3

INTRODUCTION TO PART THREE

Personal identity

Andrea Sauchelli

3.1 Criteria of personal identity

One of Parfit's motivations to discuss our beliefs about what we are is to further criticise the *Self-interest theory* (S), a theory of rationality holding that our rational aim is to bring about the outcomes that would be best for ourselves, and that would make our lives go, individually, as well as possible (R&P: 3).[1] In particular, Parfit suggests that since our identity over time is not always determinate and may hold as a matter of degree, an unqualified version of S is not plausible.[2] To motivate part of this argument, he explores several theories of personal identity, for example the *Cartesian Ego View*, the *Physical View* and the *Psychological View*. Parfit groups theories of personal identity into two families, that is, reductionist and non-reductionist views. The family of views favoured by Parfit (i.e., *reductionism,* a family of theories according to which the holding of the relation of personal identity consists in physical and/or psychological facts) can support a criticism of S on the basis that the special concern we almost instinctually have for, say, ourselves at an old age is at least not as important or deep as we think it is. Parfit's idea is that it is not irrational to regard our prudential concern for some of our 'future selves' as proportional to the degree of psychological connectedness we may have with them.[3] More on the notion of 'psychological connectedness' later. Becoming a reductionist should persuade us to revise not only our views on rationality but also on morality. For example, Parfit suggests that reductionism supports ethical theories that include prescriptions to care impartially and impersonally for the welfare of others (although it is compatible with other views).

Parfit's strategy to argue for these conclusions, especially those related to his views on personal identity, relies on a series of thought experiments.[4] For instance, consider:

> *The Teletransporter Case*. Suppose that you want to travel to Mars. Luckily, traveling from Earth to Mars takes only approximately 20 minutes now: a scanning device on Earth can record the exact molecular states of your body and brain and send this information to a receiver on Mars. Once the information is recorded and transmitted, your body and brain on Earth are destroyed. The device then recreates on Mars a perfectly functional replica of your body and brain. The new body and brain on Mars will remember the last thing that you did on Earth before the information was recorded (presumably, entering into the device) and will have your character traits, personality, interests, and so on. These facts may imply that your life, and not someone else's, will continue. Now, the device works exactly as expected: on Mars, you seem to continue the conversation you were having with your partner on Earth before the both of you were teletransported.
>
> *The Malfunctioning Teletransporter Case*. There is a new device that creates people on Mars without destroying the 'originals' on Earth. You try such a device, but it malfunctions: the transmission of your blueprint and creation of a new individual on Mars based on this information are successful but the scanning process has caused some damage to the body and brain on Earth. In a few days, the body and brain on Earth will cease functioning—an outcome you may think appropriate to describe as causing *your death*.

A few points about these cases. In both scenarios, the teletransportations are presented so as to involve causal processes. In other words, there is a causal connection between the body and brain on Earth and those on Mars—although such a connection does not imply that the body and brain on Earth are identical with the body and brain on Mars. Also, these scenarios presuppose that mental states and subsequent occurrences of mental states, for example memories, depend on at least physical states.[5] How should we understand and evaluate the previous scenarios? First of all, the kind of identity at issue is the numerical identity of persons through time, not their qualitative identity.[6] More specifically, Parfit is discussing which metaphysical criteria of personal identity correctly describe that part of reality involving the identity of numerically the same person through time, not the identity conditions of person-types. In this context, a criterion of personal identity tells us what personal identity consists in, that is, what is at issue here is a metaphysical criterion, not a criterion about how we know that personal identity holds (which would be an epistemological criterion).

Parfit claims that many would argue that the best description of the outcome of the second scenario is that *you* die. Why is that so? One intuitive answer is that you on Earth and the individual on Mars are not physically continuous and so you and the replica are different persons (i.e., you and the replica may be qualitatively identical without thereby being numerically the same). Although there is a causal process—we may even add that this causal process is usually reliable—the two individuals do not seem to have the same body or even a relevant part of it in common. This point may suggest that the identity of a person through time depends on one

form of physical continuity (e.g., brain continuity), or, at least, that physical continuity is a necessary condition of personal identity. According to what Parfit refers to as the best version of the *Physical View* (or *Criterion*), personal identity within t_1 and t_2 holds if and only if enough of the same functioning brain continues to exist within these times in a non-branching form (e.g., if the same brain continues to exist, is not split in half and both halves are not transplanted into different bodies within the period under consideration). Why is this version better than other forms of the *Physical View*? There is a sense of 'body' that does not imply that a body must be composed of exactly the same bodily parts to persist; for instance, if we regard it as an organism. If you are a yakuza, and have to make amends and cut one of your fingers, this does not mean that you or your body will cease to exist. Organisms gain and lose parts. That is one of the ways in which they survive. On this view, after having cut your finger, you have made amends (hopefully) and you still have or are the same organism (albeit slightly reduced). How many parts of your organism can you cut away and still be the same? A plausible reply is that perhaps the only thing you cannot cut away (or replace) without thereby ceasing to exist is your brain or, at least, those parts of your brain responsible for the functioning of some of your psychological activities or the functioning of your organism.[7] The parts that cannot be cut away without you thereby ceasing to exist—your brain— would be those the continuity of which, under the right circumstances, grounds personal identity through time, an idea captured by the best version of the *Physical View*.

A different type of theory is based on the idea that personal identity is essentially dependent on psychological features (the *Psychological View*). Parfit elaborates on John Locke's and Sydney Shoemaker's versions of this approach.[8] In particular, Parfit improves Locke's account in at least two ways: first, he distinguishes between *psychological connections* and *psychological continuity*, and, second, he more explicitly claims that various psychological states and connections in addition to memories are relevant to our identity over time (a point already made by Shoemaker). For example, one relevant connection can be that between an intention ('I want to read Frank Herbert's *Dune*') and the subsequent acting on this intention (my reading the novel). Other relevant psychological connections can be those between our character traits and their expressions in our behaviour.

Parfit defines psychological connectedness as 'the holding of particular direct psychological connections' and psychological continuity as 'the holding of overlapping chains of strong connectedness' (R&P: 206). Psychological connectedness can hold to different degrees, depending on the number of direct connections obtaining. For example, one direct psychological connection holds between you-yesterday and you-today if you-today remembers the experience of you-yesterday having dinner. For you-today to be *strongly* connected to you-yesterday, you-today should have at least half the number of direct connections that hold, over every day, in the lives of actual people—Parfit claims that this specific quantity of direct connections is simply a stipulative qualification of 'strong connectedness'. The relation of strong connectedness is not a transitive relation: from the fact that A at t_1 is strongly connected to B at t_2, and that B is strongly connected to C at t_3, it does not

follow that A at t_1 is strongly connected to C at t_3. For example, considering only memories for simplicity, you-today may remember much of what you-three-days-ago did, and you-three-days-ago may remember much of what you-one-week-ago did, but it is possible that you-today does not remember much of what you-one-week-ago did.[9] Parfit defines the *Psychological Criterion* as follows: P at t_1 is one and the same person as Q at t_2 iff P at t_1 is psychologically continuous with Q at t_2, in case such a continuity has the right cause and has not taken a branching form—a relation branches if for each relatum at a time, there is not at most one other relatum at any other time when the relation holds. There are various ways of specifying what a *right cause* is: on the *Narrow* version of this criterion, the right cause is the normal cause (i.e., continuity of the relevant parts of the body and brain responsible for the relevant psychological connections); on the *Wide* version, the cause only has to be reliable (e.g., a reliable teletransportation device); on the *Widest* version, the right cause can be any cause (e.g., an unreliable teletransportation device).

On the *Narrow Psychological Criterion*, it is possible that you may forget everything you did, say, in 2008 and still be the same person in 2018, provided that the changes in your character or memory are brought about by their normal causes—and that there are intermediate steps that form strong chains of psychological connections between you-in-2008 and you-in-2018. Note that this view differs from the brain-version of the physical criterion in that the latter claims that the continuity of enough of your brain is necessary *and sufficient* for your persistence as the same person over time. According to the *Narrow Psychological Criterion*, continuity of enough of a person's brain may be at most a necessary condition for this person's continuity as the same person. In a note added in 1985, Parfit claims that he has withdrawn his previous (1984) support for the *Wide Psychological Criterion* and, in the 1987 edition, he claims that we do not have to decide (or even that we *should* not try to decide) which version of the psychological criterion to adopt (R&P: x; 208–209). However, according to Parfit, there is a *family* of theories that we should accept: reductionism. Both physical and psychological theories as discussed above are forms of reductionism partly because they claim that the identity over time of persons consists in only the continuity of relevant bodily parts or psychological states (properly connected).

Let us elaborate on the distinction between reductionism and non-reductionism. In R&P, Parfit seems to draw the distinction in metaphysical terms, more precisely, in terms of the difference in the alleged metaphysical dependence or constitution of the metaphysical fact(s) of identity. More specifically, Parfit maintains that, for a reductionist, what constitutes personal identity over time is only facts regarding bodies, brains and (eventually) psychological states properly connected. Parfit says that a reductionist may also make a conceptual claim, namely, that facts of personal identity can be described in an impersonal way. If it is true that these facts can be described impersonally, then the related analysis of personal identity may not involve vicious circularity. In contrast, non-reductionists generally claim that our identity is not analysable and that it depends on a further particular and indivisible fact or entity. For example, according to the *Featureless Cartesian View*, such a

separately existing entity is a purely mental or spiritual substance the only essence of which is being conscious. This entity is said to exist 'separately' because its existence is supposed not to depend on or consist in the existence of physical and psychological facts. Some non-reductionists even argue that we *are* purely mental and indivisible subjects of experience and that we may know this solely by introspection. Parfit denies that he is introspectively or directly aware that he is an entity of that kind—and even if he were, it is not the case that he would thereby also know that such a purely mental entity is a *persisting* subject of experience. Even though Parfit rejects non-reductionism, he does not think that the concept of a Cartesian Ego is unintelligible—after all, it might have been true that such entities existed. His claim is rather that we do not have sufficient evidence for believing in the existence of such purely mental and indivisible entities.[10]

What is Parfit's preferred theory of personal identity in R&P? Parfit is a reductionist and, although he claims that we do not need to (or even should not) choose amongst reductionist theories, he consistently argued throughout his career that, if at all, personal identity only consists in physical and psychological facts or events—where the latter do not involve purely mental or spiritual substances. In addition, Parfit's version of reductionism also holds these two claims: (1) a person's existence consists solely in the existence of a body and brain, and the occurrence of several mental events, and (2) a person is an entity that is distinct from a body and brain, and the occurrence of several mental events. On this view, persons exist, but their existence consists in the existence of physical and psychological facts. So, although a person can be said to exist—we can even claim that a person *has* thoughts and desires, a body, a brain, etc.—a person is not an independently existing entity (i.e., an entity the existence of which consists in facts other than those which are physical and psychological).[11] As a nation can be said (a) to exist, (b) to be an entity that is distinct from its citizens and its territory, and (c) to be an entity the existence of which simply consists in the existence of its citizens living together in a certain way and a territory, people can be similarly said to exist in the way previously described. Given the kind of facts and entities our identity and persistence consist in, there can be cases where it is indeterminate whether a person is the same over time—Parfit also argues that 'only if we are separately existing entities, can it be true that our identity must be determinate' (R&P: 216).

3.2 An objection to the psychological criterion

Some non-reductionists have argued that memory connections must presuppose personal identity to count as proper memories. Today, John remembers having eaten an ice cream yesterday. Remembering such an experience—and not simply seeming to remember eating an ice cream—presupposes that John himself had the experience, or so the non-reductionist argues. The point can be generalised to all instances of experiential or episodic memory.[12] So, a certain kind of psychological connection (namely, experiential memory) that is essential to our persistence already presupposes personal identity and thus cannot be used in a non-circular

analysis of personal identity.[13] Parfit's defence of the *Psychological View* relies on (Shoemaker 1970).[14] In particular, against one understanding of the charge of circularity, Parfit claims that we can define psychological concepts (e.g., quasi-memory) in a way that does not presuppose the notion of personal identity. Such concepts can be incorporated into the *Psychological View* without thereby making it circular. More specifically, a quasi-memory of a past event involves (1) a person's seeming to remember such an event, (2) someone's having experienced such an event and (3) the person's current relevant mental state being causally dependent, in the right way, on the experience in question. The causal dependency that Parfit has in mind here is roughly in terms of memory traces, which are taken to involve (or be) modifications in a certain number of brain cells. Parfit also seems to believe that a memory trace, intended as a configuration of brain cells, can be replicated in different brains and produce identical or similar experiences—this presupposition is not essential to his other arguments. On this understanding of quasi-memory, it is possible that someone can quasi-remember 'from the inside' of an experience that was had by someone else. For example, suppose that Tonio has a conversation with Hans in Lubeck. Suppose that such an experience is recorded in Tonio's nervous system—it forms a memory trace. Years later, Lisaveta has the memory trace corresponding to Tonio's experience copied into her brain. Now Lisaveta can have a vivid apparent memory from the inside of what Tonio said to Hans—Lisaveta is quasi-remembering conversing with Hans. This does not imply that the conversation has for Lisaveta the same meaning or relevance as it had for Tonio. Rather, the point is simply that, granted that one is aware of having undergone such an implant, it does not seem to be contradictory to claim that it is possible to have non-delusory memories of experiences from a distinctive first-person perspective that have not been experienced by the person who is currently remembering them. What ensures that such memories are not delusions is that they represent real experiences and they are transmitted through an appropriate causal chain. This idea implies that it is not necessary that a memory is veridical only if such a memory presupposes personal identity over time.[15]

The next step is the incorporation of quasi-memory and other mental states (e.g., quasi-intending) into the *Psychological View*: Instead of saying that psychological connectedness involves memories, we should rather claim that, for example, psychological connectedness is determined also by overlapping chains of quasi-memories, quasi-intentions and so on. This statement completes what seems to be Parfit's favourite version of reductionism.

3.3 Is personal identity always determinate?

It is not necessary to believe in Cartesian Egos to claim that personal identity is always determinate. In particular, also some supporters of the physical criterion have suggested that although personal identity does not depend on a further fact, nevertheless personal identity is determinate. For example, Bernard Williams holds that we cannot even imagine or conceive of situations in which

it is indeterminate whether I personally survive—that is, situations in which it is indeterminate whether personal identity holds.[16] This claim about what is imaginable or conceivable, in turn, would make it difficult to believe that it is possible that personal identity be indeterminate. However, Parfit argues that even the bodily criterion cannot plausibly provide the required sharp cut-off point that would make personal identity always determinate. In addition, Parfit claims that only non-reductionism may imply such determinacy; however, since non-reductionism is false and that we should be reductionists, we should admit that there may be cases in which it is indeterminate whether personal identity holds.

Now, Williams's main purpose in (1970) was not that of showing that personal identity must always be determinate, rather, he wanted to argue that the physical criterion is better than the psychological one. In a nutshell, Williams holds that, in considering cases in which torture is presented to us as an inevitable future event, we would find little solace in being told that, say, prior to the ordeal, the torturer will also tamper with our brain and inscribe in it memory traces from the life of Napoleon. In general, our emotions and beliefs about our own future do not seem to align with beliefs regarding our psychological continuity. To the degree that our prudential concern for our future pain reveals the extension of personal identity, it seems that psychological continuity is not what determines our personal identity (or our attitudes towards our future). Parfit replies that Williams' reasoning seems to presuppose that questions regarding our persistence through time must always have a determinate answer (i.e., that in principle you could always affirm or deny that you exist at a specific time (as the same person)). This idea is questioned by appealing to various 'spectrum arguments'. First, Parfit re-describes part of Williams' argument in terms of what he calls the *psychological spectrum*. Call the sum of your psychological features relevant to personal identity your 'psychological profile'. Suppose that a surgeon can activate a series of switches working as follows. The flipping of switch s_1 causes a minimal change in your beliefs, memories and character traits. The change will involve only the insertion into your psychological profile of a few memories and one minor character trait that Napoleon had. The flipping of s_2 would cause another insertion so minimal that, in itself we would not regard it as sufficient to determine the end of a person. When all switches have been activated, you would eventually be psychologically similar to Napoleon—you will have his psychological profile. For Williams' argument to be a successful attempt to show that psychological continuity is not necessary for personal identity, we should rule out other evaluations of the case at issue. In particular, Parfit says that there are at least three ways of evaluating this case: (1) the two individuals at the end of the spectrum are numerically the same person, (2) there is always a sharp borderline between the two ends of the spectrum with respect to which person exists, and (3) the reductionist solution.

If we want to deny 1., Parfit and Williams seem to assume that we would need to claim that there is a sharp borderline between the two ends even in all of the central cases of the spectrum (a similar assumption is made in the evaluation of the other cases below). Now, option 2 is excluded as highly implausible—there is no

plausible candidate to play the role of this (non-conventionally established) cut-off point. The reductionist's solution (3) would be to say that there are cases in the middle of the spectrum where there is no answer regarding whether the resulting person would be you or not. However, according to Williams, we cannot make sense of this solution when it comes to, for example, the anticipation of a future agony; in fact, he believes that, on reflection, we have a strong intuition that it will be either me or not-me who will suffer. Besides, special concern for our future does not seem to become inappropriate simply because of small changes (say, one or two non-important memories). So, we may believe that 1. is the correct reply: at the end of the spectrum, the resulting person is you albeit with a different psychological profile—after all, you still fear the prospect of a future torture even after being told about the change of psychological profiles (R&P: 230).

If the problem of the supporter of the *Psychological View* with the previous reasoning is that she may have to admit the existence of sharp borderlines and that this is implausible, then, Parfit argues, this is hardly a problem only of the *Psychological View*. In fact, we can construct an analogous Physical Spectrum against the *Physical View*. In this scenario, switching s_1 will cause the replacement of 1 per cent of the cells in your brain and body with duplicates out of new matter, s_2 of 5 per cent and so on until a 100 per cent substitution. At the end of the spectrum, we would have a result similar to that for the case of teletransportation.[17] Is there in this scenario a precise threshold the trespassing of which determines the end of a person and the beginning of another? First, we may argue that it is plausible to say that the two persons at the end of the spectrum are the same—after all, they display the same behaviour, have the same interests, and so on. If plausible, this conclusion would show that the physical criterion is false (this is the counterpart of option 1 in the previous scenario). Alternatively (option 2), we may claim that there is a sharp borderline, but again this does not seem very plausible—how can a difference of just a few cells (or even an atom) determine the end of a body and the beginning of another? The last option is to argue that there are borderline cases in this spectrum—the reductionist's favourite solution. Still, we may be tempted to choose 1. over 3., since we may believe that the alleged psychological continuity between the two individuals can sustain personal identity.

The *combined spectrum* scenario includes a new twist aimed at showing that the reductionist's favourite solution is after all the best option. In this scenario, at the near end of the spectrum there are changes compatible with personal identity through time—the resulting person would be you in the same way in which, in your actual life, it is you who will wake up tomorrow with some minor difference in your psychological profile and your body. At the far end of the spectrum, there is someone physically and psychologically entirely different from you, say, Charlize Theron (I assume). Suppose that, by flipping a series of switches (s_n), a scientist can gradually substitute parts of your brain and body with corresponding parts of new organic matter modelled after Charlize Theron—for example, s_1 can replace a few bodily and brain cells, s_2 a few more and so on. In considering this case, Parfit claims that the counterpart of the previous option 1 is not plausible in this scenario: we

cannot sensibly claim that the resulting person (a replica of Charlize Theron) is you, as this person is physically *and* psychologically completely different from you. Since it is implausible to claim that there is a sharp borderline—after all, the differences generated by the activation of a switch are sufficiently small that it is hard to believe that the persistence of an adult organism or psychological profile depends on, for example, only one cell or memory—Parfit claims that our best option is the reductionist's. On this view, in at least some of the central cases of the spectrum, it is indeterminate whether the existing entity is you or Charlize Theron's replica. Parfit also claims that, according to the reductionist, asking whether the resulting person is you or not in the central cases may be an empty question in the sense that, once you know which physical and psychological facts hold, you know all there is to know. Whether the resulting person would be you in the borderline cases may have an answer, but such an answer may just be a stipulation about how to use or extend our language, in particular that part of our language involving personal identity. This extension, in turn, does not seem to be based on anything metaphysically deep.[18] Besides, Parfit says, such a stipulation cannot determine anything intrinsically or directly morally or rationally important (it may be important how we use our language in certain legal cases, but the importance of such stipulations do not derive from the facts personal identity consists in) or, at least, it may not have the importance we originally ascribed to the difference between identity and non-identity.

3.4 Divided minds

Parfit claims that 'recent' findings (recent in the 1970s and 1980s) further support the reductionist view. In particular, he discusses several cases that seem to put into question traditional views of the connections among minds, consciousness and persons. Parfit's discussion is indebted to Nagel (1971) because he deploys some of its points to argue in favour of reductionism as well as for other more radical theses.[19] First of all, it is (technically) possible to separate the two hemispheres of our brain by cutting a bundle of fibres that connect them—an operation (commissurotomy) that surgeons have performed to treat severe cases of epilepsy. Since each hemisphere controls one hand, one field of vision and several other specific abilities (e.g., linguistic abilities), it has been possible to study the effects of this 'brain splitting' operation. In particular, it seems that, in certain cases, we have evidence to conclude that the right hemisphere, which controls the left field of vision and the left hand, may perceive and express awareness of things taking place on the left side of the body in a way that can be insulated from what can be separately perceived by the other hemisphere. In addition, the expressions of awareness of one hemisphere may differ from what the other hemisphere might express; for example, one hemisphere may report, through the use of the related hand, that one object is in its relevant field of vision while the other hemisphere might report only the existence of a different object on the other side of the field of vision. This situation can be described as involving a person with two streams of consciousness, each unaware of the other.

Parfit suggests that, although our brains tend to specialise during development—that is, certain areas of the brain become specialised in performing certain tasks (e.g., the elaboration of language)—it is possible that, in some individuals, their mental abilities can be equally distributed across the two hemispheres. Now, suppose that this last possibility is true of you and imagine that you are taking a difficult physics exam in an isolated room. After having read the exam, you think that there are two different ways of solving the problem in it, but you do not know which one is correct or better. Fortunately, you have a device that allows you to momentarily split the two hemispheres of your brain, a device you can activate almost at will. You split your brain and the result is two different streams of consciousness, each separately aware of working on one way of solving the problem. We might even suppose that, in addition to the splitting device, you also have a memory trace-reduplicator—perhaps installed to amplify the vividness of your memory or as a back-up if half of your brain is compromised—that produces memory traces of your short- and long-term memories in both hemispheres (or, at least, of those memories that will make a contribution to the decision-making systems of your brain). After some time, the designated hemisphere activates the device and reunites your brain. Suddenly, you seem to remember having worked on two different calculations. While your brain (and mind) was divided, each single stream of consciousness had the feeling of being the same person before the split, apparently including the memory of deciding to split your brain to work on two separate ways of solving the problem and how to proceed in this regard. Parfit suggests that, since the brain-division is brief, reversible, completely under your control or at least under the control of one designated stream, and that each separate stream's experiences are accessible to the main stream after unification, it is plausible to claim that, in the context of this scenario, there is only one person in the room. Some have argued that each stream of consciousness is unified by a subject of experience and that each of us is a subject of experience. However, this idea is not appealing to those who claim that, in the physics-exam scenario, you had two different streams of consciousness: how can you be the *unifying* subject of both if at any time when your mind was divided each stream was not conscious of at least one of the experiences of the other stream? Perhaps we should say that when your mind was divided, there were two different subjects of experiences. We might even have to claim that the case in question involved at least three subjects of experience: the owner of the pre-division stream of consciousness (a person), and the two subsequent owners of the two different streams causally connected to the first one. A fourth individual can be added: the subject after the unification of the streams. Parfit claims that, on this interpretation, we may have to believe that the life of a person might involve two non-person subjects of experience. However, when you divide, each stream feels continuous with the allegedly different previous subject of experience—according to our scenario, each stream of consciousness need not be interrupted or experience disunity with the antecedent experiences. Explaining this unity of consciousness by saying that these experiences are experienced by the same subject that is not you seems implausible—after all, each post-division stream 'feels' that it was part of the pre-division stream. How can

this new subject not be you if the unity of each stream of consciousness is explained by the existence of one and only one corresponding subject of experience?

At this point, the reductionist may claim that there is a simpler and better explanation. In particular, the reductionist can argue that the unity of consciousness in the case of division is explained by the fact that there are two single states of awareness that 'connect' certain thoughts, perceptions and feelings. When the mind is divided, there is no single state of awareness that spans over both sets of experiences at the same time. In addition, the reductionist can say that, in the case at issue, there are two streams of consciousness each *co-conscious* of appropriately connected mental states and that this is all there is to say with regard to their unity. The unity of consciousness need not be explained in terms of a unifying subject, but can simply be explained in terms of the relations that connect the different mental states in each stream. In Parfit's words, 'a particular mental event occurs within some life in virtue of its relations to the many other mental and physical events which, by being interrelated, constitute this life' (R&P: 252).[20]

3.5 Personal identity does not matter

An extension of the previous thought experiment is the following. Suppose that you have an identical twin and that your body and your twin's brain have been fatally damaged. Surgeons connect your brain to the nerves of your twin's body successfully. Who is the person that wakes up after the operation? Both supporters of the *Psychological View* and of the brain version of the *Physical Criterion* may agree that the resulting person is you. Notice that you could have survived had only half of your brain survived the operation: after all, we believe that people who had a stroke that damaged an entire half of their brain can survive. David Wiggins gave this thought experiment a further twist.[21] Suppose that you have another sibling with a body similar to yours, that your brain has been divided into two halves, and that each half is separately transplanted into one of your siblings' bodies. Both resulting people will be psychologically continuous with you and believe to be you.[22] What happens to *you* after this person-splitting (or fission) case? (1) You do not survive, (2) You survive as one of the two people, (3) You survive as the other, or (4) You survive as both. Against 1., Parfit claims that since you would survive had only half of your brain been successfully transplanted, it is not plausible to hold, at the same time, that you do not survive if the other half were also successfully transplanted. In Parfit's words, '[h]ow could a double success be a failure?' After all, the relation between you and each of the resulting people is the same. Options 2. and 3. are implausible: given the relevant facts, there is nothing that would seem to make you one of the resulting individuals rather than another—and thus regarding 2. (or 3.) more plausible than 3. (or 2.). On behalf of 4., we may say that describing the two resulting individuals as different persons is misleading: since Parfit admitted that a person can have a divided mind, perhaps the two resulting individuals are parts of one person with a doubly-embodied divided mind, that is, you may claim that you survive as both. However, Parfit claims, this supposition would involve too

great a distortion of our common concept of a person—particularly when we also assume that the two individuals may decide to each go their separate ways and lead different lives, perhaps in two different parts of the world. Among the problems of this possibility for our concept of a person, claims of moral responsibility and other attitudes we think as connected to the concept of a person may, over time, be applicable only to one of the resulting individuals.

Parfit suggests that the reductionist has a better solution. In particular, the reductionist would say that, in a sense, the different options above do not really describe different possibilities. In knowing that each individual will have half of your brain and will be psychologically continuous with you, you know everything there is to know with regard to the identity of the individuals in question—and adding that, for example, it is *you* who survives would not add anything about reality. There is a better description of the outcome of the thought experiment, which Parfit claims to be that neither of the resulting people will be you, but this is not a description that adds anything to reality, at best it indicates what we regard as the most reasonable way of extending our concept of personal identity. In a sense, which option we choose (that is, which way we decide to describe the outcome) may simply be a linguistic or conceptual stipulation. Parfit maintains that a more important question—more important than determining how to apply the concept of personal identity (and related concepts or terms, such as 'you') to this case—concerns how you should regard, from a moral and rational point of view, your division. For example: Should you be prudentially concerned with the fate of these individuals?[23]

3.6 What matters when you divide?

Parfit maintains that, when evaluating the case of your division, it would be irrational to regard the outcome as bad as death, even assuming that we describe the outcome as one in which you are not personally identical to any of the resulting individuals. The main reason is that the relations holding between you and each of the two different resulting people do not fail to individually contain anything important that is not included in your ordinary survival. If it is rational to have prudential concern for yourself, it may also be rational to be prudentially concerned for the two products of your fission. After all, each relation to the corresponding resulting person contains what we would independently regard as prudentially mattering.

Certainly, a case of double survival is not similar to ordinary survival, but it does not mean that it does not contain what matters in normal survival. Rather, Parfit claims, what is important is relation R, where 'R is psychological connectedness and/or continuity, with the right kind of cause' (R&P: 262).[24] Parfit also suggests that, in this context, the 'right kind of cause' is *any* cause. In our world, relation R normally has a non-branching form, that is, R connects no more than one person existing with another at a particular time, whereas, in the your-division case, relation R has a branching form—it holds between you and two different individuals,

both existing at the same time—hence, there is not a single future individual with whom you are identical.

Some (e.g., Susan Wolf) objected that, far from being appealing, branching-R scenarios would have horrifying consequences.[25] In particular, according to Wolf, branching would disrupt some of the things we value most, such as some forms of exclusive personal relationships. In R&P, Parfit seems to address this issue by saying that the uniqueness of the holding of R may have some extra value, but it can make at most *some* difference—perhaps a difference due to the psychological profile of the dividing person. However, Parfit claims that, on reflection, even such an individual should not rationally regard the prospect of division as bad as death. A different reply, for example the one discussed in (Parfit 2007), draws our attention to the distinction between questions of *prudential* concerns from questions regarding what makes survival *desirable*. The point of the case of your division is to discuss whether it would be irrational to extend to the R-branching individuals the kind of prudential concern that a self-interest theory would prescribe you to have for yourself. Extending such a concern does not imply that such an extension would be always desirable—for example, you may be prudentially concerned for yourself even though you think that your future will not be desirable. Questions of the desirability of a future may not coincide with questions of prudential concern for a future. So, in evaluating your division with regard to the rationality of prudential concern for R-branching individuals, we would better suppose that their future lives are as desirable as the one the pre-branching individual is having. Wolf's worries may thus not be all relevant to the case at issue.

What are some of the consequences of accepting the previous arguments about the non-importance of personal identity? One is this: instead of saying 'I shall be dead', you may rather say that there will not be any future experiences related in certain ways to the experiences you are now having. This idea may bring you some solace at the perspective of incumbent death: death does not involve the irremediable destruction of something metaphysically deep. Rather, it should be seen as the interruption of a causally interrelated flow of experiences and bodily functions.

If we hold that what matters is both R *and* physical continuity, our position is equivalent to saying that what matters is R with its normal cause since physical continuity (at least in its brain-continuity version) is part of R's normal cause. However, Parfit claims that the continuity of your brain should not matter if it is not the carrier of relation R. Brain-continuity can be part of what matters, but its importance is only derivative. For, Parfit claims, our desire (if at all) to have a body or brain similar to those we have is not the desire to have a particular body. In fact, a functionally equivalent body—functionally equivalent in its capacity to sustain strong psychological connectedness—should, on reflection, suffice. Our 'old brain and body' may have sentimental value to us, so it may not be entirely irrational to care about their persistence, but such a care should not obfuscate what really matters or why this brain and body started to matter at all: the value of what they sustain.

In considering whether both psychological connectedness and continuity matter, Parfit seems to believe that caring about only psychological continuity may involve

deep regret (R&P: 301). It is conceivable that, given sufficiently long periods of time, we may have psychological continuity between A and B with very few psychological connections between them; however, Parfit argues, most of us would regret the loss of some of our long-past memories. A similar claim can be made about the persistence of other psychological features such as intentions and desires. Although you may regard psychological fluctuations and changes within the life of a person as compatible with psychological continuity, a significant number of them would reduce our psychological connectedness over time. Parfit's conclusion is that psychological continuity certainly matters, but it is not true that it is the only relation that matters: psychological connectedness matters as well.

3.7 Personal identity and rationality

How is the previous discussion on personal identity connected to Parfit's criticism of S? According to S, for each person, there is one supremely rational ultimate aim, namely, that things go as well as possible for her. Some (e.g., Joseph Butler) have argued that if reductionism is true, we have no reason to be concerned about our own future (*the Extreme Claim*). If this claim and reductionism were both true, we would have a direct refutation of S. However, Parfit maintains, a reductionist can also plausibly argue that relation R may give us some reasons for special concern (the *Moderate Claim*). Perhaps a better argument against S can be devised. This new argument is based on the idea that what matters are psychological connectedness and continuity. Another crucial premise is a central point of S, *the Requirement of Equal Concern*: a rational person should be equally concerned about all parts of their future.[26] Now, Parfit claims that there may be other reasons to support the idea that it may not be irrational to care less about one's future other than simply because it is in one's own future. In particular, Parfit holds that one of these reasons is that our concern for our future,

> may correspond to the degree of connectedness between me now and myself in the future. Connectedness is one of the two relations that give me reasons to be specially concerned about my own future. It can be rational to care less, when one of the grounds for caring will hold to a lesser degree. Since connectedness is nearly always weaker over longer periods, I can rationally care less about my further future.
>
> *(R&P: 313)*

Although in general it is not true that to a decrease of the degree to which a relation holds we can always rationally believe that such a relation has less importance, it may not be irrational to care less about our own future because there is a reduced degree of psychological connectedness.[27] Hence, Parfit argues, it may not be irrational for you to care less when there will be much less connectedness, given that such a connectedness is one of the reasons for having prudential concern. For example, suppose that you know that you will have a painful day tomorrow and

one in 30 years. You-now is likely to be more strongly psychologically connected to you-tomorrow than to you-in-30 years. Since connectedness is one of the relations that justifies prudential concern, you cannot be irrational to comparatively care less when there is less connectedness. One consequence of—some may say 'an objection to'—the previous reasoning is that, if Parfit's view is correct, it might not be irrational to be greatly imprudent, for example, it may not be irrational to be an imprudent youth. However, we should criticise great imprudence. Parfit's reply to this objection is that we should include the issue of imprudence in the moral territory—that is, condemn great imprudence from a moral point of view. For example, we might appeal to one form of consequentialism, in particular, to a version including an impartial or agent-neutral principle of beneficence. On this version of consequentialism, if your action now will overburden you-old, you are increasing the overall sum of suffering. This increase of suffering is morally wrong, independently of whether it is you or an individual about whom you may rationally prudentially care less than yourself who will suffer as a result of your action. So, great imprudence can still be criticised. In addition, we may also claim that you stand in a special relation of duty or obligation to your future self, akin to the relation you may have with your children, pupils etc.[28] In turn, this obligation can provide further ground to criticise imprudence.

3.8 Personal identity and morality

Reductionism gives us reasons to change some of our moral beliefs. Among the claims or consequences of reductionism that can give us such reasons, Parfit focuses particularly on: (1) personal identity over time may not be a deep truth or fact, and (2) becoming a person and personal identity over time are a matter of degree.[29] In evaluating the morality of early cases of abortion (e.g., a few days after fertilisation), the reductionist may believe that at the beginning of a pregnancy there may be nothing seriously wrong with a voluntary abortion, as it would not amount to the destruction of an existing person, and that having an abortion may become, *ceteris paribus*, gradually wrong as time passes and the foetus gradually nears personhood.

Some non-reductionists have maintained that only if personal identity is a further fact or involves a deep metaphysical truth, can we justifiably maintain a series of important moral and legal practices (the moral counterpart of the *Extreme Claim* for rationality). For example, non-reductionists argue that if reductionism is true, we would not be justified anymore in punishing people for their past crimes (at least in certain cases). The reason would be that only if non-reductionism is true can we truly say that a person is identical through time, and thus that the same person who committed a crime can be subsequently punished for it (a necessary requirement for punishment). Parfit suggests that a different claim is equally defensible, namely, that also psychological continuity may carry a certain degree of moral responsibility; hence, reductionism is compatible with some of our current beliefs about the scope of, for example, punishment and responsibility. If we also consider psychological connectedness as having moral and practical relevance, we may

also think that punishment should be proportional to the degree of psychological connectedness between, for example, the person at the time of the crime and the person at the time of conviction (or even later in their life). This conclusion seems plausible.

There are other consequences of reductionism for morality, for instance, with regard to compensation and distributive justice. To see the connection between reductionism and these issues, we need to introduce the notion of the separateness of persons—roughly, the idea that each person has a separate, individual life to live and that this fact is normatively relevant.[30] Versions of this notion seem to have influenced important figures in the history of ethics, including some utilitarians. For example, Henry Sidgwick claims that there are two rational aims, S and Benevolence. In particular, Benevolence says that it is rational to desire or act such that 'things go, on the whole, as well as possible for everyone' (R&P: 329). Sidgwick then specifies the requirement of benevolence in terms of his Hedonism; more specifically, he suggests that 'our ultimate moral aim is the greatest net sum of happiness minus misery' (R&P: 330). On his version of utilitarianism, what morally matters is the amount of happiness and suffering—how this distribution is performed among people makes no moral difference insofar as the utility is maximised. Benevolence may not be compatible with S and Parfit suggests that Sidgwick accepted the possibility of an irresolvable conflict between these two principles also because he believed in the separateness of persons. Given the separateness of persons, it may be rational to hold either of the two principles. Some deontologists also appealed to the separateness of persons. For example, John Rawls, Thomas Nagel and others variously objected to utilitarianism that, in addition to its principles of impersonal distribution, we should also introduce other principles of distributive justice the moral importance of which is not reducible to their utility and that are based on the separateness of persons. Now, utilitarians need not disregard distributive principles; rather, they may regard these principles (e.g., of equal distribution) as mere means to maximise overall or average utility, not as aims in themselves. Still, critics of utilitarianism have complained that this way of reasoning ignores 'the boundaries between lives' and their differences.[31] Ignoring such boundaries would be wrong because, among other things, the relationship between (i) one person's selves and those of other people and (ii) the selves of a single person, is radically different and incommensurable; for example, only the holding of the latter relationship can automatically compensate an individual for what she did at an earlier time. As David Brink noted in a similar context, 'diachronic, intrapersonal compensation is automatic; interpersonal compensation is not.'[32]

Parfit claims that there are different ways in which reductionism can be combined with our preferred principles of distribution; for instance, reductionism may justify changes to the scope or to the weight of these principles. With regard to the first option, a reductionist may change the scope of the principles at issue (their 'units of moral concern') from entire lives to selves or even to people's states at particular times.[33] In particular, Parfit suggests that reductionism supports regarding the subdivision within lives as, in certain ways, like the divisions between lives (since the

fact of non-identity is not a deep metaphysical truth). As a consequence, a reductionist may apply distributive principles even within lives. For example, in a situation in which a child (the present self of a person) will be burdened with some hardship for a benefit in his adulthood (a future self of the same person), the reductionist may claim that such a distribution would be as unfair, or at least similarly unfair, as the allocation of a burden on a person to benefit someone else: If all the future selves of a person are, in certain ways, like selves of other people, the reductionist may regard trade-offs within lives as potentially unfair as trade-offs between lives (R&P: 333–334). However, the reductionist can also suggest that, although the distributive principles should have a wider scope, their weight should be correspondingly reduced (R&P: 334–345). For example, if distributive principles are based on the separateness of persons (the alleged non-identity between people), the reductionist may argue that since there are cases in which intrapersonal and interpersonal boundaries are not metaphysically deep, we should give less moral weight to principles such as that of compensation in these cases. Parfit's main point is that if we accept reductionism and come to believe that persons are not separately existing entities, the fact of their non-identity should seem less important and 'it becomes more plausible to be more concerned about the quality of experiences, and less concerned about whose experiences they are' (R&P: 346). On the reductionist view, the impersonality of utilitarianism towards, for example, suffering—roughly, the idea that it is not morally important *who* is suffering but the negative quality of the experience—is more plausible. In general, Parfit's lesson here is that a change in our beliefs about what we think we are and how we persist through time has significant consequences for our beliefs about rationality and morality.

Notes

1 I use 'R&P' for references to *Reasons and Persons* (1984/87).
2 I here assume that personal ontology and the metaphysics of personal identity coincide. More recently, philosophers have argued that an answer to the question 'What are we?' is not immediately answered by an account of personal identity. For instance, some have suggested that we are entities that are not always persons (e.g., we are human organisms). See Olson (2007) and Sauchelli (2018: 13–14). Theories of personal identity are here understood as aiming at providing at least the synchronic and diachronic conditions of identity of persons.
3 Parfit employs the terminology of 'selves', 'future selves', etc., but also claims that it is simply a manner of speech. A self is merely a unit of particularly interconnected psychological connections, but is not an independently existing entity.
4 Some scholars have objected in various ways to the thought experiment methodology. I do not have the space to properly address their arguments but I will assume that thought experiments—even those that investigate the application of certain familiar concepts to situations alien to the 'normal application of such concepts'—are still useful. (R&P: 200). See Wilkes (1988) for criticism.
5 Parfit does not specify what he regards to be the metaphysical nature of mental states or of their content. Some philosophers of mind may argue that some of the presuppositions of his thought experiments are questionable on the basis of unstated internalist presuppositions on the metaphysics of content. See Burge (2003).

6 Two things may be qualitatively identical without thereby being numerically identical—or so it is generally assumed. Besides, one thing can be numerically the same at a later time without thereby being qualitatively identical to itself in the past (e.g., a baby growing into an adult).
7 See Olson (2007).
8 Locke (1694) and Shoemaker (1970). See also Shoemaker (1997) for a later development of his view.
9 See section 2 for a qualification of the kind of psychological connections Parfit has in mind.
10 Parfit has discussed and partially modified the distinction between reductionism and non-reductionism in several other publications after R&P. See Parfit (1999) and (2007).
11 This version of reductionism, a slightly different version of which Parfit would later call *Constitutive Reductionism*, resembles Shoemaker's and Baker's theories. See Parfit (1999), Shoemaker (1984), (1997) and Baker (2000). More recently (2012), Parfit proposes a different approach (although he says that he has not rejected his previous version of constitutive reductionism).
12 Psychologists and philosophers have distinguished a variety of kinds of memory (episodic, semantic, etc.). See Bernecker (2010: 11–45) for an extensive review.
13 This objection has been understood in other different ways, at least starting from Joseph Butler's and Thomas Reid's criticisms of Locke (Sauchelli 2018: 110–112).
14 In turn, Shoemaker's account is indebted to Martin and Deutscher (1966). In a later work, Shoemaker claims that Parfit's example in R&P of an episode of quasi-memory is not convincing. See Shoemaker (2004: 581).
15 See Schechtman (1990) and McDowell (1997) for further criticisms. Parfit (1999) contains a reply to McDowell.
16 The actual claim made by Williams is subtler than this. In particular, his point is that it is of little help in settling our emotions to know that there are indeterminate cases in our future when we project or imagine how such a future would be like—under the presupposition that we know that we will have to undergo such a future. See Williams (1970/73: 58–59).
17 Some may argue that it makes a difference whether the change is gradual and continuous (both in space and time) or whether the original brain and body stop functioning (even for a very short time). In fact, we can argue that if the change is gradual and does not involve any interruption of function, we may dispute Parfit's claim that at the far end of the spectrum your 'brain and body are completely destroyed' in the manner in which the teletransporter does. See Unger (1990: 123–125) and McMahan (2002: 70–72).
18 It is not always clear what the concept of 'metaphysically deep' is. One way of understanding it is that a difference is not metaphysically deep with respect to K if it is a difference that we do not generally regard as significant to determine the belonging of an entity to a kind K (whether natural or not).
19 Parfit cites some early works by Roger Sperry. See Gazzaniga (2005) on the topic.
20 Parfit calls this view 'the bundle theory' of self in Parfit (1987).
21 Wiggins (1967: 53–55). Parfit says in various places that reading Wiggins's example and the Brown-Brownson thought experiment in Shoemaker (1963) greatly influenced his choice of studying philosophy.
22 We may further add that the two 'receiving bodies' are equally similar with yours with respect to the relevant facts (e.g., if one of your character traits is inherently connected to certain features of your kind of body and a certain physical appearance, we can stipulate that such a connection is maintained after the operation).

23 The thesis that personal identity does not prudentially matter has generated considerable debate (e.g., Lewis (1976/83), Brueckner (1993), Parfit (1993) and Johnston (1997)). Recent instances are Johansson (2010) and Gustafsson (2018).
24 In a more recent paper, Parfit recognises that this formulation contains an inaccuracy. In particular, he should have said that what matters is non-branching R, where R is psychological connectedness *and* continuity (Parfit 2007: note 30).
25 See Wolf (1986).
26 See also Sidgwick (1907/81: 124, n. 1).
27 It is not entirely clear to me whether Parfit has successfully addressed the objection discussed at (R&P: 314).
28 See Whiting (1986) for relevant considerations.
29 We should distinguish two claims, namely, that (1) it can be indeterminate whether an entity at a time is a person, and (2) it can be indeterminate whether P at t is the same person as Q at t*. Parfit seems to accept both (R&P: 321–323).
30 Norcross (2009) provides an interesting (non-sympathetic) attempt to understand this notion.
31 'To sacrifice one individual life for another, or one individual's happiness for another's is very different from sacrificing one gratification for another within a *single* life' (Nagel 1970: 138).
32 Brink (1997: 108).
33 In replying to an objection raised by Bart Schultz, Parfit seems to suggest that, given reductionism and the argument on compensation, the relevant units of moral concern—the moral units that should be regarded as the proper subjects of distribution, etc—would better be selves (temporal parts or stages of people, see note 3). See Schultz (1986) and Parfit's reply at (1986: 840). Parfit clarifies his claim in Appendix H (1986). See Shoemaker (1999) for discussion.

References

Bernecker, S. 2010. *Memory*. Oxford: Oxford University Press.
Brink, D. O. 1997. Rational egoism and the separateness of persons. In *Reading Parfit*, ed. J. Dancy, 96–134. Oxford: Blackwell Publishers.
Bruekner, A. 1993. Parfit on what matters in survival. *Philosophical Studies* 70, 1: 1–22.
Burge, T. 2003. Memory and persons. *Philosophical Review* 111, 3: 289–337.
Gazzaniga, M. 2005. Forty-five years of split-brain research and still going strong. *Nature Reviews Neuroscience* 6: 653–659.
Gustafsson, J. 2018. The unimportance of being any future person. *Philosophical Studies* 175: 745–750.
Johansson, J. 2010. Parfit on fission. *Philosophical Studies* 150, 1: 21–35.
Johnston, M. 1997. Human concerns without superlative selves. In *Reading Parfit*, ed. J. Dancy, 149–179. Oxford: Blackwell.
Lewis, D. 1976/83. Survival and identity. Reprinted in his *Philosophical Papers, Volume I*, 55–77. Oxford: Oxford University Press.
Locke, J. 1694/1971. *An Essay Concerning Human Understanding. 2nd edition*. Oxford: Clarendon Press.
Martin, C. B. & M. Deutscher. 1966. Remembering. *Philosophical Review* 75, 2: 161–196.
McDowell, J. 1997. Reductionism and the first person. In *Reading Parfit*, ed. J. Dancy, 230–250. Oxford: Blackwell.
McMahan, J. 2002. *The Ethics of Killing*. Oxford: Oxford University Press.

Nagel, T. 1970. *The Possibility of Altruism*. Princeton: Princeton University Press.
Nagel, T. 1971. Brain bisection and the unity of consciousness. *Synthese* 22: 396–413.
Olson, E. 2007. *What Are We?* Oxford: Oxford University Press.
Parfit, D. 1984/87. *Reasons and Persons*. Oxford: Clarendon Press.
Parfit, D. 1986. Comments. *Ethics*, 96: 832–872.
Parfit, D. 1987. Divided minds and the nature of persons. In *Mindwaves*, eds. C. Blakemore & S. Greenfield, 19–28. Oxford: Blackwell.
Parfit, D. 1993. The indeterminacy of identity. *Philosophical Studies* 70, 1: 23–33.
Parfit, D. 1999. Experiences, subjects, and conceptual schemes. *Philosophical Topics* 26, 1–2: 217–270.
Parfit, D. 2007. Is personal identity what matters? *Marc Sanders Foundation*, 31 December. Retrieved at: www.marcsandersfoundation.org/wp-content/uploads/paper-Derek-Parfit.pdf
Parfit, D. 2012. We are not human beings. *Philosophy* 87, 1: 6–28.
Rawls, J. 1971/99. *A Theory of Justice. Revised Edition*. Cambridge, MA: Harvard University Press.
Sauchelli, A. 2018. *Personal Identity and Applied Ethics*. London & New York: Routledge.
Schechtman, M. 1990. Personhood and personal identity. *Journal of Philosophy* 87, 2: 71–92.
Schultz, B. 1986. Persons, selves, and utilitarianism. *Ethics* 96, 4: 721–745.
Shoemaker, D. 1999. Selves and moral units. *Pacific Philosophical Quarterly* 80: 391–419.
Shoemaker, S. 1963. *Self-Knowledge and Self-Identity*. Ithaca, NY: Cornell University Press.
Shoemaker, S. 1970. Persons and their pasts. *American Philosophical Quarterly* 7, 4: 269–285.
Shoemaker, S. 1997. Self and substance. *Philosophical Perspectives* 11: 283–304.
Shoemaker, S. 2004. Brown-Brownson revisited. *The Monist* 87, 4: 573–593.
Sidgwick, H. 1907/1981. *The Methods of Ethics, 7th edition*. Indianapolis, IN: Hackett Publishing.
Unger, P. 1990. *Identity, Consciousness, and Value*. Oxford: Oxford University Press.
Whiting, J. 1986. Friends and future selves. *Philosophical Review* 95, 4: 547–580.
Wiggins, D. 1967. *Identity and Spatio-Temporal Continuity*. Oxford: Blackwell.
Wilkes, K. 1988. *Real People*. Oxford: Clarendon Press.
Williams, B. 1970. The self and the future. *The Philosophical Review* 79: 161–180.
Wolf, S. 1986. Self-interest and interest in selves. *Ethics* 96, 4: 704–720.