Morality and Personal Identity

Torbjörn Tännsjö

(Gothenburg University)

1. INTRODUCTION

The problem of the nature of our personal identity may seem to belong exclusively to metaphysics. However, if one inspects the actual discussion about the notion of personal identity, one gets the impression that the problem does not only have a metaphysical aspect, but also a moral one. Mainly, there have been two ways of conceiving of the connection between metaphysics and morality. On the one hand, some thinkers have come to believe that the nature of our personal identity is of relevance to morality. Our beliefs about who we are may rationally influence our moral outlook. The most outspoken proponent of this view is Derek Parfit.[1]

On the other hand, some thinkers have suggested that there exists no exclusively metaphysical question as to who we are. The notion of personal identity is a moral notion or, to borrow the words of John Locke, a ”forensic” one. On this view, there is a connection between metaphysics and morality, but it goes the other way round. Our moral beliefs ought to determine our beliefs about who we are. The most influential version of this view has been stated, of course, by Immanuel Kant who taught that, in order to rescue the notions of moral responsibility and free will, we had to hypothesise a ”noumenal” self.

Though rarely held, a third position is obviously possible as well. Metaphysics and the notion of personal identity is one thing and morality quite a different thing. Our metaphysical views about personal identity and our moral views can be rationally held, quite independently of each other.

In the present paper I investigate the connections between metaphysics and morality. The point of departure of my discussion are the views put

---

[1] Cf. mainly his *Reasons and Persons* (Oxford: Clarendon Press, 1984), but also his subsequent paper, ”The Unimportance of Identity” in Henry Harris (ed.), *Identity* (Oxford: Clarendon, 1995).

forward by Derek Parfit in Part III of his *Reasons and Persons* and the discussion that has followed over the last decade upon the publication of this book. Some of the most important contributions to this discussion can be found in Harold Noonan's (ed.), *Personal Identity*.[2] A review of recent work on personal identity is also given by James Baillie.[3] And, eventually, the *Reading Parfit* volume, edited by Jonathan Dancy, has seen the day of light.[4] So time is ripe for a critical assessment of Parfit's contribution to this classical philosophical problem.

The upshot of my investigation is a defence of the view that personal identity is a moral notion. In the present context, metaphysics and moral philosophy cannot be pursued in isolation from each other. However, Parfit to the contrary notwithstanding, moral considerations ought to determine our view of personal identity, not the other way round.

2. THE DIFFERENT POSSIBILITIES

Which are the most important views of personal identity?

In his book, *Reasons and Persons*, Derek Parfit distinguishes roughly between two main alternatives, a reductionist view of personal identity and a non-reductionist view. Then there are various different versions of both a reductionist and a non-reductionist variety. There are, for example, reductions of personal identity to mental as well as to physical entities, and to combinations of these. And there are various different ways of thinking of the notion of a person as not "reducible".

This seems to me to be to simplify things too much. For, to be sure, in the discussion, a lot of other distinctions seem to play a crucial role. In particular I think of the distinction between the view of personal identity as something that is or is not always determinate and the distinction between the view of personal identity as a "deep" or a "shallow" fact.

One reason that Parfit simplifies the discussion is that he believes that a reductionist view implies that there may be cases where personal identity is indeterminate, while a non-reductionist view does not imply this:

---

[2] Harold Noonan (ed.), *Personal Identity* (Aldershot: Dartmouth, 1993).

[3] James Baillie, "Recent Work On Personal Identity", *Philosophical Books*, No. 2, Oct. 1993, pp. 193-205.

[4] Jonathan Dancy (ed.), *Reading Parfit*, Oxford: Clarendon, 1997)

> If we accept a Reductionist View, there may be cases where we believe the identity of such a thing to be, in a quite unpuzzling way, *indeterminate*. We would *not* believe this if we reject the Reductionist View about this kind of thing.[5]

But first of all, it is far from clear what kind of indeterminacy is intended in the quoted passage. I will return to this problem. Furthermore, it is not at all clear why *all* reductionist views should allow this kind of indeterminacy. And it should not simply be taken for granted that a non-reductionist view could *not* allow it.

Another reason that Parfit simplifies the discussion is that he believes that, if reductionism is a correct view of personal identity, then personal identity is not a ”deep further fact”.

I agree that, if reductionism is correct, then personal identity is not a *further* fact, i.e. a fact further than what it has been reduced to (whether this be something mental or something physical or some definite combination of both), but I question the claim that this means that personal identity is not a ”deep” fact. Whatever this may mean, and the problem of its meaning in the present context will be the main focus of this paper, I find it obvious that, for all we know, depth may coexist with a reductionist view.

We have to make more finely grained distinctions, then. I will do so by addressing, in order, the problem what it means to ”reduce” a notion such as personal identity, what it means for the fact of personal identity to be ”determinate”, and what it means for the notion of personal identity to mark off a ”deep” distinction.

3. REDUCTION

The problem of personal identity can be phrased in two different ways. One way is as follows. Let me refer to it as ”The Three Dimensional View”: When are two temporally separate occurrences of a person occurrences of the same person? On this view, persons are, at each time, fully present. They ”endure” in time. We may call the other way of

---

[5] Ibid., p. 213.

phrasing the problem the ”The Four Dimensional View”. On the Four Dimensional View the question is formulated thus: When are two separately existing temporal ”person-slices” slices of the same person? On this view, persons have temporal parts, they ”perdure” in time.

To see the difference, consider, for example, me now when I an writing this, and me when I started to study philosophy in Stockholm in the Fall of 1966. On the Three Dimensional View, under what conditions are these occurrences of a person occurrences of the same — enduring — person (me)? On the Four Dimensional View, under what conditions are these two, separately existing temporal slices slices of the same (perduring) person? Note that my person is not identical to my body. As will be explained below, my person and my body are made up by the same kind of stuff, however. And my person and my body overlap, to a large extent. They do not match each other perfectly, however. There are early temporal slices of my body that are not slices of my person (no consciousness exists in those embryological slices), and for all I know there may exist late (brain dead) temporal slices of my body that are not slices of my person.

In the present context I adopt the four-dimensional stance. The four-dimensional view of persons as entities with temporal parts fits well into a naturalistic general outlook which I find intellectually attractive.

Is a four dimensional view compatible with counterfactual talk typical of moral contexts; what if I had acted differently from how I did? Will we not have difficulties in explaining such talk in terms of possible worlds?

I think it clear that, even under the four dimensional analysis, such talk is meaningful. One way of understanding it is not to take the possible world metaphor literally, but to construe modal talk as talk ascribing characteristics to *actual* objects. This is how I have argued elsewhere.[6] Another way is to take the possible world metaphor literally and to supplement it with a *counterpart* theory of physical objects and persons. This is how David Lewis has argued.[7] His counterpart theory may seem extravagant but, to be sure, it is not the main obstacle with his modal ontology. If we believe in the existence of merely possible worlds, we should not hesitate too much to believe, too, in counterparts.

---

[6] Cf my "Morality and Modality", *Philosophical Papers*, Vol. XX, 1991, pp. 139-153.

[7] Cf his *The Plurality of Worlds*, (New York: Blackwell, 1986).

On a reductionist view, how do we answer the kind of question now raised? When are two temporally separate occurrences of a person occurrences of the same person?

Now, this is not quite easy to say, for there are various different sorts of reductionism. At least we ought to distinguish a semantic variety of reductionism from a metaphysical one. We should also note that there exists what has been called ”eliminative” reductionism.

On a semantic version of reductionism, ”these two temporal slices are temporal slices of the same person” means the same as ”these two temporal slices are related to each other in the following mental and/or physical way...” It is far from clear about what language or speech community such a theory could purport to be true. And it is hard to believe that there exists any language or speech community about which it would, as a matter of fact, be true. In the following discussion I think we can just dismiss semantic reductionism.

Then there is a metaphysical kind of reductionism. We are familiar with it from the sciences. It comes to something like the following. We use a certain (folk) notion in our study of nature. I think of a notion such as ”water”. We recognise water through its smell, taste, and looks. We find evidence for certain empirical laws, where the notion plays a role, either in the antecedent or in the consequent of them. We find, for example, that if something is water, then it freezes at 0°C, boils at 100°C, and so forth. And we find that if salt, which we also recognise because of its smell, taste, and looks, is thrown into water, it dissolves. As science advances, we come to hold more sophisticated views. We find that matter is composed of atoms that compose molecules. We discover that there exists a substance in nature with the molecular structure H_2O. It is true of H_2O as well that it freezes at 0°C and boils at 100°C. Furthermore, something's being H_2O connects in an interesting way with various other scientific facts. We realise that this is no coincidence. Something being water seems to be supervenient upon its exhibiting the molecular structure H_2O, we think. If we are daring, we even venture the hypothesis that water *is* H_2O. This means that we have ”reduced” water to H_2O.

The reduction is not metaphysically innocent. In order to be able to achieve it, we must make the bold conjecture that, apart from things there exist properties as well. We must have come to think that, while ”water” and ”H_2O” differ in meaning, *they both refer to the same property*.

Whatever this may mean, it must include the conjecture, I suppose, that there is no possible world where water is not H_2O.

Or, if explicitly we do not want to commit us to the existence of properties, we may have recourse to facts. We say that the fact that something is water is identical to the fact that something is H_2O. This is the line taken by Parfit, who writes that, on reductionism, "the fact of a person's identity over time just consists in the holding of certain more particular facts".[8]

But how do we identify facts? I suggest that two facts are identical if, "in them", the same property is being ascribed to the same individual. So we have to acknowledge properties all the same.

Can we achieve something similar to metaphysical reduction in the discussion of personal identity. Since Parfit defends what he calls a ”reductionist” view of personal identity, one may get the impression that this is what he believes. As we saw, according to Parfit "the fact of a person's identity over time just consists in the holding of certain more particular facts." Which are these more particular facts? Parfit describes two possible sets of such facts, one set of physical facts and one set of psychological facts. He seems to be inclined to adopt the psychological criterion. These sets of facts are described in the following manner:

> *The Psychological Criterion.* (1) there is *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) there does not exist a different person who is also psychologically continuos with Y. (5) Personal identity over time just consists in the holding of facts like (2) to (4). (207)

What, then, does strong connectedness mean? The answer is given as follows:

> ... we can claim that there is enough connectedness if the number of connections, over any day, is *at least half* the number of direct connections that hold, over every day, in the lives of nearly every

[8] Ibid., p. 210.

> actual person. When there are enough direct connections, there is what I call *strong* connectedness. (206)

And the physical criterion is as follows:

> *The Physical Criterion:* (1) What is necessary is not the continued existence of the whole body, but the continued existence of *enough* of the brain to be the brain of a living person. X today is one and the same person as Y at some past time if and only if (2) enough of Y's brain continued to exist, and is now X's brain, and (3) there does not exist a different person who also has enough of Y's brain. (4) Personal identity over time just consists in the holding of facts like (2) and (3).

Personally, I find it hard to see how these two criteria can come apart. My cerebrum seems to go where my person goes.

Parfit gives science fiction examples of, say, teletransportation and concludes that, if it succeeds we have the psychological criterion fulfilled but not the physical criterion. In teletransportation a Scanner on Earth destroys my brain and body, while recording the exact states of all my cells. It then transmits this information by radio. The message reaches the Replicator on Mars. This creates, out of new matter, a brain and body exactly like mine. In this body I wake up.

But why does he not say that the body on Mars is identical to the body that existed (earlier) on Earth? Why do we not conclude that successful teletransportation is not only a way of transporting persons but a way of transporting their bodies as well? Or, on the four dimensional view, why don't we say that the body of the teletransported person has one temporal part on Earth and another temporal part on Mars?

To be sure, the body on Mars is composed of other pieces of matter than the body on Earth. But *our* bodies on Earth undergo radical changes as well. The fact that the body I have on Mars is *my* body (after I have been teletransported to Mars and wake up in it) would suffice, it seems to me, to warrant the assertion that it is *the same* body as the body I possessed before teletransportation, on Earth.

Something similar can be said about the examples Parfit gives where new memories and projects are instilled in an old brain. Must we not think of some changes in the brain as well, in order for the new mental

traits to become possible? But then should we not say that, what we have now is not a new person "inhabiting" another person's brain, but a new brain as well as a new "resulting" person inhabiting this brain?

I will not go any deeper into these problems, however. My query is different. In the present context, the important thing is to know how we should conceive of Parfit's claim that his view of personal identity be reductionist (irrespective of whether the reduction be made along the psychological criterion, the physical criterion or, along both). How should this claim be taken?

It is not plausible to interpret Parfit as holding the view that a semantic reduction of our talk of personal identity is possible. This view is so implausible that considerations of charity forbid us to ascribe it to anyone, unless it be very explicitly stated.[9] And this is not the interpretation that Parfit's own words suggest. He seems to be defending metaphysical reductionism. Personal identity *is* the holding of facts like the ones mentioned in the respective criteria. More specifically, on Parfit's view, personal identity is the holding of psychological facts of the kind mentioned in the psychological criterion. To say that these relations hold between to spatiotemporal person-slices, and to say that the slices are slices of the same person, are only two different ways of referring to the same fact.

Even if we can identity the fact of personal existence over time with a certain psychological fact, this does not settle what kind of entities persons are. Are we identical with (parts of) our bodies? Or, are we separately existing entities, constituted by certain truths about our bodies? The former position is called "identifying" reductionism by Parfit and the latter "constitutive". He goes explicitly for the latter. I am not sure that I understand the difference. This distinction, however, plays no role in the argument to be examined in the present context.[10] For my own part, I would like to say, as was stated above, that my person is different from my body but a part of it. So they are both physical objects overlapping each other to a large extent.

However, as is noted by Parfit in "The unimportance of identity", another way of taking talk about "reduction" is possible as well, what we

---

[9] This is not only due to the four-dimensional form that the view of personal identity takes in the present context. It would be implausible also on a view of persons as "enduring" rather than "perduring" objects.

[10] "The unimportance of identity", p. 16.

can speak of as ”eliminative reduction” or, for short, ”elimination”. When science advances, some empirical notions, like the notion of water, are, so to speak, ”saved” through reduction to more secure notions, such as H_2O. Other notions tend to disappear, however. They get eliminated rather than reduced. This is what has happened to the notion of being a witch. We gave up the notion when we discovered that it did not refer to any property at all. Something similar could be held to have happened to the notion of being a person, one may think. According to Parfit, some Buddhist texts come close to this position. Parfit himself, however, guards against it. On this interpretation of Buddhism, persons do not exist. Parfit, however, believes that persons exist.

In the following, it is of the utmost importance to distinguish clearly between the reductionist and the eliminative position. For, obviously, they do not sit well together.

Perhaps we should mention also another possible ”deflationist” view of persons, less radical than the eliminative one. We could conceive of the notion of a person in a nominalist manner. This means that the notion, although shallow and far from precise, plays *some* role, not in science, but in our ordinary thought in our ordinary lives. In this it is similar to the notion of something being a chair or a table. We take such notions at face value. There are some clear applications of them. And it does not matter that there are borderline cases, where it is an open question whether they apply. They raise no difficult metaphysical questions. Tables and chairs are physical objects of a well-known kind.

While elimination is the sole way of *getting rid* of notions we find in folk sciences, there exist more than one way of *rescuing* a notion. Reduction is but one of them. Theorising is the other main alternative. Not all respectable scientific properties are identical with empirical ones. Some scientific properties are theoretical. Theoretical properties are not straightforwardly empirical, but they are, none the less, real.

Characteristic of theoretical properties are that, while we attribute them on the basis of some empirical evidence, the evidence underdetermines our attributions. Theoretical properties are connected to empirical, observational properties, through (often rather esoteric) ”background” hypotheses.

Why do we believe in theoretical entities? We need them in our best explanations of some empirical data. They play an important role in the

scientific laws we use in our explanations. The property of being an elementary particle is an example in place.

The view discussed by Parfit, that personal identity may be a ”further fact”, could be understood in various different ways. The most charitable interpretation, however, is to take the view to be to the effect that the property of being a person may be, in a similar vein, theoretical. But this idea comes in two versions. The point could be that persons are further *entities* (Cartesian egos). Or, the point could be only that personal identity is a further *fact*, that being a person is a *property* that is different from the properties mentioned in the so called ”criteria” of personal identity; this is consistent with both kinds of properties being exhibited by the same entities.

The latter interpretation is the one that yields the most plausible (and, ontologically speaking, the most economical) view of personal identity. It makes perfectly good sense of the idea that being a person is a further fact, without stipulating that there exist mystical entities, outside space (Cartesian Egos). On this view, evidence of the kind cited in the physical and psychological criteria underlies our ascriptions of personal identity. If certain psychological and physical relations hold between two spatiotemporal slices of persons we claim that the two slices are slices of the same person. However, the evidence at hand typically underdetermines this conclusion. We draw the conclusion that the slices are slices of the same person, if we do, because of its superior explanatory value. Moreover, we attribute personhood on the basis of empirical evidence of various different kinds. Typically, these attributions are underdetermined by the evidence. Sometimes we have difficulties in reaching a definite opinion. What about a grown up chimpanzee, for example?

Having distinguished between three kinds of reduction (semantic, metaphysical, and eliminative) and between two ways to save a notion, through reduction and theorising respectively, and having, furthermore, noted the possibility that the notion of being a person be ”nominal”, I turn now to the questions of indeterminacy and depth.

4. INDETERMINACY

As we have seen, Parfit argues that if we can reduce personal identity to various mental or physical relations between actually existing things in the universe, then personal identity may well be indeterminate. There may exist cases where it is an open question whether a certain person did survive or not. Examples of this are such things as teletransportation, where the original as well as the ”copy” survive, or ”division”, where both halves of a person’s symmetrically functioning brain are successfully transplanted to bodies similar to the original one. Or, in his combined spectrum case, where I am being tampered with so that (in the middle of the spectrum) the resulting person is both physically and mentally a combination of me and Greta Garbo.

I admit that, in the science fiction cases cited by Parfit, it is sometimes difficult to say whether a person would survive or not. And there are cases of senile dementia, where we are reluctant to say whether a living body is occupied by any person at all. This means that, epistemically speaking, these cases are indeterminate. This does not mean, however, that, ontologically speaking, they are indeterminate. There may be a fact of the matter (in a certain fictitious case, I either live or do not live), in spite of the fact that we cannot gain secure knowledge about it.

Epistemic indeterminacy (in science fiction, hypothetical contexts, in ”thought experiments” or in extreme cases of brain damage and senile dementia) is probably compatible with reductionism, since it is difficult in these examples to tell whether the psychological and physical criteria apply, but epistemic indeterminacy does not as such bring with it ontological indeterminacy.

Why should it? Even if we cannot *tell* whether, in these examples, our physical or psychological criteria of personal identity are fulfilled, there may *be* a fact of the matter.

Parfit thinks otherwise, however:

> Suppose that we accept the Psychological Criterion. We can describe cases where the psychological connectedness between me now and some future person will hold only to a reduced degree. If I imagine myself in such a case, I can always ask, 'Am I about to die? Will the resulting person be me?' On this version of the Reductionist View, in some cases there would be no answer to my question. My question

> would be *empty*. The claim that I am about to die would be neither true nor false. If I know the facts about both physical continuity and psychological connectedness, I know everything there is to know. I know everything, even though I do not know whether I am about to die, or shall go on living for many years.[11]

As Parfit himself notes, this is hard to believe. It is not hard to believe that we cannot *tell* in these examples whether we would survive. But it is hard to believe that *there is no fact of the matter*. But why should we accept this conclusion? I fail to see why this is so, even on reductionism.

On reductionism, the question whether I have survived or not is not empty. It is a question as to whether the criteria of personal identity are satisfied in the example in question.

But if we know enough about the case to know whether the criteria are satisfied or not, is it not *then* an empty (further) question whether I have survived or not?

No, this further question is a genuine one. In order to answer it, we must not only know that the criteria are satisfied, *we must also know that personal identity is what is captured by these criteria.*

If the reduction is correct, the truth of the identity of a person and what is captured by the criteria may be necessary. However, it is not a truth that we know a priori. We have to come by it in the way we come by truths like the one that water is H_2O.

What if the suggested criteria are vague? To the extent that they are, and to the extent that we want to stick to reductionism, we must make them more precise, it seems to me. In the discussion that has followed upon Parfit's first contribution to it, the number of such suggested amendments to, or revisions of the criteria, is legion. These amendments and revisions are true to what could be called "the reductionist program".

Parfit seems to think that there is no point in making the criterion more precise. For there is no definite property that, by doing so, we can capture. The question whether I survive in these circumstances or not is "empty". So, if we make a more sharp delineation of the notion of personal identity, we just settle a matter of meaning.

Is this so? I cannot see that Parfit has given us any reason to believe that it is. To be sure, it does not follow from reductionism that it is. Quite

---

[11] *Reasons and Persons.* p. 214.

to the contrary. The conclusion that the question sometimes is empty seems hard to reconcile with reductionism. How can we "reduce" an indefinite property to any definite property in particular? How can we "identify" it with any other definite property?

If there is no fact of the matter, then this speaks in favour of an eliminative view of personal identity rather than a reductionist one. If there are no (definite) conditions of identity, then there is no entity. Or, at least, the notion is of a ”nominal” character.

On the other hand, if, as we tend to believe, ontological determinacy is a fact, then this is consistent both with reductionism (best conceived of as a program) and with the view that the notion of a person is theoretical.

Epistemic indeterminacy is compatible too with both a reductionist and a theoretical view of persons. As a matter of fact, epistemic indeterminacy is only what we should *expect* to find, if persons are theoretical entities. After all, if our ascriptions of personal identity are underdetermined by our evidence, then it is very natural to believe that there may exist cases of (epistemic) indeterminacy as well.

To some extent, the choice between the reductionist program and the theorising position may be a matter of philosophical taste. If it is, I must confess that I find the theorising alternative more tasty. This means that, in my discussion, I will concentrate on it. However, I believe that what I say about personal identity can be transformed just as well to apply also to the reductionist position.

Note one again, however, that the theorising view comes in two versions, one presupposing that there are separately existing entities, Cartesian Egos, and the other merely presupposing that facts of personal identity are "further" facts (that may well concern, say, ordinary physical objects). The latter version is not taken seriously by Parfit who notes that we believe that personal identity is determinate and adds:

> For that to be true, personal identity must be a separately obtaining fact of a peculiarly simple kind, it must involve some special entity, such as a Cartesian Ego, whose existence must be all-or-nothing.[12]

This is doubly incorrect. As we have seen, the belief that the fact of personal identity is always determinate is compatible with the reductionist

---

[12] "The unimportance of identity", p. 27.

program (as a matter of fact, it is a presupposition of that program). Moreover, even if we take a theorising stance to the fact of our personal identity over time, we need not assume that there exist Cartesian Egos. So even if we have a metaphysical prejudice against this kind entities, we may adopt the theorising position in a, metaphysically speaking, more "innocent" version. We may conclude that persons are different from bodies but that both kinds of entities are physical (they are parts of space-time) and, to a large extent, overlapping each other. The exact delineation of a person represents a theoretical decision, based on, but not completely determined by, empirical evidence. A comparison with biology may be in place here. According to Richard Dawkins, a gene is a theoretical entity.[13] We identify genes with functional criteria, but the criteria underdetermine the exact delineation of a single gene. However, this does not mean that genes are Cartesian entities, existing outside space. They are parts of the DNA of concrete living individuals.

It has surfaced from the foregoing discussion that the reductionist and the theorising positions have something in common. They are both realist views of persons. In this they contrast with the eliminative position.[14]

How do we decide between a realist and an eliminative position, then? This question brings us to the problem of the ”depth” of the fact of personal identity, be this fact a ”further” or an empirical one.

5. DEPTH

It seems to me that existence is a limiting case of depth. Even nominal notions refer to existing things. There are chairs and tables. So these notions have some depth. However, depth comes in degrees. In what way can one notion be deeper than another one?

I suggest that the notion F is deeper than the notion G if F (or, the property referred to by ”F”) connects with other properties in more interesting and systematic and general ways than does G (or, more properly, the property referred to by ”G”).

---

[13] Cf. *The Selfish Gene* (London: Granada, 1978), Chapter 3.

[14] Perhaps there are quasi-theoretical, projectionist middle positions between on the one hand reductionism and the theorising view and, on the other hand, the eliminative view. In the present context, I will not go into such subtleties, however.

One example of this is if the notion fits into various lawlike hypothesis capable of best explaining empirical data. In this sense, even though chairs and tables exist, it is a shallow fact that something is a chair or a table. We do not refer to chairs and tables in any scientific explanations. However, it is a deep fact that something is water (H_2O). We refer to this fact in some explanations. It was because something happened to be H_20 that it boiled at 100° C.

Now, is being a person a deep fact? Does the property of being a person enter the antecedent (or consequent) of any lawlike hypotheses capable of best explaining any empirical data?

Few would say that it does. Some say that in our explanations of free human action we must have recourse to the notion of a person. This is the position taken up by Roderick M. Chisholm.[15] According to Chisholm, actions are caused, not by events (desires and beliefs, for example) but by *agents*. However, it seems to me that no *actual* explanation of any action takes this form. To the extent that we can explain human actions, we do explain them in terms of desires and beliefs. Any further reference to the agent being the *cause* of the action is redundant. So it is tempting to conclude that the notion of being a person, if it is at all real, is a very shallow one indeed. However, there are other kinds of law-like hypotheses that the property of being a person can fit into. Most notably, the property does fit into certain *moral* hypotheses, thought of as being capable of explaining important *moral* data.

Could not these connections underpin the view that personal identity is a deep notion? I think they can. And this is the main thrust of this paper. If these moral hypothesis are plausible, then the fact of personal identity is indeed a deep one, or so I will argue. We should not only believe that entities that we need to have recourse to in our best (scientific) explanations of empirical data are real, we ought also to believe that entities that we need to have recourse to in our best (moral) explanations of particular moral facts (the rightness and wrongness of particular actions) are real. Metaphysically speaking, moral and scientific depth are on a par.

One example of the kind of moral hypotheses I am thinking of is egoism. This is the view that each individual ought to promote his or her

---

[15] C.f. for example his ”Human Freedom and the Self”, in Gary Watson (ed.), *Free Will* (Oxford: Oxford University Press, 1982), pp. 24-35.

own well-being. Sometimes this theory is referred to as a theory of ”rationality” but, to the extent that it is claimed by the theory of rationality in question that we *ought* to be rational, the theory is in competition with various moral theories and it coincides with egoism and should be discussed under this heading. (If it gives only a stipulative definition of ”rationality”, then there is little reason to discuss it at all).

On egoism, each person has duties towards his or her ”future self”. On egoism, it is of (normative) importance whether a certain good is *your* good or the good of *someone else*. The theory need not be ”pure”, i.e., it need not pretend to be specifying the only source of your duties. Perhaps you have also some (weaker) duties towards other people. But to the extent that egoism stipulates *some* special concern for your own well-being (no matter how well-being is conceived of), it makes essential use of the notion of being one person rather than another. This means that, if the theory is plausible, then this fact (the plausibility of the theory) adds to the *depth* of the fact of personal identity. The theory renders it important (to me) in difficult cases to try to find out whether a certain person (me) would survive a radical change (such as a severe brain damage) or not. The fact that this person would be me could help to explain (on the assumption that there is some truth in egoism) why I have a certain obligation towards this person.

Other examples are given by various different ideas of retributive and distributive justice. In obvious ways these theories make essential use of the difference between different persons. If is fair, on a retributive view of punishment, to punish you, only if you are identical to the person who committed the crime for which the punishment is meted out. The fact that you are identical to this person can help explain why you deserve to be punished. And, on, say, an equalitarian view of distributive justice, it may be obligatory to give a smaller good to a person who leads a life that is (overall) unhappy than to give a greater good to another person who leads a life that is (overall) very happy. The fact that it is this person, and not another person who has lead a miserable life, may help explain why it would be right to give a certain good to this person. More generally, if any of these theories, which make essential use of the notion being a person, is plausible, then this adds to the depth of the fact of personal identity. We need to find out in hard cases where one person ends and another person begins. And we need to have recourse to the notion in our

(moral) explanations of why certain actions are right and other actions wrong.

Further examples are found in various different ideas about special obligations to persons we have committed us to, or to persons who are near and dear to us. If any of these normative outlooks, making essential use of the notion of being a person is plausible, then this adds to the depth of the fact of personal identity.

Finally, even a utilitarian, with a subtle theory of value, may need to have recourse to the idea of personal identity. I think of a view to the effect that the value profile of a life of a person may be of moral importance (it may be better to experience more happiness at the and of the life than at the beginning), or a view to the effect that it could be better for a person to live a ”full” life (to see one’s grand-children) than a short life, even if the latter contains more pleasure on the whole (provided that the longer life is, all the time, worth living). If there is *anything* to *any* of these evaluative suggestions, then this means that the fact of personal identity possesses some depth. We need to find out where one persons ends and another persons begins.

According to Parfit, reductionism means that all moral outlooks, making essential use of the notion of being a person, are undermined. They may survive somewhat revised, in the form of claims that empirical aspects of personal identity (psychological continuity and/or connectedness) are of moral importance, but not, according to Parfit, of any very serious importance.

This argument seems to me to be mistaken. Parfit puts the cart before the horse. He argues that, since reductionism is a true view, personal identity is not a deep fact. And, if it is not a deep fact, it cannot play any very important moral role. However, in the first place, reductionism does not as such show that personal identity is not a deep fact. It shows that personal identity is a empirical fact. This empirical fact may very well be a deep one. If persons exist at all, the fact of personal identity possesses at least *some* depth.

Secondly, how deep this fact is depends, I suggest, on whether the moral views here discussed, making essential reference to persons, are plausible or not. If they are *all* plausible, then this shows that the fact that a person survives, or that a certain good comes inside rather than outside the life of a certain person, or that a certain person really did a forbidden act, is indeed a deep fact. We have to refer to this kind of fact

when we settle moral disputes: This person ought to be punished, this person rather than that person ought to have this gratification, and I have a moral responsibility to prepare specially for my *own* old age, and so forth. And we need not only have recourse to the notion of being a person when we try to *find out* what is right and wrong. We must have recourse to the notion of being a person in our (moral) *explanations* of *why* certain actions are right rather than wrong too. And this means that the fact of personal identity demarcates an *important* difference, a *deep* difference, between persons.

Parenthetically, it should also be remarked that Parfit's *modified* theories of egoism, retributivism, distributive justice, and commitments, making use, not of the notion of personal identity, but of psychological relations of continuity and/or connectedness, are only plausible as views about *instrumental* value. If I have a certain preference, then I may be interested in there being someone in the future having the same preference and trying to have it satisfied since, if there is not, maybe the preference will not be satisfied. If I want to save Venice, to use Parfit's own example, I may want there to be *a* future person who wants to save Venice too. Otherwise there is a risk that no one will bother to do anything to this effect. But this instrumental interest is very different from the interest presupposed in egoism that *I* be there in the future. On egoism, it is important in itself (to me) that *I* be there in the future (at least if I lead a good life; otherwise it is important that I *not* be there in the future). It is strange to believe, however, that it would be of interest *in itself* (to me) that *someone* be there in the future and have my preference. Perhaps Venice is already saved and the preference is no longer needed.

I suggest that it is only this kind of instrumental interest that Parfit captures in his discussion of the importance of psychological relations *without* identity.

6. AN OBJECTION

To the foregoing argument it might be objected that what I am discussing is only the *moral* depth of the fact of personal identity. I have suggested how personal identity might connect with important moral outlooks, such as egoism, theories of justice, and moral theories about special

commitments. However, the crucial aspect of Parfit's use of the depth metaphor may be different. He may want to argue that, unless personal identity marks off a deep *scientific* fact, it cannot carry any moral weight. In order to possess moral depth a notion must possess scientific depth as well. And the fact of personal identity lacks scientific depth. Therefore, it must also lack moral depth.

This argument is suggested by some passages of Parfit's text. I am not sure that he is committed to it, however. Be that at is may, the argument may seem compelling and needs to be discussed. However, the argument is flawed. Moral facts may *supervene* upon empirical or theoretical facts, but, typically, they do not as such possess scientific depth. For example, consider the notion of an experience being (to a certain extent) more pleasant than another experience. To any moral outlook that pays at least some attention to human happiness, this notion is crucial. However, this notion plays no scientific role whatever. And yet, for all that, it is a deep further fact about an experience that it is more pleasing than another experience. So we have to acknowledge that moral depth and scientific depth of our notions do not always coincide.

The upshot of this is that moral and scientific depth of our notions may very well come apart. Typically, they do not coincide. But, to be sure, moral depth of a notion is enough to warrant a realistic stance to it.[16] It remains to be shown, however, if we want to retain the notion of being a person, that this notion really has moral depth. I will not try to settle this issue in the present context. My aim is more restricted. I will examine some arguments put forward by Parfit to the effect that the difference between persons is of no moral importance.[17]I find them wanting.

---

[16] C.f. also my "Moral Doubts About Strict Materialism", *Inquiry*, Vol. 30, 1987, pp. 451-58, where I argue that the (moral) plausibility of hedonism makes materialism of the eliminative variety implausible.

[17] Being an adherent of hedonistic utilitarianism I am not suited to this task. As a matter of fact, I believe that the difference between persons is of no moral importance. However, in the present context, I play the devil's advocate. I do not believe that what Parfit has said about personal identity strengthens the position I want, for other reasons, to defend. Cf. my *Hedonistic Utilitarianism* (Edinburgh: Edinburgh University Press, 1998) for a defence of a moral outlook making no essential reference to persons.

7. PARFIT'S PARTICULAR ARGUMENTS

Parfit does seem to assume that since the fact of personal identity is not a deep one, it must lack moral importance. This seems to me to be the main thrust of his argument. I have already noted the main problems with this line of argument. However, he also tries to show that personal identity is not what matters in more particular arguments. I now turn to them.

One of his argumentative strategies is as follows. He considers our interest in our own future. We seem to have a special concern for our own future. If we know that some one will feel pleasure or pain, then we want to know if we are identical with this person who will feel pleasure or pain. Or so we think, at any rate. Parfit tries to show, however, that on a reductionist view of personal identity, personal identity cannot be what matters when we are (especially) concerned whether a certain pleasure or pain will be *our* pleasure or pain or the pleasure or pain of *someone else*. He runs through various different suggested criteria of personal identity, identifying our identity over time with physical or mental facts (or with a combination of both). In relation to each of them he concludes that, what is captured by them, cannot really be of any importance. And he concludes that this shows that the *sort* of facts captured by these criteria must lack direct importance. This is not convincing. The reason that we are not satisfied with the examples may be that they are based on incorrect reductions. This may be true of all suggested reductions. The reductionist program may not yet have been consummated.

Suppose now that the program has been successfully completed. We know that two temporal slices of a person are temporal slices of one of the same persons if, and only if, they are X-related to each other. We used to believe that personal identity is of great importance. We now ask ourselves. Is X, as such, of great importance?

If the reduction is correct, we must conclude that it is. But this may come as a surprise. Suppose I learn all the things a chemist can tell me about a certain liquid. I then know a lot about it. But suppose that someone adds the information that it is a certain *claret,* that I have once enjoyed. This may come as a surprise to me. All of a sudden, I know that it is nice to drink it. The chemical composition about this liquid now takes on a new significance to me. Something similar can happen when we learn that a certain empirical fact is identical to the fact that a certain

experience will be an experience of mine. Then we realise that it is of great importance (to me).

A similar, negative point, has been made by Sydney Shoemaker[18], and it has often been repeated in the discussion about this subject. A materialist should be no less concerned to prevent pain than a dualist.

Then there is another argumentative strategy taken up by Parfit. It takes as its point of departure the putative fact that physical and psychological continuity and/or connectedness can take a branching form. This would be the case if both halves of my brain would be successfully transplanted, into different bodies that are just like mine. Two people would wake up, each of whom has half my brain, and is, both physically and psychologically, just like me. Parfit now makes the following claim:

> How should I regard these two operations? Would they preserve what matters in survival? In the Single Case, the one resulting person would be me. The relation between me now and that future person is just an instance of the relation between me now and myself tomorrow. So that relation would contain what matters. In the Double Case, my relation to that person would be just the same. So that relation must still contain what matters. Nothing is missing. But that person cannot here be claimed to be me. So identity cannot be what matters.

We can summarise the argument as follows:

(1) Physical and psychological continuity and/or connectedness can take a branching form.
(2) Identity cannot take a branching form.
(3) When a person is split into two different persons (not identical to each other), then everything that matters is preserved in the relation between the original person and each of the resulting persons.
---
(4) Therefore, personal identity cannot be what matters.

This argument is problematic on many counts.

---

[18] Cf. his ”Critical Notice of Derek Parfit, *Reasons and Persons”*, *Mind*, Vol. 94, 1985, p. 451.

In the first place, it is not clear that (the relevant kind of) psychological continuity and/or connectedness *can* take a branching form. In the examples given by Parfit it is plausible to hold, with Nozick, that some of the resulting persons is in some manner more closely connected to the original person.[19] Then the relevant relation of continuity and/or connectedness holds only to this person, the ”closest continuer”.[20]

However, suppose there exists a case where we have two resulting persons and no difference between them as to who is most closely related to the original person. Is it true, then, that what matters is preserved in the relation between the original person and any one of the resulting persons? This is not obvious. In the example there is no unique person carrying memories and other mental traits like mine. Why could not this matter?

As a matter of fact, Parfit himself seems to believe that something is lost. According to Parfit, the *instrumental* value of survival is taken care of in the example. If I have two future copies, then all the books I intend to write will be written, my children will be taken care of, and so forth. But yet, for all that, Parfit himself seems to admit that something is lacking, compared to what we believe to be the case in standard cases of survival. There is no possibility for the original person to "anticipate" what will be felt later on, if branching is a fact. This seems to me to confirm that personal identity is what matters (fundamentally, on egoism). True branching, if it can come about, is death, then. And, if egoism is correct, and if the dead person would have had a good life if he or she had not died, then death was bad for him or her.

Parfit does say, of course, that when branching takes place, no *loss* comes about since, anticipation is impossible anyway. There is no deep fact of future experiences being mine, even in ordinary cases, where no branching takes place.

---

[19] Cf. his *Philosophical Explanations* (Cambridge, Massachusetts.: Harvard University Press, 1981), pp. 29-37, about this.

[20] Another way of avoiding the conclusion that mental division is possible is to go for a theory to the effect that two persons can be present in one and the same body before the bodily ”split”. This line has been taken up by both John Perry, ”Can the Self Divide?”, *The Journal of Philosophy*, Vol. LXIX, pp. 463-488 and David Lewis, ”Survival and Identity”, in A. Rorty (ed.), *The Identities of Persons* (University of California Press, 1976), pp. 17-40, but, since I do not find it plausible, I will not rely on it in my argument.

But how does Parfit know this? Once again, when Parfit denies the importance of personal identity, he has taken for granted, rather than shown, that egoism (or any other moral theory referring essentially to persons) is wrong. In particular, it does not follow from reductionism that it does not matter (fundamentally, to me) that there be a *unique* future person having my memories and my mental traits.

It might be tempting to argue as follows. Let us concede that fission means personal death. But, considering a possible science fiction case of fission, we do not find it as fearful as ordinary death. So continued personal existence cannot be what matters.

Sydney Shoemaker accepts this argument.[21] So does David O. Brink.[22] The latter, however, develops a version of egoism with a new kind of agents, what he calls *continuants*. Continuants are maximal series of R-related person stages. In the fission case, there is one continuant consisting of the parts making up the person before fission and the parts making up one of the two persons after fission and another continuant consisting of the parts making up the person before fission and the parts making up the other one of the two persons after fission. As is noted by the author, this generates ambiguity about the subject of practical deliberation at the time of fission. It is not a very attractive response, it seems to me.

Or, in a different respons, Brink argues that our concern for the results of fission may have a rationale in the view that their good is part of our good, in the manner that the good of my spouse may be said to be part of my good.[23] This is not very convincing either, however. On a hedonistic view of well-being this response is simply incomprehensible (when my spouse is happy this can *make* me happy, but her happiness is not *experienced* by me), and even on preferentialism the response is hard to understand.

Instead of making the kind of concession Brink makes we ought to note therefore that the argument he concedes to is flawed. In the first place, some people may find fission as fearful as death. Personally, I am inclined to take up the same kind of attitude towards both. And some people tend to believe that, even if fission is not as bad as death, it is "something

---

[21] Shoemaker and Swinburne, *Personal Identity*, p. 121.

[22] ”Rational Egoism and the Separateness of Persons”, in Jonathan Dancy (ed.), *Reading Parfit*, p. 125.

[23] Ibid., pp. 126-128.

horrible". This is the view of Susan Wolf.[24] But be that as it may, however, our attitude is irrelevant. Most people care less about the distant future than they do about the near future. This does not make it rational to care less about the distant future than about the near future. According to egoism, we *ought* to be equally concerned about both our near and our distant future. We can perhaps find an evolutionary explanation of the difference in attitude. But this explanation does not rationalise our attitude. The same may be true about our attitude to science fiction cases of fission. The attitude may have a good explanation (though hardly an evolutionary biological one) but it may still be irrational.

If egoism is a correct moral view, then we *have* reasons to care about our distant future — and to fear our personal death (even if it comes about through fission rather than in the "usual" way). And egoism cannot be rebutted with the claim that we do not always abide by it.

Finally, there is the following argument put forward by Parfit. According to various different moral outlooks such as egoism, various different theories about retributive and distributive justice, and about our special commitments, personal identity plays a crucial moral role. Now, consider something like the middle of the "combined spectrum", where I have been tampered with mentally and physically so as to end up as something *as* similar to me now as to Greta Garbo at 30. According to Parfit's view of the received view, somewhere in the middle of the combined spectrum there must be a sharp dividing line where I cease to exist, a line of importance to the moral outlooks just mentioned. But how can a very little difference be of such moral importance?

> It is hard to believe ... that the difference between life and death could just consist in any of the very small differences described above. We are inclined to believe that there is *always* a difference between some future person's being me, and his being someone else. And we are inclined to believe that this is a *deep* difference. But between neighbouring cases in this Spectrum the differences are trivial. It is therefore hard to believe that, in one of these cases, the resulting person would quite straightforwardly be me, and that, in the next case, he would quite straightforwardly be someone else.[25]

---

[24] Cf. her "Self-Interest and Interest in Selves", *Ethics*, Vol. 96, 1986, p. 715.

[25] *Reasons and Persons*, p. 239.

This is not convincing. First of all, it is difficult to believe that, in the middle of the combined spectrum, there is a person at all. The memories and ambitions of this individual, if we were to produce him or her (does the person has a sex in the middle of the spectrum?), do not form a coherent pattern. But perhaps Parfit's point could be restated. On the received view, somewhere in the spectrum Parfit ceases to exist. And, at some other point in the spectrum, Greta Garbo starts to exist. The latter case is exotic. We see examples of the former kind of case when people suffer from progressive forms of senile dementia, however. Then we could state Parfit's point as follows: It is hard to believe that a *small* difference (where the person with dementia has lost a few more memories or personal characteristics) can mean a difference between life and death.

But is this so hard to believe? It is hard to believe that the difference between life and death can mean a little difference from the way life is (to me) *now*. But the point where I cease to exist means a *great* difference from how life is (to me) now. And it is not hard to believe that exactly *how* great a certain *great* difference from how life is (to me) now can be of the utmost importance; it could mean the difference between life and death.

Exactly where are we to draw the line? Well, Parfit himself has suggested an answer to this question. This questions bears at least some plausibility. According to the psychological criterion, the line should be drawn, as we saw above, whether the number of connections, over any day, is *at least half* the number of direct connections that hold, over every day, in the lives of nearly every actual person. If we are not satisfied with this answer, we can try to refine the criterion.

8. CONCLUSION

The conclusion of the foregoing argument is as follows. If we believe in any of the various different moral theories referring essentially to the fact of personal identity, then we ought to take a realist stance to persons. If we refer to persons in some of our moral explanations (it was wrong for a certain person to perform a certain action because it meant that he or she did not maximize his or her well-being and each person ought to maximize his or her own well-being, for example), then we should

countenance the existence of persons. We must also conclude that, ontologically speaking, there is always a fact of the matter as to whether a certain person did survive a certain treatment or not. For this is presupposed by the moral theories in question.

This realist stance to persons is compatible with our taking either a reductionist view of personal identity or a theoretical one. Furthermore, both these views are compatible with epistemic indeterminacy. Especially if we take the fact of my continued personal existence to be theoretical, as my philosophical taste advises me to do, if at all I shall take a realist stance to persons, then epistemic indeterminacy is only what we should expect.

On the other hand, if we find moral reasons to reject all moral views making essential reference to persons, defending, for example, classical hedonistic utilitarianism, i.e., if we can make all the moral explanations (of the rightness and wrongness of actions) we want to make, without having recourse to persons, then it is only natural to think of the notion of personal identity, not only as indeterminate, but as confused. If the notion plays no crucial role either in science or in morality, then the notion must be on a par with the notion of being a witch, so we ought to get rid of the notion of being a personal altogether. We should stop believing that there are persons.

Or, if this (Buddhist) alternative seems to radical, we may conceive of persons as on a par with chairs and tables. These notions (of chairs, tables, and, perhaps, persons) have fuzzy edges, they play no role in our scientific (or moral) hypotheses, but we still have much use for them, in our ordinary dealings in our ordinary life. They play the same role that the notion of a nation plays to a methodological individualist. The methodological individualist uses the notion in descriptions of the world, in his or her everyday dealings with the world, and so forth, but not in scientific explanations. This is to take the notions at face value but to adopt what could be called a ”nominalist” stance towards them.

The former, eliminative claim, as well as the latter, less radical, nominalist claim, rests on the presupposition that there are no systematic uses *other* than the moral ones for the notion of personal identity in, say, the behavioural sciences.

I do not believe that there are any such uses, but, in the present context, this is nothing I will attempt to argue. It must suffice to note that my conclusion rests on this presupposition.