with its normal cause, and the continued existence of enough of his brain. Moreover as a reductionist, Parfit can claim that his relation to the person waking up in Jack's body comprises of what fundamentally matters, and this claim stands irrespective of what happens to other people elsewhere. With one revision, this claim meets Williams' requirement 1. Where Williams claims that whether Parfit shall be some future person depends only on his relation to this future person. Parfit instead asks whether his relation to this future person comprises of what matters. Like Williams, Parfit claims that the answer must depend only on the *intrinsic* features of this relation. (p. 271)

If the other operation were to have succeeded, someone would have awoken in Bill's body. But this does not change Parfit's relation between him and the person who awoke in Jack's body. Neither does it change the importance of this relation; since it still comprises of what fundamentally matters. But since this relation now exists in a branching form, we can't call each branch "personal identity"; therefore we are forced to give it a new *name*, even though the change of name bears no significance. (p. 271)

According to Parfit, this reductionist view also meets the analogue of Williams' requirement 2. Because judgements of identity are so obviously important, Williams claims that we should not make one such judgement and deny another without an important difference in our grounds for doing so.

On the reductionist view, what is important is relation R, *i.e.* psychological connectedness and/or continuity, with the right kind of cause. Unlike identity, this relation cannot fail to obtain because of a trivial difference concerning the facts. If this relation fails to obtain, it must be because there is a fundamental difference in the facts. This meets Williams' requirement 2. (p. 271)

In the case that Parfit divides, although we may not call his relation to the resulting people "identity", it does comprise of what fundamentally matters. But since his relation to the resulting people is about as good as identity, it probably carries most of the ordinary implications of identity. Thus, even though the person who wakes up in Jack's body may not be called 'Derek Parfit', he is just as deserving of punishment or reward for Parfit's actions. According to Wiggins (1976 p. 146): "a malefactor could scarcely evade responsibility by contriving his own fission". (p. 271)

This case raises some difficult questions: Suppose the malefactor were sentenced to twenty years in prison. Should each of the resulting persons serve twenty years, or only 10 years each? Parfit discusses some of these questions in his Chapter 15. However these questions do not detract from the general claim, *i.e.* "if we accept the reductionist view, it is R and not identity which is what matters". (p. 271 - 272)

According to Parfit, if this is the case, it may be thought that we ought to afford relation R the importance we now afford to personal identity; however this does not follow. If we believe that personal identity is more important, it may be because we still embrace a non-reductionist view. If however we instead change our view and embrace reductionism, we may also change our view about the relative importance of personal identity, to wit that relation R has nearly all the importance that, on the reductionist view, personal identity has. We may then accept that what fundamentally matters is relation R, not personal identity. Nevertheless, we may believe that both relation R and personal identity have much less importance than personal identity *would* have on the non-reductionist view. Parfit discusses this belief in his Chapters 14 and 15. (p. 272)

This belief does not change what Parfit has already claimed about Williams' requirements: If we assume that identity is what matters, neither of these requirements can be met. If we reject the non-reductionist view, our identity criterion could be either the physical or psychological criterion. However, recall that there is no plausible version of either criterion that meets both of Williams' requirements. Parfit adds that if we acknowledge that identity is not what fundamentally matters, we should not try to extend our criterion of identity so that it coincides more with what matters. As the case of division shows, any such coincidence can only be partial. Besides which, revising our criterion might then misleadingly suggest that it is again identity that matters. (p. 272)

Williams (1973, p. 24) offers one further reason why the criterion of identity should be logically one-to-one: "Unless there is some such requirement, I cannot see how one is to preserve and explain the evident truth that the concepts of identity and of exact similarity are different concepts". On the reductionist view however, we can see that these are indeed different concepts. Recall that on the branch-line case, where Parfit talks to his replica on Mars, two people can be exactly similar, but not numerically identical. The question, "Is he one and the same person as me, or is he merely another person, who is exactly similar?" is meaningful and one we can understand. In some cases, such as those in the middle of the physical spectrum, there is no real difference between a resulting person being me and him being someone else, just exactly like me. The reductionist view implies that, in some cases at least, there is no real difference between numerical identity and exact similarity. But it

also recognises other cases in which there is a real difference; therefore to paraphrase Williams, it preserves and explains the truth that these are different concepts. (p. 272)

In summary, Parfit's discussion of Williams' claims demonstrates that there is no "third option" between the reductionist view, on the one hand, and the idea that we are separately existing entities, on the other hand. This provides further grounds for accepting the reductionist view. Even if Williams' requirements are plausible, they cannot be met if we believe that identity is what matters; however on the Reductionist view where relation R is what matters, we *can* meet the analogous requirements.

92. Wittgenstein and Buddha

According to Parfit, Wittgenstein would have rejected the reductionist view. Wittgenstein believed that our concepts depend on certain facts obtaining, and that we should not entertain imaginary cases where such facts do not hold. However, as we have seen, the arguments for the reductionist view appeal to just such cases. But this disagreement is only partial. Most people do have beliefs about imaginary cases. Some of these beliefs imply that that we are separately existing entities, distinct from our bodies, and that existence must be all-or-nothing. According to reductionism, we ought to reject such beliefs, and Wittgenstein would have concurred. Given this agreement, Parfit feels that he is not obliged to discuss Wittgenstein, or similar views on this matter. (p. 273)

With two exceptions discussed below, Parfit is confident that he has discussed and considered all the points of view relevant to this debate. He admits that there may be other published views that, at the time, he may not have been aware of. Also, he has not considered historical views held at various times or by various civilizations. Obviously, Parfit intends for his claims be universal across all times and peoples. It would therefore be disturbing if he were to discover that they may merely be part of one cultural line of thought, say that of Modern Europe and America. (p. 273)

Fortunately, Parfit claims, this is not the case. When we ask what persons are, and how they continue to exist, the choice is between one of two views: According to one, we are separately existing entities, distinct from our bodies, and whose existence is all-or-nothing. The other is the reductionist view, which Parfit's arguments support. In Appendix J below (reproduced verbatim), Parfit sets out to show that the Buddha would have agreed. Therefore the reductionist view cannot be merely part of only one cultural tradition.

Appendix J: Buddha's View³

At the beginning of their conversation the king politely asks the monk his name, and receives the following reply: 'Sir, I am known as "Nagasena"; my fellows in the religious life address me as "Nagasena". Although my parents gave (me) the name "Nagasena"... it is just an appellation, a form of speech, a description, a conventional usage. "Nagasena" is only a name, for no person is found here.' (*The Milina Panha*: Collins, 1982 p. 182 - 183)

A sentient being does exist, you think, O Mara?

³ This section is reproduced verbatim from Parfit's Appendix J.

You are misled by a false conception.

This bundle of elements is void of Self,

In it there is no sentient being,

Just as a set of wooden parts,

Receives the name of carriage,

So do we give to elements,

The name of fancied being. (*Cila Mara*: Stcherbatsky, 1919 p. 839)

Buddha has spoken thus: 'O Brethren, actions do exist, and also their consequences, but the person that acts does not. There is no one to cast away this set of elements and no one to assume a new set of them. There exists no Individual, it is only a conventional name given to a set of elements.'

(*Vasubandhu*: Stcherbatsky, *Op. cit.*, p. 845.)

Vasubandhu: . . . When Buddha says, 'I myself was this teacher Sunetra', he means that his past and his present belong to one and the same lineage of momentary existences; he does not mean that the former elements did not disappear. Just as when we say 'this same fire which has been seen consuming that thing has reached this object', the fire is not the same, but overlooking this difference we indirectly call fire the continuity of its moments. (Stcherbatsky, *Op. cit.*, p. 851)

Vatsiputriya. If there is no Soul, who is it that remembers? *Vasubandhu*: What is the meaning of the word 'to remember'? *Vatsiputriya*. It means to grasp an object by memory. *Vasubandhu*. Is this 'grasping by memory' something different from memory? *Vatsiputriya*. It is an agent who acts through memory. *Vasubandhu*. The agency by which memory is produced we have just explained. The cause productive of a recollection is a suitable state of mind, nothing more. *Vatsiputriya*. But when we use the expression 'Caitra remembers', what does it mean? *Vasubandhu*. In the current of phenomena which is designated by the name Caitra, a recollection appears. (Stcherbatsky, *Op. cit.*, p. 853)

The Buddhist term for an individual, a term which is intended to suggest the difference between the Buddhist view and other theories, is *santana*, i.e. a 'stream'. (Collins, 1982 pp. 247-61)

Vatsiputriya. What is an actual, and what a nominal existence? *Vasubandhu*. If something exists by itself (as a separate element) it has an actual existence. But if something represents a combination (of such elements) it is a nominal existence. (Stcherbatsky, 1923 p. 26)

The mental and the material are really here,

But here there is no human being to be found.

For it is void and merely fashioned like a doll,

Just suffering piled up like grass and sticks. (The *Visuddhimagga*, Collins, 1982 p. 133)

Clinical Cases 3 – Ventromedial Prefrontal Cortex Lesions

Patients with damage to a part of the brain known as the **ventromedial prefrontal cortex (vmPFC)** located in the frontal lobe at the bottom of the cerebral hemispheres, have difficulty with tasks related to the **self-reference effect (SRE)**, the tendency for people to encode information differently depending on whether they are implicated in the information. Compared to controls, patients with injuries to the vmPFC have little or no ability to recall references to the self regardless of the context of time. In addition they have less confidence about their ability to possess traits compared to controls. This points to the central role of the vmPFC in the formation and maintenance of identity. Lesions to the vmPFC are also associated with altered personality, blunted emotion and executive function. Injury to this area is also associated with **confabulation**, *i.e.* false memories that are recounted with great confidence – confabulators are genuinely unaware that their stories are false. (Martone, 2022)

Whether we take a reductionist (Parfit) or even an eliminativist (Buddhism) view of the self, we cannot deny that we genuinely *do* have a *sense of self and identity*, and this it turns out to be a function of the vmPFC.

93. Am I Essentially My Brain?

Thomas Nagel, whom we met in Classic Text 23 on qualia, suggests a view that Parfit has not discussed so far, namely that what persons really are is the cause of their psychological continuity. According to Nagel, what persons really are are their bodies, and more specifically, essentially their brains. Nagel justifies this view in the form of two arguments that Parfit discusses in his Appendix D. In addition Nagel claims that his view is intuitively plausible. He writes of his brain that it,

seems to me to be something without which I could not survive – so that if a physically distinct replica of me were produced who was psychologically continuous with me though my brain had been destroyed, it would not be me and its survival would not be as good as my survival. (Nagel, 1986 p. 44)

Nagel's is a case very much like that of teletransportation; however he believes that teletransportation is not merely as good as survival but nearly as bad as death. After describing his imagined case he claims,

I will not survive the night... the replica will not be me. Trying to summon my courage, I prepare for the end.

According to Parfit, this suggests that personal identity is what matters, and that in the case of teletransportation, many people would accept Nagel's view, that what matters is the survival of their brains. Apart from Appendix D, Parfit provides a different sort of answer to this and similar views in Section 98 below.

94. Is the True View Believable?

According to Parfit, Nagel once claimed that even if the reductionist view were true, it would be psychologically impossible to believe it. In what follows, Parfit briefly reviews his arguments and

then asks whether he can honestly believe his own conclusions. If he can, he assumes he cannot be unique in believing so, *i.e.* there would be other people who can believe the truth of the reductionist view. (p. 274)

Recall Parfit's claim that a person is not like a Cartesian Ego, a being whose existence is all-or-nothing. Rather a person is more like a nation – an account of their identity over time would be similar in their essential features. Parfit also distinguished two views about the nature of persons: On the non-reductionist view, persons are separately existing entities, distinct from their bodies and experiences. The best known example of this view is the Cartesian Ego. On the reductionist view, persons do exist and are distinct from the bodies and experiences but they are not separately existing entities. The existence of a person over any time-interval just consists of the existence of the body, the thinking of its thoughts, the doing of its deeds, and the occurrence of many other physical and mental events.

Since these opposing views differ about the nature of persons, they also disagree about the nature of personal identity over time. According to the reductionist view, personal identity just involves physical and psychological continuity, both of which can be described in an impersonal way, without reference to any experiences had by a person. Reductionists also claim that personal identity is not what matters. What does matter are certain one-to-one relationships of connectedness and continuity that are involved in personal identity. For the non-reductionist however, personal identity is what matters. Over and above physical and psychological continuity, identity is a separate, further fact, that either obtains fully or fails to obtain at all. Psychological unity is explained by ownership, while the unity of consciousness at any time is explained by the fact that several experiences are being had by a person at the time. The unity of a person's life is explained in the same way. According to Parfit, these claims must either stand or fall together. (p. 275)

Parfit concedes that the non-reductionist view might have been true, if for example there were compelling evidence for reincarnation; however there is none, but instead much evidence against it. Some evidence is supplied by actual cases of patients who have undergone commissurotomies. Such patients have two parallel streams of consciousness, each unaware of the other. We might claim that that in such patients there exist two different people in the same body, much like the case of division discussed above and reviewed below. For Parfit however, it is more plausible to claim that there is a single person with two streams of consciousness. (p. 275 - 276)

If so, how do we account for the unity of consciousness within each stream? We cannot explain this by claiming that the different experiences in each stream are had by the same person or subject, because this conflates the two streams as one. According to Parfit, if we believe that the unity of consciousness is explained by ascribing different experiences to a particular subject, then in commissurotomy patients, although there is a single person, there must be two subjects of experience. It follows that in the life of a person, subjects of experience are *not* persons. This is hard to believe, but perhaps such cases are better explained by the reductionist psychological criterion. Accordingly, at any time, there is one state of awareness of experiences in one stream of consciousness and another state of awareness in the other stream. (p. 276)

Although the existence of commissurotomy cases militates against the non-reductionist view, they comprise of only a small part of the evidence against such a view. There is, moreover, no evidence that the harbinger of psychological continuity is an entity, such as a Cartesian Ego, that must be all-

or-nothing. On the other hand, there is every reason to believe that the harbinger of this continuity is the body, especially the brain. Equally, there is every reason to believe that our psychological features depend on bodily states, especially brain states. And a brain's continued existence need not be all-or-nothing. Physical connectedness can be a matter of degree, and there are numerous actual cases in which psychological connectedness obtains only certain ways, or to a reduced degree. (p. 276)

Therefore, we have sufficient evidence to reject the non-reductionist view and embrace the reductionist view as the only alternative. Parfit did consider other possible alternatives but concluded that none was both non-reductionist and for which there are sufficient reasons to accept it. Although these other views differ in other ways, the plausible ones do not deny reductionism's central claim, to wit that we are not separately existing entities, distinct from our bodies and experiences, and whose existence is all-or-nothing. (p. 276)

Besides appealing to the facts, Parfit made several claims about our beliefs by appealing to our intuitions about several imaginary cases. In one such case, a neurosurgeon gradually erases his psychological continuity by tampering with his brain. Parfit described three other ranges of cases: In the psychological and physical spectra there would be, between him and some future person, all possible degrees of either psychological or physical connectedness. In the combined spectrum there would be a matrix of all possible degrees of both kinds of connectedness. (p. 276 - 277)

According to Parfit, most of us would be strongly inclined to believe that any future person would be either ourself or someone else, and that there would always be a profound difference between such outcomes.

In the same range of imaginary cases, Parfit supposes that he is about to undergo one of the operations that occupy the middle of the spectrum. Between the person before and after the operation there would be certain kinds and degrees of physical and psychological connectedness. On the reductionist view, knowing all the facts about the operation exhausts the full truth of what will happen. Suppose that before losing consciousness Parfit asks, 'Am I about to die? Or shall I be the resulting person?' Parfit believes that there must always be two possibilities, one of which must be the truth. On the reductionist view, this is an empty question. Sometimes it is the case that there is a real difference between some future person's being him, and he being someone else. But for cases occupying the middle of the combined spectrum there would be no real difference. What difference could there be? Nothing could make it the case that either the resulting person would be him or that he would be someone else. Because persons are not separately existing entities, there is nothing that could make either of these differences true. We could say that the resulting person would be him or we could say that he would die and become someone else, but these are simply different descriptions of the same outcome. (p. 277)

By way of illustration, Parfit appealed to Hume's comparison of persons with nations, clubs or political parties. If we consider these other entities, most of us would take a reductionist view. If a political party split to become two rival parties we might ask, 'Did the original party cease to be, or did it continue to exist as one or other of the resulting parties?' But this is not a real question about different possibilities, one of which must have happened. Instead this is an empty question. Even if we have no answer to this question, we could still know all that there is to know about what happened. (p. 277 - 278)

Since we can make sense of and accept the reductionist view about political parties, clubs and nations, we can make sense of what is being claimed by the reductionist view of persons; although most are inclined to reject this view. Most would insist that there must always be a difference between some future person being him and him being someone else. Parfit is somewhat sceptical that merely reflecting on the combine spectrum may be sufficiently persuasive on its own; therefore he provides further arguments. (p. 278)

One scenario he imagines is division. Commissurotomy patients represent actual cases of the division of one stream of consciousness into two; therefore Parfit's scenario is a natural extension of such actual cases. In the imagined case, each half of his brain is successfully transplanted into another body. We might ask, so what happens to him? The only possible answers are that he will be one of the resulting people, or the other, or neither. If we believe that identity is what matters, then each of these answers is difficult to accept. Given that, in the original scenario, both of the recipient bodies were identical twin brothers, it is hard to believe that he will be *one* of these two people. If, however he is neither of the two people, and identity is what matters, then we ought to regard his imaginary division as equivalent to death. But this is also hard to believe, especially since his relation to each resulting person contains everything needed for survival. Strictly speaking, his relation between each of the resulting people cannot be called identity because it obtains between one and *two* resulting people. In the case of death this relation holds between one and *no* future person. So, although double survival cannot be described in terms of identity, it cannot be equivalent to death. Numerically, two does not equal zero. (p. 278)

The imagined case of division endorses another aspect of the reductionist view. Not only are the possible answers above hard to believe, it is hard to conceive of different possibilities, any of which might be true. If we are not Cartesian Egos, what could make it the case that, if we divided, we would be one or the other of the resulting people? Or if there were other possibilities, in what could the difference consist? Although each resulting person would have half the original person's brain and be fully psychologically continuous with him, it seems there is no right answer to this question, even if we had a full description of all the events. This is yet another empty question, nor are there any different possibilities which could be true. These are merely different descriptions of the same outcome. (p. 278 - 279)

According to Parfit, the best description is that he shall be neither of the resulting persons. But this does not imply that division is as bad, or nearly as bad as death, when it should be regarded as about as good as ordinary survival, *ceteris paribus*. Although Parfit would not be the same person as one of the resulting people, his relation to each of these comprises of what fundamentally matters in the case of ordinary survival; therefore identity cannot be what fundamentally matters. What matters is what Parfit has called relation R *i.e.* psychological connectedness and/or psychological continuity, with the right kind of cause. (p. 279)

Having reviewed the main arguments in favour of the reductionist view, Parfit now asks whether he finds it impossible to believe. On an intellectual or reflective level he is convinced, but he admits that on some other level he will always have some doubts. Specifically, his belief is strongest when considering some of his imagined cases. He is convinced that if he divided, it would be an empty question whether he would be one, or the other, or neither of the resulting people, and that there could be nothing that could make these different possibilities. Similarly, he is convinced that in the central

cases of the combined spectrum, it would be an empty question whether the resulting person would be him. (p. 279)

Concerning other cases, he admits his conviction is less firm. In the example of teletransportation for one, Parfit imagines that he is in the cubicle about to press the green button, but then has doubts and changes his mind. He admits that reviewing his arguments could never wholly remove all doubts. Parfit compares this to looking through a window at the top of a skyscraper. Although he knows he is in no danger, looking down from this dizzying height makes him afraid, and he would have similar irrational fears if he were about to push the green button. (p. 279)

As to Nagel's claim that it is psychologically impossible to believe the reductionist view, well Parfit for one does, and presumably others do too. Similarly the Buddha claimed that, although it is very hard to do so, it is possible. Parfit claims that his remaining doubts or fears seem to be irrational: "We can believe the truth about ourselves", he says. (p. 280)

13 What Does Matter

95. Liberation from the Self

Parfit's sustained arguments above show that the truth about the self and personal identity is not what we are inclined to believe. But he asks, "Is the truth depressing?" and responds with his most moving and oft quoted passage:

Some may find it so. But I find it liberating, and consoling. When I believed that my existence was a such a further fact, I seemed imprisoned in myself. My life seemed like a glass tunnel, through which I was moving faster every year, and at the end of which there was darkness. When I changed my view, the walls of my glass tunnel disappeared. I now live in the open air. There is still a difference between my life and the lives of other people. But the difference is less. Other people are closer. I am less concerned about the rest of my own life, and more concerned about the lives of others. (p. 281)

Parfit also claims that when he belied in the non-reductionist view, he was more concerned about his inevitable death. Since his death in 2017 there has been no one living who is him, but this fact can be redescribed. Although subsequently there have been many experiences, none of them is connected to his experiences while he was alive by chains of such direct connections as those involved in experience-memory, or in the carrying out of an earlier intention. Some of the present experiences may be connected to those while he was alive in less direct ways. There are some present memories about his life, and there are others that were influenced by him, or things done as a result of his advice while alive. While Parfit's death has severed the link between his more direct experiences and those that obtain now, it has not extinguished various other relations. According to Parfit, "This is all there is to the fact that there will be no one living who will be me. Now that I have seen this, my death seems to me less bad". (p. 281)

So, instead of the live Parfit saying, "I shall be dead", he equally said, "There will be no future experiences that will be related, in certain ways, to these present experiences"; a redescription of the fact that he found less depressing. This is similar to the person undergoing some ordeal like one of the

operations described above. Instead of saying, “The person suffering will be me” he could equally have said, “There will be suffering that will be related, in certain ways, to these present experiences”; again a redescription of the fact that seems less bad. (p. 281 - 282)

When Hume thought about his radical sceptical arguments he recalls that he was thrown into “the most deplorable condition imaginable, environed with the deepest darkness”. The cure for Hume was to dine and play backgammon with his friends. Parfit’s arguments for reductionism had the opposite effect, removing the “glass wall” between him and others and caring less about his death. According to him, “This is merely the fact that, after a certain time, none of the experiences that will occur will be related, in certain ways, to my present experiences. Can this matter all that much?” (p. 282)

96. The Continuity of the Body

Parfit admits that he is glad that the reductionist view is true, but that this is simply a report of psychological effects, which may be different in others.

But there are still other questions to be addressed that require more than a report of his reactions. Instead, the answers to these questions depend on the force of certain arguments. First he addresses what, as reductionists, we ought to claim to be what matters. Then he asks how, if we have changed our view about the nature of personal identity, we ought to change our beliefs about rationality, and about morality.

Recall that the case of Parfit’s Division shows that personal identity is not what matters; it just so happens that, in most cases, personal identity does coincide with what matters. Parfit asks, “What does matter in the way in which personal identity is, mistakenly, thought to matter? What is it rational to care about, in our concern about our own future?” (p. 282)

According to Parfit, this question can be restated. Assume that, for simplicity’s sake, that it could be rational to be concerned only about one’s self-interest. Suppose that Parfit were such an egotist and that he could be related in one of several ways to some resulting person. What relation to this resulting person would justify his egocentric concern? If the rest of this person’s life were worth living, in what way would he wish to be related to him? Alternatively, if the rest of this person’s life were much worse than nothing, how should he *not* wish to be related to him? In short, what sort of relation, for an egoist, should fundamentally matter? Presumably, this relation should also be, for all of us, what fundamentally matters, in our concern for our own future. But since we may be concerned for the resulting person, *whatever* his relation to us, it is probably clearer to ask, what should matter for an egoist. Here are Parfit’s simplest answers:

- 1) Physical continuity,
- 2) Relation R with its normal cause,
- 3) R with any reliable cause,
- 4) R with any cause.

Recall that R is psychological connectedness and/or continuity, with the right kind of cause; therefore if we decide that R is what matters we must consider what is the relative importance of connectedness and continuity. If it is suggested that *both* R *and* physical continuity are what matter,

then this is the same as 2) above, since physical continuity is part of R's normal cause. However, is 1) defensible? Can we claim that, if Parfit will be physically continuous with some resulting person, then is this what matters, even if he will not be R-related to this person? (p. 282 - 283)

In Williams' earlier example, the surgeon destroys any kind of psychological continuity. Suppose that the surgeon were about to operate on Parfit, without pain, such that the resulting person will have a life that is much worse than nothing. If he were an egotist, Parfit might regard this prospect as being no worse than a painless death, since he would not care what happened to the resulting person. Alternatively he might regard this prospect as much worse than death because he would be egoistically concerned about this person's dreadful future. What should his attitude be? (p. 283)

Well, he ought to be concerned about this person's future if he could justifiably believe that this person will be *him*, rather than *someone else*, who is merely physically continuous with him. But Parfit has already argued that this belief is not justified. Recall that Williams' example lies as the far end of the psychological spectrum where, together with cases in the centre of the spectrum, there is no real difference between the resulting person's being him, and his being someone else. In Williams' example, the full facts are as follows: The resulting person will be physically but not psychologically continuous with Parfit. It doesn't matter whether we call him Parfit, or someone else. On the wide psychological criterion of personal identity, we could call him someone else; however both of these descriptions could not be factually wrong because they are both descriptions of the same fact. If we resort to one of these descriptions in order to justify some view about what matters, it might turn out to be a bad description because it may imply something unjustifiable about what matters. Therefore we must decide on what matters *before* we choose one description over another. (p. 283 - 284)

If Parfit were to accept these claims, then, as a reductionist, we might ask, should he be egotistically concerned about this future of this person? If he knew that physical continuity could not make it true, as a further fact, that this person would be him. Should he be concerned? In deciding what matters, we cannot appeal to notions of identity, because identity here is an empty question. Instead, we must ask whether physical identity *ipso facto* justifies egoistic concern. According to Parfit, the answer is No. Recall that those who believe in the physical criterion cannot demand whole body continuity. It does not matter if one receives a transplanted organ, so long as the transplant functions just as well or better. All that could matter is that at least enough of one's brain continues to exist.

Parfit proposes that the reason that the brain is singled out in this way is because the brain is the harbinger of psychological continuity, or relation R, and if it weren't it would be no more important than the continuity of any other part of the body. If relation R did not obtain, the continuity of the brain would have no significance for the person whose brain it is, and certainly would not justify egocentric concern. Reductionists cannot plausibly claim that physical continuity is all that matters. At most, they can defend 2) above, to wit that relation R would not matter if it did not have its normal cause, part of which is physical continuity. (p. 284)

But Parfit believes that 2) above is also indefensible because he believes that physical continuity is the least important factor of a person's continued existence. What we value about ourselves and of others, is not the continued existence of the particular brains and bodies, but the various relations between ourselves and others whom we love and what we love about them and ourselves. Some of us of course, may want ourselves or others to have bodies very similar to our present bodies, but

this is not the same as wanting the continuity the same particular bodies. According to Parfit, if there will be some person R-related to him as he is now, it hardly matters whether that person will have his present body, including his present brain. What fundamentally matters is relation R, with or without its normal cause. Therefore it ought not to matter whether his brain were to be replaced by an exact duplicate. (p. 284 - 285)

Of course, if someone were to be R-related to Parfit, or any other person, their body should be sufficiently alike to allow for full psychological connectedness. This should be especially important if this body were of the opposite sex, and almost inconceivable, if it belonged to another species. For some people who are exceptionally beautiful or possess one or more extraordinary athletic abilities there should be exact physical similarity. Parfit passes over these demands of exact physical similarity for the remainder of the discussion.

According to Parfit, "Whether we accept this view may affect our beliefs and attitudes about our own lives"; however the imagined case of teletransportation is quite clear. On Parfit's view, his relation to his replica comprises of what fundamentally matters, being about as good as ordinary survival. From the non-reductionist standpoint however, ordinary survival is little better than, or about as bad as, being destroyed and replicated. For Parfit then, it would be irrational to pay much more to be physically transported to Mars by a spacecraft. However, many people Parfit included, would harbour some fears about teletransportation. But since this is an imaginary example, we know exactly what will happen and we cannot fear that the worse of two outcomes will occur.

According to Parfit, his relation to his replica is R without its normal cause; the abnormality of the case seemingly trivial. Reconsider the case of artificial eyes implanted to restore sight to those gone blind. If these eyes gave recipients visual sensations just like those of normally sighted individuals, and if these eyes provided their recipients true beliefs about what can be seen, then this would be as good as normal sight. It would not be rational to reject these synthetic eyes just because they are not the normal cause of human sight. There may be some grounds for disliking them if they appeared unsightly to others; however the analogy does not carry over to the case of teletransportation. Parfit's replica would be just like him in every detail; indeed, he would have a normal body with a normal brain. (p. 285)

Perhaps we should not call Parfit's replica by his name. If we decide to do so, we should regard this as a case of division, where there is at best an answer to an empty question. If one is about to divide, it is probably best to say that neither resulting person will be me, but this does not imply that division is like death. Since we also know what will happen with division, it is not as if we are deciding between outcomes; we will merely be choosing one of several descriptions for a single outcome. The same is true for teletransportation. Given that one would already be in possession of all the relevant facts beforehand, one's attitude to its outcome should not be swayed but what to call one's replica. If Parfit were to decide not to call his replica me, the fact

- a) that my replica will not be me would just consist in the fact
- b) that there will not be physical continuity, and
- c) that, because this is so, R will not have its normal cause.

According to Parfit, since a) would just consist in b) and c), we can ignore a). Parfit's attitude should therefore depend on the importance of facts b) and c), and these facts are all there are to his replica not being him. (p. 285 - 286)

But even if we acknowledge that c) is true, Parfit is convinced we cannot rationally believe that it matters much, *i.e.* that the cause is not normal. Rather it is the *effect* that matters and that this effect, *i.e.* Relation R, is *ipso facto* the same. If it happens that this effect has a cause that is not normal, we can simply describe it in a different way. We could say that, although my replica is psychologically continuous with me, he will not be me. But this is not a further difference in what happens, beyond a difference in the nature of the cause. If we were to opt for the more expensive journey via spacecraft, we could not rationally appeal to the abnormality of the cause.

Similar considerations apply to the continued existence of one's present body, including one's brain. It surely is not unreasonable to want one's replica to be like one's present body, but this is a desire for a certain kind of body, not a desire for the same particular body. Does it really matter that *this* very brain gets to Mars? As before, perhaps the natural fear is that this is the only way to ensure that I shall get to Mars; however this assumes that whether or not I get to Mars is a real, rather than an empty question. Even if this question had a best answer, we shall know in advance exactly what will happen before deciding what the answer is. So, according to Parfit, it might be rational to care a little whether or not this body will be his present body, including his brain, but it would be irrational to care a great deal about it. The reason it might not be irrational to care a little could be similar to the reason that one may wish to keep the same wedding band, rather than identical band forged from the same metals. Most of us would probably have a sentimental wish to keep the very band that was involved in our wedding ceremony. Similarly, it might not be irrational to mildly prefer that the person on Mars should have my present body. (p. 286)

Finally for this section, there remains the question of whether, if some person were R-related to me, it would matter whether this relation did or did not have a reliable cause. According to Parfit, there is an obvious reason why we might, in advance, prefer that the cause will be reliable. If teletransportation were unreliable and only worked in a few cases it would be rational to pay the much larger fare for the journey via spacecraft, but this not the point. We should rather ask, "In the few cases, where my replica will be fully R-related to me, would it matter that R did not have a reliable cause?" Again the answer should be No. If there were an unreliable treatment for some disease, that in most cases achieves nothing but results in a cure for a few cases, only the outcome matters. In other words, only the effect matters and this effect is just as good, even though its cause were unreliable. According to Parfit, we should claim the same about Relation R, namely that we should accept 4) above, and that in our concern for our own future, "*what fundamentally matters is relation R, with any cause*". [Original emphasis] (p. 286 - 287)

97. The Branch-Line Case

Here Parfit returns to the Branch-Line Case that he discussed in the opening pages of Part III of *Reasons and Persons*. Recall that he regards teletransportation as about as good as ordinary survival. However, recall that he introduced a challenge in which a new scanner successfully transmits his blueprint to Mars, but in doing so damages his heart, so that he could be expected to die within, at most, a few days. Recall also, that in the time remaining Parfit uses the Intercom to converse with

his replica on Mars. He reassures Parfit that he will continue his life where he leaves off. What should Parfit's attitude have been, given that he was about to die? Does his relationship to his replica comprise of what matters? Even though his replica is fully psychologically continuous with him before the malfunction that damaged his heart, he asks again if this relation is about as good as survival. According to Parfit, "It may be hard to believe that it is. But it is also hard to believe that it can matter much whether my life briefly overlaps with the life of my Replica." (p. 287)

But consider the following:

The Sleeping Pill. Certain actual sleeping pills cause retrograde amnesia. It can be true that, if I take such a pill, I shall remain awake for an hour, but after my night's sleep I shall have no memories of the second half of this hour. I have in fact taken such pills, and found out what the results are like. Suppose that I took such a pill nearly an hour ago. The person who wakes up in my bed tomorrow will not be psychologically continuous with me as I am now. He will be psychologically continuous with me as I was half an hour ago. I am now on a psychological branch-line, which will end soon when I fall asleep. During this half-hour, I am psychologically continuous with myself in the past. But I am not now psychologically continuous with myself in the future. I shall never later remember what I do or think or feel during this half-hour. This means that, in some respects, my relation to myself tomorrow is like a relation to another person. Suppose, for instance, that I have been worrying about some practical question. I now see the solution. Since it is clear what I should do, I form a firm intention. In the rest of my life, it would be enough to form this intention. But, when I am on this psychological branch-line, this is not enough. I shall not later remember what I have now decided, and I shall not wake up with the intention that I have now formed, I must therefore communicate with myself tomorrow as if I was communicating with someone else. I must write myself a letter, describing my decision, and my new intention. I must then place this letter where I am bound to notice it tomorrow. I do not in fact have any memories of making such a decision, and writing such a letter. But I did once find such a letter underneath my razor. (p. 287 - 288)

This case is similar to the branch-line case in one way. Indeed, it would be just like a variant of the case in which Parfit lived for a few days after leaving the cubical, with the creation of his replica postponed until after his death. However, In the original case Parfit's life overlaps with that of his replica: recall them talking on the Intercom. But in the original case, there is no analogue to the sleeping pill case.

There is however an analogue in the imagined case of the physics exam. Recall that in that case Parfit divides his mind for ten minutes, being aware that, in both streams of consciousness, he is having thoughts and sensations in the other stream, but not knowing what those other thoughts and sensations are. His relation to himself in each of the other streams is similar to his relation to another person in which he would have to communicate in some overt way. Perhaps he would write a letter to himself in the other stream, placing it in his other hand. (p. 288)

This situation is very much like the branch-line case. We can imagine our minds divided; therefore we need not assume that Parfit's replica on Mars is someone else. Here on Earth, he would not be aware of what his replica on Mars is thinking. This is similar to each of his two streams of consciousness in the physics exam. Neither would he be aware of what, in the other stream, he is thinking. Similarly, Parfit could say that in the branch-line case, he has two streams of consciousness, one

here on Earth, the other on Mars. When he talks to his replica on Mars it is much the same as the communication between his two different streams of consciousness in the physics exam. (p. 288)

The actual, rather the imagined, case of the sleeping pill provides a close analogue to one of the special features of the branch-line case, namely that Parfit is on a psychological branch-line. The imagined physics exam case provides another close analogy to the other special feature, namely that his life overlaps, for a time, with that of his replica. Taken together, they support the case that, when on the branch-line, his relation to his replica contains everything that matters. If the extent of the overlap were large enough, this *would* make a difference.

Suppose now that Parfit were an old man, about to die, but who will be outlived by someone who was once his Replica. When this person came into existence, say, some 40 years ago, he would have been psychologically continuous with Parfit, as he was then, but that he has since lived his own life for 40 years. Old man Parfit's relation to his erstwhile Replica, though better than death, would not be nearly as good as ordinary survival. However, had only ten days or ten minutes elapsed, then this relation would be just about as good. Parfit cites Nozick (1981, p. 44) who argues that overlaps as brief as this cannot be rationally thought to have much significance.

Parfit believes that his two analogies above are enough to defend this claim, but that they may be hard to believe. *Before* he pressed the green button, it would have been easier for him to believe that his relation to his replica contains what fundamentally matters in ordinary survival. He could have looked ahead along the main line towards 40 years of life ahead. *After* he pressed the green button and talked to his replica however, he could not have done so. Instead he would have had to redirect his concern back along the branch-line, beyond the point of division, and forward again along the main line. Parfit admits that this psychological manoeuvre would be difficult, but that it would not militate against what he has argued about this case. (p. 289)

98. Series-Persons

To revise, Parfit denies that personal identity is what matters. Instead he believes that what fundamentally matters, in our concern for our own future, is relation R obtaining, with any cause, and this would be what matters, even when it does not coincide with personal identity. On Nagel's view however, what I am is essentially my brain, and what fundamentally matters is the continued existence of my brain. Parfit addresses Nagel's arguments in Appendix D, but he is not convinced that they may be persuasive. Parfit therefore explains how both Nagel's view (in a revised form) and his own view could both be true. (p. 289)

Supposing Nagel's view is true, then if one is essentially one's brain, one cannot decide to take a different view of oneself, but one can do something else. Nagel (1986 p. 45) discusses the concept of a **series-person**: while a person is essentially a particular embodied brain, a series-person is potentially an R-related series of embodied brains. But there is a problem in that our bodies, including our brains, age and decay. Nagel, however imagines a community in which technology provides a solution. In this community, everyone over 30 enters a scanning replicator once a year. The machine scans and destroys a person's body and creates a replica who is R-related to the person and who's body is almost identical, except that it has not aged or decayed. According to Nagel, the series-persons in this community would not consider it irrational to use such a scanning replicator. On their

criterion of identity, each series-person would continue to exist in a new body each year. Indeed, each series-person would continue to have a body (and brain) with their, youth, appearance and vigour that they had at age 30. (p. 290)

Of course, some of these series-persons would have fatal accidents; therefore Parfit takes the liberty of introducing a detail. As a precaution, each series-person would have a blueprint made of their body every day. With this precaution, series-persons would be potentially immortal and ever youthful. According to most cosmologists however, the Universe will either expand forever ending in a “heat-death” or expand at an accelerating rate and end with space-time itself being ripped apart, the so called “big-rip”, or reverse its expansion and collapse into a singularity, “the big crunch”. Therefore, so long as we are composed of matter, it would be impossible for even series-persons to live literally forever, though they could live for potentially billions of years.

Parfit assumes that even if he were to accept Nagel’s view, he may not be able to change his view, but he could do something else. Although his manuscript was typed up in November 1982, this sentence tells you that, in the rest of his book, *pronouns are used to refer to series-persons*. Even if Nagel’s view is false, this would not change what the pronouns mean. Every person may be a series-person and this would still be so even if our criterion of identity were relation R, with any cause, since this would coincide with the criterion of identity of series-persons.

If Nagel’s view were true however, the rest of Parfit’s book simply uses pronouns in a new sense. Thus ‘I’ and ‘me’ no longer refer to the person Derek Parfit, but to the series-person who’s present body (and brain) is also Derek Parfit’s body (and brain). Since the pronouns ‘I’ and ‘me’ would be used in their new sense, their old sense could be expressed as ‘Old-I’ and ‘old-me’. Similarly for other pronouns. (p. 290)

Parfit asks, what would the relation be between his old-me and his series-person. Recall the mythical Phoenix. Ordinarily, on the criterion of identity of birds, a bird ceases to be if it is burned to ashes. A Phoenix however, would not be a particular bird. It would be a series of birds, or a *series-bird*, even though a Phoenix at any time would be the body of a particular bird. When it burns to ashes however, it is animated again as the body of a new bird, rising from the ashes. Just like a series-person, a particular Phoenix would exist as a series of different bodies. (p. 290 - 291)

Of course, there never was a Phoenix, but there are many series-persons. According to Parfit, the sentences he was typing would have been typed by a series-person, ‘me’ in the new sense. But they were also being typed by a person, the ‘old-me’. This person is named Derek Parfit, but the series-person he dubs *Phoenix Parfit*. Since his body then was also Derek Parfit’s body, both he and his series-person were typing those sentences, and both persons were having the very same thoughts and experiences. So even if these two persons were intimately related in this way, *if* Nagel’s view is true, then they would be two different individuals. The difference is this: On Nagel’s view, if the Old-I were teletransported, this would kill the old-me, but not the series-person. Thus, the old-me and me would be different individuals.

It is of course doubtful whether this series-person exists just because of the invention of a new concept. Given what is meant by “series-person”, I (Parfit), the series-person, started to exist just at the same time as the Old-me did, and it is very likely that both will have ceased to exist at the same time. This is all the more likely because teletransportation has not yet become a reality. (p. 291)

According to Parfit, when we invent a new concept, we may find it applies to parts of reality. The concept of a Phoenix applies to nothing, but the concept of a series-person applies just as often as the concept of a person. If Nagel's view is correct, for every person there exists a series-person who is *very* closely related to this person, so that the distinction between these persons is hardly worth drawing, except in discussing what for Parfit really matters. If again, Nagel's view is correct, then for Parfit, what matters for the old-me is the continued existence of his present brain. But even if we believe this, we can still believe that the continued existence of our present brains is *not* what matters. We could claim that what matters instead is relation R *for us as series-persons*. (p. 291)

The invention of some new concepts, such as 'atom' and 'molecule', can give a better description of reality. If Nagel's view is correct, the concept of a series-person also provides a better description of reality, but such an improvement is not so straightforward because it enables *different* beings both to proclaim their existence *and* to give this better description. (p. 191 - 192)

Nagel also introduces the concept of a "day-person". Such a person necessarily involves an uninterrupted stream of consciousness, but for whom sleep is death. This concept also applies to reality. At any given time there are as many day-persons as there are conscious living persons, but over the course of a year, the number of day-persons who have lived vastly exceeds the number of persons living. However, the concept of a day-person is worse than the concept of a person, because what matters is not the uninterruptedness of a stream of consciousness, but relation R. Although the concept of a day-person applies to reality, it picks out parts of reality with irrelevant boundaries. The daily interruptions to our stream of consciousness do not matter because they do not destroy psychological continuity.

If Nagel's view is true, then why, when we ask ourselves what is important about ourselves, our lives and our relations, does the continued existence of a particular brain not seem to matter to us? If instead, what is more important is relation R, psychological connectedness and/or continuity, with any cause taken together with Nagel's view, then the concept of a person is worse off than a series-person and worse in a similar way. For what matters to a series person is relation R, with any cause. (p. 192)

If again, Nagel's view is true, then a person cannot become a series-person, but a series-person can speak through the same mouth that both share:

This series-person can proclaim his existence, give himself a name, and claim that all the pronouns that he writes or utters will refer to series-persons. All other series-persons could do the same. All future human lives would then be lived by beings who regard themselves as series-persons. These lives would also be lived by persons. But persons would now have a subordinate role, since they would seldom refer to old-themselves. (p. 292)

If, however Nagel's view is not true, then what Parfit has just described makes no difference. Parfit's criteria of identity may fail to cover a few individuals, such as commissurotomy cases and some imaginary cases. But since people are not Cartesian entities, questions about personal identity in these imaginary cases are simply empty. In such cases, we may decide to *give* answers to these questions, thereby extending our beliefs about the criterion of personal identity. One way to do so might be to

make the this criterion the non-branching obtaining of Relation R, with any cause. On this criterion, persons *are* series-persons, so that the distinction disappears. (p. 292 - 293)

If Nagel's view, however is correct, we cannot make such claims. If so, then persons cannot be series-persons and the events Parfit described *would* make a difference. According to Parfit, If series-persons proclaimed their existence and began using pronouns to refer to themselves, this would be an improvement. Since every series-person is very closely related to a particular person, it would be better, again according to Parfit, if the series-person took the leading role. Because the concept of a series-person picks out parts of reality less arbitrarily, series-persons could deny that, what matters for them, is the continued existence of their brains, and we could claim that what fundamentally matters for us is relation R, with any cause. (p. 293)

99. Am I a Token or a Type

Parfit introduces a case described by Williams, in which a person could have many co-existing replicas. Williams describes the concept of a **person-type** which could best be explained by way of an example. Suppose there is some particular person, Mary Smith. The old scanning replicator produces many replicas of her at a particular time but destroys her body in the process. The replicas will all be Mary Smiths; indeed they will all be different tokens or instances of the same person-type. This would raise several questions about what matters. What should Mary Smith believe about her relation to her future replicas before she pressed the green button? Is what matters the continued existence of her body, including her brain? Or would it be about as good if there will be later living tokens of her type? (p. 293)

According to Parfit, this is the question about what ought to matter to the original Mary Smith, a question that Williams does not discuss. Williams does however discuss a related question:

Since we are not supposing that the token-persons, once printed off from the prototype, have intercommunicating experiences... they will be divergently affected by different experiences, and will tend to get increasingly dissimilar. Looked at as copies of the prototype, they will become copies which are increasingly blurred or written over; looked at in their own right, they will become increasingly individual personalities. This might be welcomed. For someone who loved one of these token-persons might well love her not because she was a Mary Smith, but despite the fact that she was a Mary Smith... The more the *Mary Smiths* diverged, the more secure the hold the lover might feel he had on what particularly he loved.

If someone loved a token person just as a Mary Smith, then it might well be unclear that the token-person was really what he loved. What he loves is *Mary Smith*, and that is to love the type-person. We can see dimly what this would be like. It would be like loving a work of art in some reproducible medium. One might start comparing, as it were, performances of the type; and wanting to be near the person one loved would be like wanting very much to hear some performance, even an indifferent one, of *Figaro* – just as one will go to the scratch provincial performance of *Figaro* rather than hear no *Figaro* at all, so one would see the very run-down Mary Smith who was in the locality, rather than see no Mary Smith at all.

Much of what we call loving a person would begin to crack under this, and reflection on it may encourage us not to undervalue the deeply body-based situation we actually have. While in the present situation of things to love a person is not exactly the same as to love a body, perhaps to say that they are basically the same is more grotesquely misleading than it is a deep metaphysical error; and if it does not sound very high-minded, the alternatives that so briskly grow out of suspending the present situation do not sound too spiritual, either. (Williams, 1973 p. 80 - 81. See also Brennan, 1982 & 1984)

Parfit asks if loving a person is basically the same as loving that person's body. Although Williams, above, admits that this claim is "grotesquely misleading", he suggests that it is better than any alternative, and that we should not undervalue "the deeply body-based situation that we actually have". But, according to Parfit, persons who had many co-existing replicas would threaten much of what we value. Nevertheless, we should accept this last claim, but not accept that loving a person is basically the same as loving their body, especially when there *is* a better alternative.

Parfit summarises Williams' reasoning as follows: Unless what I love is a particular body, I cannot love an individual. But suppose that I loved the original Mary Smith; however a machine destroys her body and produces a replica. If I then transfer my love to the Replica, this suggests that what I love is not an individual, but a person-type. And if we consider what such love entails, we find it disturbing. Parfit agrees that such love would be very different, and disturbing, but he rejects the reasoning just given.

Instead, Parfit asks us to imagine *two* kinds of imaginary cases. One is the community that Nagel imagined in which, although there is replication, relation R never takes a branching form. Suppose that once Mary Smith turns 30, she uses the youth-preserving replicator once a year, as do many other individuals in her community. If such a replicator existed, it would be *possible* to produce several co-existing replicas of a single individual; however we can suppose that in this community, the individuals choose not to use it for this purpose. Perhaps for the reasons Parfit set out in Section 90 and for those that Williams expresses, these individuals believe that division is not as good as ordinary survival. (p. 294)

Parfit supposes that he moves into this community and falls in love with Mary Smith. How should he react after she first uses the replicator? He claims that he would and *ought* to love her replica. This is not a moral "ought" but an ought based on the best conception of the best kind of love. She would be fully psychologically continuous with the Mary Smith he fell in love with and have a body that is exactly similar to that person he fell in love with. According to Parfit, If he did not love her replica, this could only be for one of several bad reasons. (p. 295)

One bad reason might be because one believes in the non-Reductionist view, and that personal identity is a further fact, which could not be produced by replication. Perhaps such a person would not love Mary Smith's replica because he did not believe that, in this profound way, she is not Mary Smith. But this reaction is unjustified because there is no further fact. Another bad reason might be because, even if one accepts the reductionist view, one might be grief-stricken at the destruction of her body when she pressed the green button. Perhaps he would come to love her replica, but this love would be tinged by grief. But this reaction is also unjustified. On the reductionist view, we should regard replication as being as good as ordinary survival. Since Mary Smith chose to be replicated, we can assume that this was her view also, and that Parfit should transfer his love to her rep-

lica. Moreover, since this community would be one of series-persons, Mary Smith's replica *is* Mary Smith.

But if Parfit's love is not transferred, there could be two further explanations. Recall that Williams reluctantly suggested that loving a person is, basically, loving a particular body. But this kind of love, or lust, or fetish, Parfit supposes, is extremely uncommon. What he regards as more common is a purely physical or sexual obsession with a person's body, that is not concerned with the psychological life of the person. But this is not *love* of a particular body. According to Quinton (1962), in the case of such obsessions, "no particular human body is required, only one of a more or less precisely demarcated kind". If Parfit were obsessed with Mary Smith's body, his obsession would transfer to her replica. This would be like being obsessed with the body of an identical twin. If this twin died, his obsession could simply be transferred to the other twin.

But ordinary love could not so easily be transferred. Love as we know it is concerned with the continuously changing psychological life of a person. And loving someone is a process, not a fixed state of being. Indeed, mutual love involves a shared history (and a mutual identity as a couple). For this reason, if Parfit loved Mary Smith for many months or years, her place could be simply taken by her identical twin. But twins are not replicas. If Parfit loved Mary Smith for many months or years, her replica would have full quasi-memories of their shared history together. (p. 295)

So if Parfit does not love Mary Smith's replica and was not simply in lust with her particular body, the remaining explanation may be that his love has ceased for no reason. And no reason is a bad reason. According to Parfit "Love can cease like this, but only an inferior kind of love".

Williams suggests that in a world of replication, we should distinguish person-types from tokens of these types. But for Parfit, in this world where relation R always takes a one-to-one form, such a distinction would be unnecessary. It would be a misdescription of what happens if we called each new Replica another token of a person type. Moreover, this description ignores what is most important, namely psychological continuity and the development of a life. Parfit believes that we could better describe what happens in either of two ways:

If Nagel's view is false, we could extend our criterion of personal identity to the non-branching obtaining of relation R, such that each individual in Nagel's imagined community would be a person – they would simply move to new bodies once a year. Since Mary Smith would have many such new bodies, his love for her should be directly transferred. Though such individuals would have many such bodies, love for particular people would not be in peril. If however Nagel's view is true, then individuals in this community are not persons, rather they are, and believe themselves to be, series-persons. So, the series-person, Mary Smith, moves to a new body each year. But loving a series-person is not the same as loving a person-type. It would be loving a particular individual, who has a continuous history.

Consider next the other alternative to the actual world: the one imagined by Williams. Here there would be many co-existing replicas of a single person. If there were fifty replicas of Greta Garbo as she was at age 30, these would be correctly described as different tokens of one person-type. But as Williams claims, if the object of love is a person-type, this would be very different from ordinary love, not placing much store in shared history. (p. 296)

Parfit claims that if he lived in such a world and was one of a set of replicas, he might regard himself as a token of a type. But what if he regarded himself as *the type*? On one sense of the word 'type', if he were a person-type, he could not simply cease to exist. Even if there were not any tokens of this type, there would still be this person-type. Parfit goes on to claim that a person-type would survive even the destruction of the universe because a type is an abstract entity, like a number and we could not possibly regard ourselves as abstract entities. (p. 296 - 297)

On any sense of the word 'type' there would be a significant difference between ordinary love and love of a person-type. Love of the latter kind would not be mutual. According to Parfit, one may love some person-type, but they could not reciprocate. A type can not love any more than a number could. One cannot be loved by the English Rose or by the Blonde Bombshell. What may be true is that love for one may involve some of the features of some person-type. All the many tokens of this type would be loved by me and perhaps be reciprocated. But any such love would, as Williams claims, lead away from loving the person-type because of the lack of shared histories.

Parfit now returns to Williams' chief claims. Williams suggests that, though it is misleading, there is a profound truth in that to love a person is to love a particular body, and that if it were not so, we would be loving a person-type. Such a love, Parfit claims, would be very different from ordinary love, and would be disturbing, threatening much of what we value. Following Quinton, Parfit doubts that anyone loves a particular body. A purely physical obsession would be an obsession with a body-type, and would have the same disturbing features of love of person-type. If one were obsessed with the body of an identical twin and that twin died, then the obsession would be transferred, without grief, to the remaining twin's body.

Parfit also denies that, if the object of our love is not a particular body, then we must love a person-type. This can be seen in Nagel's imagined alternative to the actual world, in which people are replicated often but in a one-to-one form. In this world relation R winds its way through many different bodies, but never takes a branching form. According to Parfit, in such world, ordinary love would survive unchanged. And if Nagel's view is false, people in this society would move to new bodies each year, but would still be particular people. However, if Nagel's view is true, it would be series-people who move to new bodies, but love would still be that of a particular individual because series-persons are still individuals. Parfit concludes that, "If these claims are correct, I can again keep the view that I have defended. What matters is not the continued existence of a particular body, but relation R with any cause". (p. 297)

100. Partial Survival

Before examining actual lives, Parfit considers several additional imaginary cases that reveal some further intuitions we have about identity. The first is fusion, which is the opposite of division. Logical identity is one-to-one and all-or-nothing; however just as division shows that what matters to survival need not take a one-to-one form, so fusion shows that it can have degrees. Recall that there was fusion in the central cases of the combined spectrum. In these cases, the resulting person would be psychologically connected, to about the same degree, to both of the original persons. (p. 298)

Fusion does sometimes occur at a microscopic level, but we can imagine two people coming together and while they are unconscious, their two bodies fusing together so that only one body and one

person wakes up⁴. This one person could quasi-remember once living the lives of the original two people. According to Parfit, no quasi-memories need be lost but some properties would be. The fused person would have different characteristics, desires and intentions. Some features would be compatible and would co-exist in the resulting person; others incompatible but of equal strength may cancel each other out, or if of differing strength may combine weakly. Potentially, the effects of fusion might be as predictable as the laws governing the combination of dominant and recessive genes.

Parfit imagines some examples: suppose he admires Palladio, and intends to visit Venice, while another person with whom he is about to fuse, admires Giotto, and intends to visit Padua. The resulting person would, presumably, have both tastes and intentions. And since Padua is close to Venice, both these intentions could be fulfilled. Suppose instead that Parfit loves Wagner and always votes Conservative, while the other person with whom he is about to fuse hates Wagner and always votes Liberal. The resultant person, according to Parfit, would be a tone-deaf floating voter.

Like the division of bodies, fusion does not fit the logic of identity. The resulting person would not be the same as each of the two original people; instead the resulting person would be neither of the original two. But this does not mean that either of these two persons should regard fusion as equivalent to death. Whereas division results in two identical persons, fusion results in one person, not wholly similar to each. As Parfit argued, there is no fact involved which is all-or-nothing; both physical and psychological connectedness could hold to any degree. How ought we to regard cases in which such relations obtain to a reduced degree? (p. 298)

Parfit anticipates one reaction but immediately dispenses with it:

Suppose that, between me and some resulting person, there would be about half the ordinary amount of these two relations. This would be about half as good as ordinary survival. If there would be nine-tenths of these ordinary amounts, this would be about nine-tenths as good.

This view is too crude because it fails to take into account how close the relation was between the two fusing persons and between them and the resulting person. We must also know whether the resulting person has qualities that either fusing person would have regarded as good or bad. (p. 299)

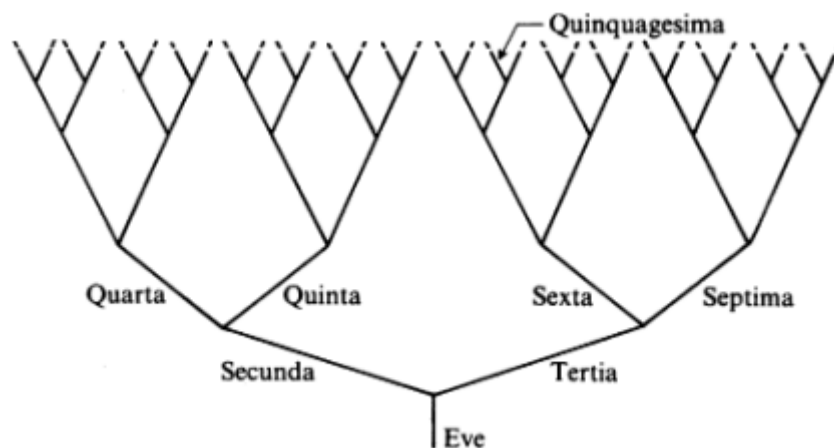
Looking at it from one side, Parfit suggests the following: In judging the value of one person to the resulting person depends both 1) on the degree of connectedness to this person, and 2) on the value, in this person's view, of this person's physical and psychological features. Suppose, by way of example, that hypnosis causes Parfit to lose five unwanted features *i.e.* untidiness, laziness, fear of flying, nicotine addiction, and all memories of a wretched life. These represent much less than psychological connectedness, but this is more than outweighed by the removal of bad features.

Only the narcissistic view themselves as perfect. Most of us would welcome some changes to our physical and mental features. So long as the changes were improvements, we would welcome the partial reduction in both kinds of connectedness above. If however fusion involved subtracting val-

⁴ The reverse of such a process is depicted in the B-grade movie *Doppelgänger* (1993) starring Drew Barrymore and George Newbern. The protagonist, fused with the body of her evil twin which was absorbed during pregnancy, untwine to become two separate bodies.

ues that we value or adding those that we find repugnant, we should avoid fusion. Parfit supposes that if there were only two things that gave his life meaning: his struggle for Socialism and his love of Wagner, he would dread fusion with a Wagner-hating Conservative. If the resulting person would be a tone-deaf floating voter, his relation to him would be nearly as bad as death. However, on another scenario, while involving as much but different change, he might regard it as better than ordinary survival. Such changes might all be regarded as improvements. According to Parfit, “Fusions, like marriages, could be either great successes, or disasters”.

Parfit introduces yet another kind of creature, just like us except for their mode of reproduction which is fission, just like an amoeba. The lives of these people are shown in the diagram below.



The line segments joining each node represent the spatio-temporal paths traced by the bodies of these people, while the nodes represent points of division. A line segment between two points of division Parfit calls a *branch*. The whole structure he refers to as a *tree*. Each branch then corresponds to the life of one person, with the first person being Eve. The names of subsequent persons are given by feminine Latin ordinals: Secunda, Tertia, Quarta, etc. The fiftieth person along the tree is Quinquagesima. (p. 299)

At the start of their lives Secunda and Tertia are fully psychologically connected to Eve, just before she divided. As Parfit argued before, Eve’s relation to each of her divided selves is about as good as ordinary survival. The same is true of every other division represented in this community. But what about the distant descendent Quinquagesima? Between Eve and Quinquagesima there will be direct continuous chains of overlapping psychological connections. Thus Eve has some quasi-intentions that are carried out by Tertia, who in turn has some that are carried out by Sexta, and so on all the way down to Quinquagesima. Similarly, Quinquagesima can quasi-remember most of the life of her immediate predecessor, who can quasi-remember most of her life of her immediate predecessor, and so on, right back to Eve. Though Quinquagesima and Eve are psychologically continuous, there is no *direct* connectedness between the two. Like us, Eve cannot be strongly connected to every person in an indefinitely long tree. With subsequent divisions, shared quasi-memories and quasi-intentions will gradually weaken and fade away. Similarly quasi-ambitions, once realised, will be replaced by others, and quasi-characteristics will gradually change. Persons further along the tree will have fewer direct connections with Eve, and those sufficiently remote will have *no* direct and distinctive psychological connections. (p. 300)

Parfit emphasises *distinct* because there would be some kinds of direct connection. Quinquagesima would presumably inherit from Eve many memories, such as the fact that she and every other person in the community reproduces by division. She would also inherit many general abilities, such as how to speak, or swim, but not any of the psychological features that distinguish Eve from most of the other members of the community.

According to Parfit, both relations of psychological continuity and psychological connectedness matter, and that, for want of good argument to the contrary, neither relation matters more than the other.

Because of its importance later, Parfit introduces another view on which connectedness does not matter but continuity does. If there is some later person psychologically continuous with him as he is now, it would not matter at all if there were no direct psychological connections with that person. Consider first memory. If our lives have been worth living, we would highly value our ability to remember many of our past experiences. However, the loss of such memories need not destroy the continuity of memory, which only requires overlapping chains of memories. Parfit supposes that, if two days hence, the only memories he will have will be his episodic memories of tomorrow. On the view just stated, this is all that matters because there will be a continuity of memory. On this view again it should not matter that he will soon lose all his memories of his past life; although most of us would strongly disagree: the loss of all such memories would be something we would deeply regret.

Next Parfit asks us to consider the continuity of our desires and intentions. If he now loves certain other people, he could cease to love all of those people without any break in psychological continuity; although he would greatly regret such changes. Suppose that he also wants to achieve certain aims. Given that these would be cherished, he would also regret their replacement by other desires. But if they were replaced, he would have to care more *now* about the achievement of what he *now* cares about, and since he would now care more about the fulfilment of these present desires, he would regret losing them. More generally, Parfit claims, he would want his life to have certain kinds of overall unity. If he were in a constant state of flux *viz.* his desires and concerns his psychological continuity would not be diminished, however his psychological connectedness would be reduced. This is a change most of us would regret. (p. 301)

Clinical Cases 4 - Manic-Depressive Episodes

People in the midst of a **manic-depressive episode** display very obvious mood swings from depression to elation, sometimes within the same hour. What can also be very distressing is their diminished *psychological connectedness*. In the manic phase, apparent intentions, interests, expectations and stable values can fluctuate, sometimes within mid-sentence. Thoughts may occur at a rate too fast to process or articulate, and the person may lose touch with reality that would otherwise be their stable reference. In extreme cases, elation may give over to anger as others apparently fail to understand them or to instantly meet their unreasonable demands. Fortunately, there are drugs available to treat and prevent such episodes, but they must be taken chronically, sometimes for a lifetime.

Finally, consider continuity of character. There will be continuity of our character if it changes in a natural way, but most of us value aspects of our character that we do not want to change. Again, we prize connectedness over mere continuity. (p. 301)

Summarising this section, Parfit has described three reasons why most of us would reject the view that psychological connectedness does not matter. Therefore we can agree that connectedness is not all that matters but that psychological continuity also matters, but neither that it only matters. (p. 302)

101. Successive Selves

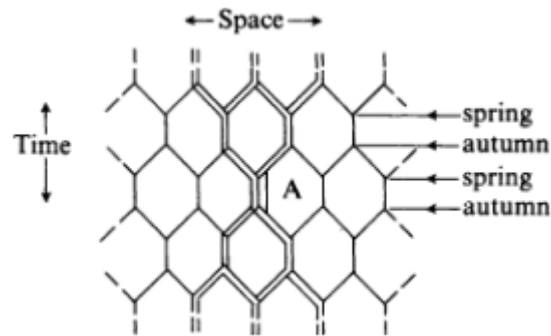
In this section Parfit describes how people who reproduce by natural division could describe their interactions. Each person is a *self* and Eve could think of any person, anywhere along the tree, as one of her *descendant selves*. This term implies a future-directed continuity. Unlike Eve however, Tertia has descendant selves only along the right half of the tree. Parfit suggests that these people could use the phrase *an ancestral self* to imply past-directed continuity. For example, Quinquagesima's ancestral selves are all of those that lie on the single line connecting her to Eve. (original emphasis)

Since the relation of psychological continuity is transitive, in either temporal direction, *being an ancestral self of*, and *being a descendent self of* are also transitive relations. However psychological continuity is not a transitive relation in both directions. Quarta and Septima, for example, are both psychologically continuous with Eve, but it does not follow, and indeed it is false, that they are psychologically continuous with each other. (original emphasis)

Next Parfit suggests how these people could describe their different degrees of connectedness. In ordinary parlance the phrases *my past self* and *my future self* refer to myself in the past or the future, respectively. However among people who reproduce by natural division these phrases could be used to refer to other people whose relation to me is psychological connectedness. Thus the phrase 'one of my past selves' implies some degree of connectedness. And we could imply different degrees of connectedness by the use of such phrases, in descending order, as: 'my closest past self', 'one of my close past selves', 'one of my more distant selves', 'hardly one of my past selves (I can quasi-remember only a few of her experiences)' and, finally, 'not one of my past selves, merely an ancestral self'. Quinquagesima could use any of these past-directed phrases, while Eve could use any similar future-directed series of phrases. (original emphasis)

Parfit's terminology could also be used to describe his earlier imagined division. Though he would not survive his division, the two resulting people are two of his future selves and they are as close to him as his self is tomorrow. Similarly, each could describe him as an equally close past self and they can share him as a past self without being the same self as each other. (p. 302)

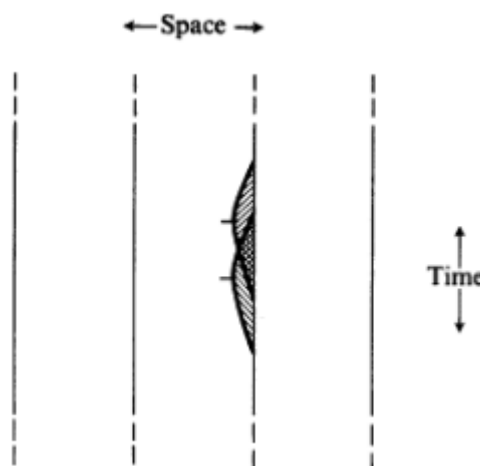
Consider now Parfit's penultimate kind of imaginary people. They reproduce twice a year by fusion every autumn as well as by division every spring. Their relations are shown diagrammatically below.



Person A is represented by the triple branched line in the centre. The double branched lines represent those lives that are psychologically continuous with A's. Each person has his own double-lined tree which overlaps with, but is different from other such trees. For these people the terms 'an ancestral self' and 'a descendent self' would be too broad to be of much use. According to Parfit, "There could be pairs of dates such that everyone who ever lived before the first date was an ancestral self of everyone who will ever live after the second date". Also the term 'I' would not be the same as our usage, since each imagined person's life lasts for only half a year. Much of what they mean by their identity would have to be couched in terms of talk about past and future selves.

These phrases, however cannot be used just as we do. 'A past self' implies psychological connectedness, which for us can be used to suggest varying degrees. But in this case successive selves are not distinguished by reduced degrees of connectedness – they reproduce by both division and fusion. This difference would probably not concern them because they divide and unite so frequently, and their life between such events are comparatively so short, that within any single life psychological connectedness would always hold to a very high degree. (p. 303)

Finally, Parfit considers another kind of imaginary people. They are just like us, except they do not reproduce at all, neither sexually, nor by division, nor by fusion. Their bodies are everlasting, but they do change gradually in appearance. As with us there are direct and distinctive psychological connections over limited periods of time, say five hundred years. This is shown in Parfit's diagram, below. The two shaded areas represent degrees of psychological connectedness to two central points (p. 303 -304)



These people could not use the vocabulary Parfit proposed above. Since there is no branching of psychological continuity, they would regard themselves as immortal, and in one sense they are. The parts of each 'life-line' are all psychologically continuous, but there are direct and distinctive connections only between parts that are close to one another. These people therefore, would not regard each 'line' as corresponding to a single, undifferentiated life. If they did, they would have no way of picking out these direct psychological connections. According to Parfit,

When such a person says, for example, 'I spent a period exploring the Himalayas', his hearers would not be entitled to assume that the speaker has any memories of this period, or that his character then and now are in any way similar, or that he is now carrying out any of the plans or intentions which he then had. Because the word 'I' would carry none of these implications, it would not have for these immortal people the usefulness which it has for us. (p. 304)

Parfit suggests a revisal to his earlier proposed way of talking. Instead of distinguishing between successive selves by reference to the branching of psychological continuity, we could make reference to degrees of psychological connectedness. Since this is a matter of degree, such distinctions can be left to the choice of the speaker, and be allowed to vary according to context. Such distinctions, drawn within a single life, come much closer to what we mean in ordinary parlance by the phrases 'my past self' and 'my future self'. On this revisal, the use of 'I' and other pronouns refer to the parts of our lives, to which, when speaking, we have the strongest psychological connection. When the connections have been weakened over time, or due to a significant change of character, lifestyle, beliefs and ideals, we could say with them, 'It was not I who did that, but an earlier self'. We, and they, might even describe in what ways, and to what degree, we are now related to this past self. (p. 304 - 305)

In the following extract Parfit quotes two very different authors in which this way of describing our own lives is both natural and useful.

We are incapable, while we are in love, of acting as fit predecessors of the next persons who, when we are in love no longer, we shall presently have become... (Proust 1949, p. 226 slightly edited)

Our dread of a future in which we must forego the sight of faces, the sound of voices, that we love, friends from whom we derive today our keenest joys, this dread, far from being dissipated, is intensified if to the grief of such a privation we reflect that there will be added what seems to us now in anticipation an even more cruel grief: not to feel it as a grief at all — to remain indifferent: for if that should occur, our self would then have changed. It would be in a real sense the death of ourself, a death followed, it is true, by a resurrection, but in a different self, the life, the love of which are beyond the reach of those elements of the existing self that are doomed to die... (Proust, 1949, p. 349)

It is not because other people are dead that our affection for them grows faint, it is because we ourself are dying. Albertine had no cause to rebuke her friend. The man who was usurping his name had merely inherited it... My new self, while it grew up in the shadow of the old, had often heard the other speak of Albertine; through that other self... it thought that it knew her, it found her attractive... But this was merely an affection at second hand. Nadya

had written in her letter: 'When you return...' But that was the whole horror: that there would be no return... A new, unfamiliar person would walk in bearing the name of her husband, and she would see that the man, her beloved, for whom she had shut herself up to wait for fourteen years, no longer existed... (Solzhenitsyn, 1969 p. 232)

Innokenty felt sorry for her and agreed to come... He felt sorry, not for the wife he lived with and yet did not live with these days, but for the blond girl with the curls hanging down to her shoulders, the girl he had known in the tenth grade... (Solzhenitsyn, 1969 p. 393)

These passages suggest that the object of our emotions may not be another person "timelessly" considered but another person within a different period during their life. Parfit provides, what he believes to be, a common example: A couple may clearly love each other. But if asked if they are still *in love with* each other, they may find the question perplexing. It may seem to them that they are in love, yet their feelings and behaviour towards each other, in each other's presence, may not be consistent with what it seems. However, their perplexity might be resolved if they distinguished between successive selves. They might recognise that they love each other, but are in love with each other's earlier selves. (p. 305)

Parfit cautions that talk about successive selves can easily be misunderstood or taken too literally. He suggests that it should be compared with the way we subdivide a nation's history. Indeed we meaningfully subdivide human prehistory itself into successive material cultures *e.g.* Early Stone Age, Middle Stone Age, Late Stone Age *etc.* There is another caveat to this way of talking. It is only suited for cases where there are sharp discontinuities marking the boundaries between successive selves. Where there are reduced degrees of psychological connectedness without such discontinuities, it is better to talk of the degrees of connectedness directly. (p. 305 - 306)

Finally, after 107 pages of closely argued text Parfit does not explicitly state his overall conclusion so far, namely that in terms of bodily continuity and personal identity, we are most similar to these last imagined people.

The Self as a Center of Narrative Gravity

In the 1992 translation of the Dennett's 1986 article, available [here](#)⁵, he draws an analogy between selves and something much simpler that has some salient properties in common with selves, namely the **centre of gravity** of an object, defined as the theoretical point within an object where its total weight can be thought of as being concentrated. Although centres of gravity are a well-behaved concept in Newtonian physics, they do not correspond to anything physical in the world. Apart from spatio-temporal location, they have no properties at all, neither mass, nor momentum, nor colour *etc.* In fact, they are a purely theoretical fiction, albeit a very useful one. Dennett reminds us how robust and familiar the idea of a centre of gravity is:

Consider a chair. Like all other physical objects, it has a center of gravity. If you start tipping it, you can tell more or less accurately whether it would start to fall over or fall back in place if you let go of it. We're all quite good at making predictions involving centers of

⁵ Note the original pagination is missing in this copy.

gravity and devising explanations about when and why things fall over. Place a book on the chair. It, too, has a center of gravity. If you start to push it over the edge, we know that at some point [it] will fall. It will fall when its center of gravity is no longer directly over a point of its supporting base (the chair seat). Notice that that statement is itself virtually tautological. The key terms in it are all interdefinable. And yet it can also figure in explanations that appear to be causal explanations of some sort. We ask “Why doesn’t that lamp tip over?” We reply “Because its center of gravity is so low.” Is this a causal explanation? It can compete with explanations that are clearly causal, such as: “Because it’s nailed to the table,” and “Because it’s supported by wires.”

And of course, we have all learned as children, the hard way, to keep our centre of gravity *over* our bicycle’s centre of gravity and not to ride chairs. But we can also manipulate the centre of gravity of an object by shifting our position, as in riding a bike, or by shifting the position or weight of an object such as a jug of water. Tip some of the water out or just tilt the jug and its centre of gravity shifts so that all the weight of the water and jug combined can now be thought of as being centred at a new point somewhere within the jug.

Although centres of gravity are purely abstract, they do have a spatio-temporal history which may even be discontinuous. Dennett imagines suddenly sticking a wad of gum to a jug. Its centre of gravity will shift from one point to another without travelling through all the intervening positions. As an abstract entity it is not bound by the constraints of physical motion.

Dennett goes on to imagine the centre of gravity of a more complicated, highly unlikely object, say a steam powered unicycle with many turning gears, camshafts and reciprocating rods. He imagines further that we had a theory of the machine’s operation that allowed us to precisely plot the trajectory of its centre of gravity. If it were discovered that the trajectory of this point corresponded to that of a particular atom of iron in the camshaft, we would be wrong in thinking that the machine’s centre of gravity was identical with this atom of iron. A fiction cannot be identical to something physical. But when Dennett says “it’s a fictional object” he does not mean to disparage it. On the contrary he says, “it’s a wonderful fictional object, and it has a perfectly legitimate place within serious, sober, [authentic] physical science”.

But a self is also an abstract object – a theoretical fiction – not a theory of particle physics, but a theory of what might be called people-physics⁶. While the physicist carries out an *interpretation* of the behaviour of object, such as a chair, and comes up with a theoretical abstraction of its centre of gravity, so the hermeneuticist, phenomenologist or anthropologist sees human beings and other animals moving about in the world and is faced with a similar problem of interpretation. It turns out to be theoretically perspicuous (and parsimonious) to organise the interpretation around a central abstraction: a *self* (in addition to the body’s centre of gravity). Even within folk psychology, we posit a self for *ourselves* as well as others. Dennett observes that “[t]he theoretical problem of self-interpretation is at least as difficult and important as the problem of other-interpretation”.

Now compare a self with a centre of gravity. For one, a self is a much more complicated concept. Dennett introduces another fictional object by way of elucidation, namely a literary fictional charac-

⁶ Dennett also uses the terms ‘phenomenology’, ‘hermeneutics’ and ‘*Geisteswissenschaft*’ or ‘soul-science’ which are similar in some contexts but not equivalent.

ter. On page 1 of Melville's *Moby Dick* we read, "Call me Ishmael". Whom are we to call Ishmael? Melville? No, Ishmael, the character Melville created. As we read the book we learn about Ishmael, his life, beliefs, desires, acts and much more besides that Melville never explicitly mentions. These latter have to be "read in" either by implication or via extrapolation. Beyond these limits, fictional worlds are simply *indeterminate*. Dennett borrows the following example from David Lewis (1978) "Truth and Fiction":

Did Sherlock Holmes have three nostrils? The answer of course is no, but not because Conan Doyle ever says that he doesn't, or that he has two, but because we're entitled to make that extrapolation. In the absence of evidence to the contrary, Sherlock Holmes' nose can be supposed to be normal. Another question: Did Sherlock Holmes have a mole on his left shoulder blade? The answer to this question is neither yes nor no. Nothing about the text or about the principles of extrapolation from the text permit an answer to that question. There is simply no fact of the matter. Why? Because Sherlock Holmes is a merely fictional character, created by, or constituted out of, the text and the culture in which that text resides.

This kind of indeterminacy is fundamental to fictional objects, which distinguishes them from the sorts of objects that scientists talk about, namely theoretical entities or what Hans Reichenbach called *illata* (from the Latin *inferre* 'to carry into') i.e. inferred entities, such as atoms, molecules, fundamental particles *etc.* Contrast these with **abstracta**: calculation-bound entities or logical constructs. (Dennett, 1989 p. 53) Logically, the principle of bivalence that states that every proposition is either true or false (see Critical Reasoning 21), simply does not apply in fictional worlds. In the real world a question such as 'Did Aristotle have such a mole?' is either Yes or No but unknown, but not indeterminate. In a fictional world however, such questions may have no answer at all.

According to Dennett, centres of gravity, as fictions, exhibit the same feature. They have only the properties that the theory that posits them endows them with. If you wonder whether centres of gravity may be a particle, you have misunderstood their theoretical status. How is it then that Parfit, The Buddha, Dennett and most contemporary philosophers and some psychologists regard the real self as a fiction? Aren't fictional selves dependent for their existence on real selves. Ishmael is a fictional character but his existence is dependent on the very real historical novelist, Melville. "Doesn't this show that it takes a real self to create a fictional self?" Dennett thinks not and back in 1982 he introduced an imaginative exercise to support his belief. Fortunately, what could only be imagined in 1982 has already been realised.

Artificial Intelligence (AI) networks trained on vast corpora of text are now capable of writing coherent, fictitious stories, given a limited amount of information about potential characters, contextual clues, happenings *etc.* All this is done by algorithms encoded in software running on mindless machines. Admittedly, no AI has produced anything that would pass for Dostoyevsky or Shakespeare, but they are getting better. Similar AIs can already compose music in any given style: Bach, Mozart, Beethoven *etc.* In Dennett's imaginative example of a fiction generating AI he supposes that the first sentence the computer produced was: "Call me Gilbert", followed by an apparent autobiography of some fictional character, Gilbert. Clearly Gilbert is a fictional self whose creator is no self. Even the creators of the AI and the manufacturers of the computer did not design Gilbert. Gilbert is the product of trained algorithms that neither the designers nor the manufacturers understand and behind which there is no self.

We can do with the autobiography of Gilbert what we please, including subjecting it to the same literary analysis that we would with any other text. And if you are a post-modern critic, so much the better. You have an interpretable text, and what the author intended is strictly irrelevant – the “author” doesn’t have intentions. According to Dennett, “Your expectations and predictions, as you read, and your interpretive reconstruction of what you have already read, will congeal around the central node of the fictional character, Gilbert”.

But now Dennett tweaks the example somewhat. Suppose that the computer running the AI is given a set of wheels and a video camera and begins moving about in the world. This mobile computer also begins its tale with “Call me Gilbert”, followed by an autobiography. If we do what post-modern literary critics say we should never do and *look outside the text*, we might find that there is a truth-preserving interpretation of the text in the real world. The adventures of Gilbert, the central character, begin to bear a striking resemblance to those of the robot wheeling about in the world. Dennett takes up his imaginative exercise:

If you hit the robot with a baseball bat, very shortly thereafter the story of Gilbert includes his being hit with a baseball bat by somebody who looks like you. Every now and then the robot gets locked in the closet and then says “Help me!” Help whom? Well, help Gilbert, presumably. But who is Gilbert? Is Gilbert the robot, or merely the fictional self created by the robot? If we go and help the robot out of the closet, it sends us a note: “Thank you. Love, Gilbert.” At this point we will be unable to ignore the fact that the fictional career of the fictional Gilbert bears an interesting resemblance to the “career” of this mere robot moving through the world. We can still maintain that the robot’s brain, the robot’s computer, really knows nothing about the world; it’s not a self. It’s just a clanky computer. It doesn’t know what it’s doing. It doesn’t even know that it’s creating a fictional character. (The same is just as true of your brain; it doesn’t know what it’s doing either.) Nevertheless, the patterns in the behavior that is being controlled by the computer are interpretable, by us, as accreting biography – telling the narrative of a self. But we are not the only interpreters. The robot novelist is also, of course, an interpreter: a *self*-interpreter, providing its own account of its activities in the world.

Suppose we take this analogy seriously. We might ask, “Where is the self?” According to Dennett, it is a category mistake to start looking around for a self in the brain. (See Ryle in Classic Text 06) At least centres of gravity have a spatio-temporal position, but selves are spatio-temporally only crudely defined. Roughly speaking and *ceteris paribus*, if there are three humans seated side by side on a bench, there are three selves there. Advances in neuropsychology have already identified *some* finer-grained localizations of the sense of self and self-functions but this does not imply that one day we will be able to accurately pin point the “seat of the self” within the brain.

Of course, there is a big difference between fictional characters and our selves. One that Dennett stresses is that fictional characters are usually encountered as a *fait accompli*. By the time we read a novel it has already been written, edited and published. We cannot, for example, ask Dostoyevsky what *else* Raskolnikov thought while sitting in the police station. The existence of collaborative novels today written by many authors and published with regular updates, revisions and multiple drafts on the internet makes much of what Dennett was claiming in 1982 redundant. Collaborative fictional

characters evolve. There is no one *true* story that we can reference. Sometimes they may be inconsistent, in which case the temptation is to *bifurcate* characters in order to resolve the conflicts.

Something like this happens to real people with dissociative identity disorder (DID) characterised by the presence of at least two distinct and relatively enduring personality states, inexplicable memory gaps, recurrent episodes of dissociative amnesia, intrusions into consciousness, depersonalization and derealization, amongst others. Consider the cases of Eve, detailed in the book *The Three Faces of Eve* (Thigpen & Cleckley, 1957) and Sybil in *Sybil* (Schreiber, 1973). Eve's three "faces" represent three distinct personalities, while Sybil apparently had sixteen. In his discussion of what, at the time, he thought happened to these women, Dennett conflates the terms "personality" and "self" and also seems to have had the concept of identity in mind. Therefore we should first define these terms explicitly as used in psychiatry and, in the spirit of charitable interpretation, allow Dennett to proceed with his argument. The **self** is the psychophysical total of a person at any given moment, including conscious and unconscious attributes. **Identity** is one's global role in life and the perception of the sense of self. **Personality** is the characteristic configuration of behaviour response patterns that each person evolves as a reflection of his or her individual adjustment to life. (Sadock *et al.*, 2017 pp. 4611, 4586, 4602)

Clinical Cases 5 - Dissociative Identity Disorder (DID)

The diagnosis of DID has been controversial, especially since the hysteria surrounding alleged cases of scores of unique personalities within an individual, caused by severe physical, sexual and religious abuses that were impossible to verify or are alleged to have been remembered from infancy. Certain psychotherapeutic techniques have also been implicated in artificially producing the disorder by creating false memories and beliefs appropriate to the diagnosis in vulnerable patients. Fortunately, the diagnostic criteria today are much more robust and the condition, although rare, is recognized by the DSM-5-TR as well as by the ICD-11 and Merck Manual for diagnosis. Although there are at present no medications *specifically* for treating DID, various other forms of treatment involve a multidisciplinary task team.

On the advice *some* sceptical psychotherapists, Dennett proposed that the first time Sybil consulted her therapist she was not one body with several personalities. Sybil then was akin to a novel-writing machine that fell in with a very ingenious questioner and eager reader. Together they collaborated, unwittingly, to write many chapters of a new novel of which Sybil was the living embodiment. She went out and engaged with the world with these new selves, "more or less created on demand, under the eager suggestion of [her] therapist".

In the following paragraph, Dennett concedes that this scenario is overly sceptical. According to him, the multiplicity of characters that followed the onset of psychotherapy for what was then known as Multiple Personality Disorder in the '80s and '90s could probably be explained along the sceptical lines above. However today, there is compelling evidence for the existence, within diagnosed individuals, of just a few identities that had already begun laying down narrative biographies before they were exposed to the "readings" of a therapist. Sybil, for example, is an extreme pathological case of something we engage in quite normally *i.e.* confabulation, telling and retelling ourselves our own life's narrative with little regard to the question of objective truth.

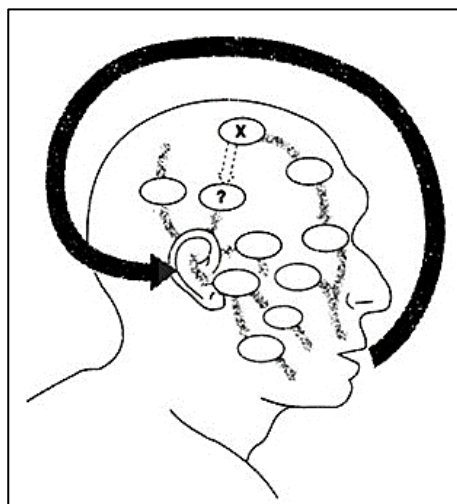
Why then are we all “such inveterate and inventive autobiographical novelists?” It took biologist and philosopher to Humberto Maturana to state the obvious when he wrote, “Anything said is said by an observer. In his discourse the observer speaks to another observer, who could be himself...” (Maturana, 1970/1980 p. 8) But why should we talk to ourselves, asks Dennett. “Why isn’t that an utterly idle activity, as systematically futile as trying to pick oneself up by one’s own bootstraps?”

Recall that in commissurotomy patients, the self is not split in two. Although, for the sake of discussion, we have emphasized the differences, they exhibit no signs of psychological splitting except in highly contrived laboratory conditions. According to Gazzaniga, this does not show that these patients have preserved their pre-surgical unity, but rather that the unity of normal life is an illusion.

...the normal mind is not beautifully unified, but rather a problematically yoked-together bundle of partly autonomous systems. All parts of the mind are not equally accessible to each other at all times. These modules or systems sometimes have internal communication problems which they solve by various ingenious and devious routes.

Dennett thinks this much is true and that it may provide an answer to the puzzling question about what conscious thought is for. Such a question begs an evolutionary answer, but we were not present at the dawn of human consciousness and therefore such an answer will have to be speculative.⁷ Dennett draws on Julian Jaynes’ (1976) *The Origins of Consciousness in the Breakdown of the Bicameral Mind* in adapting his account. Accordingly, our distant ancestors weren’t full conscious.

They spoke, but they just sort of blurted things out, more or less the way bees do bee dances, or the way computers talk to each other. That is not conscious communication, surely. When these ancestors had problems, sometimes they would “ask” for help (more or less like Gilbert saying “Help me!” when he was locked in the closet), and sometimes there would be somebody around to hear them. So they got into the habit of asking for assistance and, particularly, asking questions. Whenever they couldn’t figure out how to solve some problem, they would ask a question, addressed to no one in particular, and sometimes whoever was standing around could answer them. And they also came to be designed to be provoked on many such occasions into answering questions like that – to the best of their ability – when asked. Then one day one of our ancestors asked a question in what was apparently an inappropriate circumstance: there was nobody around to be the audience. Strangely enough, he heard his own question, and this stimulated him, cooperatively, to



Dennett (1991 p. 196) uses this diagram to illustrate how self-talk can provide a communication link between two not otherwise interaccessible regions of the brain and a possible explanation how human consciousness might have evolved

⁷ There are, at present, insurmountable ethical obstacles to creating consciousness *in vivo* or even *in silico* even if we had the technology. Moreover we do not know where “genetic evolution and transmission breaks off and cultural evolution and transmission takes over” but this does not touch on Dennett’s speculative answer.

think of an answer, and sure enough the answer came to him. He had established, without realizing what he had done, a communication link between two parts of his brain, between which there was, for some deep biological reason, an accessibility problem. One component of the mind had confronted a problem that another component could solve; if only the problem could be posed for the latter component! Thanks to his habit of asking questions, our ancestor stumbled upon a route via the ears. What a discovery! Sometimes talking and listening to yourself can have wonderful effects, not otherwise obtainable. All that is needed to make sense of this idea is the hypothesis that the modules of the mind have different capacities and ways of doing things, and are not perfectly interaccessible. Under such circumstances it could be true that the way to get yourself to figure out a problem is to tickle your ear with it, to get that part of your brain which is best stimulated by *hearing* a question to work on the problem. Then sometimes you will find yourself with the answer you seek on the tip of your tongue.

This would be enough to establish the evolutionary endorsement (which might well be only culturally transmitted) of the behavior of talking to yourself. But as many writers have observed, conscious thinking seems – much of it – to be a variety of particularly efficient and private talking to oneself. The evolutionary transition to thought is then easy to conjure up. All we have to suppose is that the route, the circuit that at first went via mouth and ear, got shorter. People “realized” that the actual vocalization and audition was a rather inefficient part of the loop. Besides, if there were other people around who might overhear it, you might give away more information than you wanted. So what developed was a habit of sub-vocalization, and this in turn could be streamlined into conscious, verbal thought.

Recall how commissurotomy patients could use nonvisual clues from both hemispheres to guess the identity of object presented to only the right hemisphere, which has no direct access to left, which controls language. Something analogous is happening according to Jaynes’ hypothesis where information from one modality is used to bridge the gap between cognitive modules that are, for what ever reason, “not perfectly interaccessible”. And it makes sense for this, or these routes if there are others, to be “streamlined” into conscious thought. If this hypothesis is correct, an external route of self-communication could have enabled the kind self-referential, self-aware consciousness characteristic of human beings.

Now it could be that commissurotomy patients have developed the talent of bridging their extreme interaccessibility problem, brought about by their operation, or it could be that the operation itself reveals – but does not create – this ability to be found in otherwise normal people. Gazzaniga claims that the latter is the most likely hypothesis to investigate. According to Dennett, it does seem that we are all virtuoso novelists who try to put the best presentation on sometimes unified, but other times disunified behaviour. “We try to make all of our material cohere into a single good story. And that story is our autobiography.”

Of course, the chief character at the centre of any autobiography is the narrative *self*, so that asking what besides the self *really* is, is to make a category mistake. Dennett proposes that when a person’s behaviour control system becomes seriously impaired, the best hermeneutical story that we can spin about the person is that there may be more than one character “inhabiting” their body. This is consistent with the view of the self that Dennett has been offering and it does not require an appeal to

metaphysical miracles. All that is required is that the story does not cohere around one self, but that it better coheres around two or maybe more selves.

We sometimes *do* encounter psychological disorders or surgically created disunities where the best way to interpret or make sense of them is to posit two centres of narrative gravity. It is not that we are discovering another metaphysical self, only that we are creating another abstraction – an abstraction that we use to understand, predict and make sense of complex human behaviour. According to Dennett, “The fact that these abstract selves seem so robust and real is not surprising. They are much more complicated theoretical entities than a center of gravity... [Yet] no one has ever seen or ever will see a center of gravity”. The same is true of the self, as David Hume observed”:

For my part, when I enter most intimately into what I call *myself*, I always stumble on some particular perception or other, of heat or cold, light or shade, love or hatred, pain or pleasure. I never can catch *myself* at any time without a perception, and never can observe anything but the perception... If anyone, upon serious and unprejudiced reflection, thinks he has a different notion of *himself*, I must confess I can reason no longer with him. All I can allow him is, that he may be in the right as well as I, and that we are essentially different in this particular. He may, perhaps, perceive something simple and continued, which he calls *himself*; though I am certain there is no such principle in me. (*Treatise on Human Nature*, I, IV, §6)

Finally, see below some recent developments in the functional localization of the sense of the self or embodiment in the human brain.

Clinical Cases 6 - The Sense of “I-ness” or Embodiment and the Anterior Precuneus

Josef Parvizi *et al.* (2021) published a case study of an epilepsy patient who reported that during seizures he would enter a state of dissociation that caused him to lose his sense of coordination and feel disconnected with his inner self. The source of his seizures was localized to a part of the posteromedial cortex known as the anterior precuneus. In a subsequent study (Lyu *et al.*, 2023) the researchers recruited an additional eight patients with electrodes implanted bilaterally in the precuneus and other locations. In all the patients, stimulation of the anterior precuneus caused dissociative changes in physical and spatial domains. These included a feeling of floating, dizziness, a lack of focus and a sense of detachment from themselves. The researchers propose that this subregion is “integral to a range of cognitive processes that require the self’s physical point of reference, given its location within a spatial environment”. (Kwon, 2023) Note the researchers have not claimed to have found the seat of the elusive Cartesian *ego* in the brain; rather they have identified a region that appears to be necessary for the creation of the sense of bodily self or “I-ness”.

Task:

What, for you, are the key “take home” messages, both philosophical and scientific, from the discussion of Parfit’s and Dennett’s contributions above?

Feedback:

Naturally, each reader will have their own take on the discussion. Some may be in partial or full agreement with the authors' conclusions; others may be more sceptical. Here are some points about which we believe there can be little doubt.

- The terms 'self', 'person', 'personal identity', and 'personality' are related but not simply interchangeable.
- The self is not a thing or an entity apart from or separate from the body (including the brain) but neither is it *identical* with the body or part of the body. At most the self is a complex function of the body or at least it is a highly useful fiction; however, according to Buddhists, it is an illusion we ought to get over.
- There is no "seat" of the self in the brain, or anywhere else in the body. There are however various regions of the human brain that are to a greater or lesser extent involved in producing one or other aspect of the *sense* of self including the sense of embodiment.
- As it happens with humans, personal identity contingently involves the continuity of the body (including the intact brain) over time, but this is not necessary because it could better be accounted for by psychological continuity with any cause. It is not impossible that technology will one day allow us to transfer our relation of psychological continuity so that it obtains across other substrates or media.
- While the continuity of memory is one component of psychological continuity, persons with various defects of memory do not cease to be persons or necessarily lose their identity or sense of self.
- The sense of self is plastic *e.g.* it may be inflated as in mania, dissociated as in various dissociative disorders or induced by certain psychoactive substances, or bifurcated as in Dissociative Identity Disorder.
- Both personal identity and the sense of self can be affected by disease, surgical intervention, physical and psychological trauma and abuse, hypnosis and/or suggestion and chemical agents, especially hallucinogens.

References

BRENNAN, A. (1982) Personal Identity and Personal Survival. *Analysis* **42**(1): 44-50.
doi.org/10.1093/analys/42.1.44

BRENNAN, A. (1984) Survival. *Synthese* **59**: 339-61 (1984). doi.org/10.1007/BF00869339

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

- COLLINS, S. (1982) *Selfless Persons*. Cambridge University Press
- 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): 308-18. doi:10.1076/jhin.10.3.308.9086
- DENNETT, D. (1989) *The Intentional Stance*. The MIT Press
- DENNETT, D. (1991) *Consciousness Explained*. Little, Brown & Co. : Boston, MA. Paperback, 1993 Penguin Books, U.K
- 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
- JAYNES, J. (1976) *The Origins of Consciousness in the Breakdown of the Bicameral Mind*. Houghton Mifflin: Boston
- KWON, D. (2023) How the Brain Creates Your Physical Sense of Self. *Scientific American Newsletter* July 12, 2023
- LEWIS, D. (1978) Truth and Fiction. *American Philosophical Quarterly* **15**: 37-46
- LOCKE, J. (1695) *Essay Concerning Human Understanding*.
- LYU, D. *et al.* (2023) Causal evidence for the processing of bodily self in the anterior precuneus. *Neuron* 8 June 2023. doi.org/10.1016/j.neuron.2023.05.013
- MATURANA, H. (1970) Biology of Cognition. Reprinted In: *Autopoiesis and Cognition: The Realization of the Living*. D. Reidel Publishing Co.: Dordrecht , 1980: 5-58
- MARTONE, R. (2022) Creating Our Sense of Self. *Scientific American* **327**(2): 86-7
- 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
- NAGEL, T. (1986) *The View from Nowhere*. Oxford University Press
- NOZICK, R. (1981) *Philosophical Explanations*. Harvard University Press: Cambridge, Mass.
- PARFIT, D. (1984, 1987) *Reasons and Persons*. Ch. 10-13: 199-302. Clarendon Press, Oxford
- PARVIZI, J. *et al.* (2021) Altered sense of self during seizures in the posteromedial cortex. *PNAS* doi/full/10.1073/pnas.2100522118
- PEACOCKE, C. (1983) *Sense and Content*. Clarendon Press, Oxford

- PROUST, M. (1949) *Within a Budding Grove*. trans. C. K. Scott-Moncrieff. Chatto & Windus: London
- QUINTON, A. (1962) The Soul. *The Journal of Philosophy*, **59**(15)
- 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
- SADOCK, B. *et al.* Eds. (2017) Kaplan & Sadock's Comprehensive Textbook of Psychiatry, 10th Ed. Wolters Kluwer Health
- SCHREIBER, F. (1973) *Sybil*. Warner paperback
- SOLZHENITSYN, A. (1969) *The First Circle*. Bantam Books: New York
- SPERRY, R. (1964 *sic.*) 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
- STCHERBATSKY, T. (1919) 'The Soul Theory of the Buddhists'. *Bulletin de l'Academie des Sciences de Russie*, 1919
- STCHERBATSKY, T. (1923) The Central Conception of Buddhism. *Royal Asiatic Society, London*, 1923, p. 26.
- THIGPEN C. & CLECKLEY, H. (1957) *The Three Faces of Eve*. McGraw Hill
- WIGGINS, D. (1967) *Identity and Spatio-Temporal Continuity*. Basil Blackwell: Oxford,
- WIGGINS, D. (1976) Locke, Butler and the Stream of Consciousness, In Rorty, A. ed., *The Identities of Persons*. University of California Press: Berkeley
- WILLIAMS, B. (1970) The Self and the Future, *Philosophical Review* **79**(2): 161-180
- WILLIAMS, B. (1973) *Problems of the Self*. Cambridge University Press