METAPHYSICS AND EPISTEMOLOGY

Personal identity over time

1. Persons are the entities that, at least during their mature age, can think of themselves as themselves and, more generally, possesses the ability to think and decide. Human beings are persons (apart from some pathological cases). If there is a God or there are extraterrestrials, they are persons too. Each person is a distinct bearer of responsibility; that is, no person bears responsibility for an action performed, or an obligation assumed, by another person.

The problem of personal identity over time is the question what are the necessary and sufficient conditions in order for a person x who exists at a time t to be identical with a person y who exists at a later time t´, that is, under what conditions x and y are one and the same person and t and t´ correspond to different phases in that person's life. The question is interesting because, on the one hand, it is connected with issues about moral responsibility and, on the other, some answers to it imply controversial views about the relation between us and our bodies.

2. The relevant discussion starts with J. Locke (17^{th} century), who considered that x and y are identical iff, at the time t´, y preserves a memory of some experience that x had; more precisely, iff, for some experience that x had at t, y at t´ remembers having that experience. Locke concluded that a person is not identical with their body; for it is logically possible for one person to succeed another in the same body, and it is also logically possible for a person to change bodies. (However, Locke also disagreed with the view that a person is identical with their soul, as well as with the view that the person is identical with the sum body + soul.)

T. Reid (18^{th} century) adduced the counterexample of the old general who is losing it. He remembers some heroic deeds he performed as a young officer, but no longer remembers the apple he had stolen as a child, nor anything else from his childhood. Locke's view leads to the wrong conclusion that the old man is not the same person as the child. (By saying 'the same person', we mean 'one and the same person'; we don't mean that he has the same personality.)

Another counterexample is that of amnesia. I may suffer amnesia and so the person subsisting after the amnesia may not remember any one of my experiences before it. Locke's view leads to the conclusion that that person will not be me and hence I shall no longer exist. But this seems to be wrong; won't that person bear responsibility for what I have done?

The deeper problem with Locke's view is that identity is a transitive relation; that is, if x = y and y = z, then x = z. (The symbol =, in logic and philosophy, is usually read 'is identical with'.) But the relation to which Locke reduces personal identity is not; it may be that y had a memory of some one of x's experiences, z has a memory of some one of y's experiences, but z has no memory of any one of x's experiences.

3. Coming to the 20^{th} century, we see that some philosophers, such as A. Quinton, modified Locke's analysis and considered that x and y are identical iff there is a series of persons p_1, …, p_n where p_1 is x, p_n is y, and every p_{a+1} preserves, at some time, a memory of an experience that p_a had (at an earlier time). Of course, if the philosophers in question are right, and there is such a series, then p_1, …, p_n are the same person; the

plural ‘persons’ was used conventionally. The relation to which they reduce personal identity is transitive.

Locke’s analysis is based on memory. But why confine ourselves to it and not go into the other mental operations? Modifying the analysis further, we can say that x (who exists at time t, but possibly at other times as well) and y (who exists at time t´, but possibly at other times too) are identical iff there is psychological continuity from x to y. What does that mean? It means that there exists a series of persons p_1, … p_n where, for some times t_1, …, t_n, the following are the case: p_1 is x; p_n is y; t_1 is t; t_n is t´; each t_{a+1} is a short while after t_a; p_{a+1}’s mental state at t_{a+1} does not differ very much from p_a’s mental state at t_a; and, for each p_a, p_a’s mental state at t_a, together with any external influences, brings about p_{a+1}’s mental state at t_{a+1}. (Of course, if this modified analysis is correct, and there exists such a series, then p_1, …, p_n are the same person. The plural ‘persons’ in ‘a series of persons’ is used conventionally.) By talking here about a person’s mental state at a time, we mean their total psyche, which consists of experiences, memories, beliefs, emotions, desires, character traits, etc. From the birth to the death of each real human being (perhaps with very few exceptions) there is psychological continuity.

4. Let’s accept that the existence of psychological continuity is a necessary and sufficient condition in order for x and y to be identical, and let’s take any person, e.g. me. Let c be my body. We can conclude that it is logically possible that I should not be identical with c. Here is how:

We imagine that all the information stored in my brain is transferred to a storage device and c is destroyed. Simultaneously, a new human body is constructed; its brain is so formed that all the information is transferred to it from the storage device. The term ‘information’ is here used in a broad sense: it includes experiences, beliefs, desires, character traits, etc. (Our imaginary story, like the other such stories that will follow, presupposes that a human’s mind is somehow stored in their brain. This proposition, even if it eventually proves wrong, is at least logically possible.) As the new body begins moving and talking, we have a person that, according to the criterion of psychological continuity, is me (consider that n = 2; take some time just before the transfer begins as t_1 and some time right after the transfer as t_2; take the person that one sees in c as p_1 and the person that one sees in the new body as p_2). Hence I am not identical with c; for I exist, with the new body, when c has been destroyed. And the circumstances we imagined are logically possible.

Let’s now accept that it is logically possible that I should not be identical with c. We can draw the further conclusion that, in actual fact too, I am not identical with c. For I have the following property: I am something for which it is logically possible not to be identical with c. But c does not have that property: it is not something for which it is logically possible not to be identical with c. If, now, something x and something y are identical, then every property that x has is a property that y has; this principle, which is generally accepted, is called ‘Leibniz’s principle of the indiscernibility of identicals’. Consequently, c and I are not identical.

(Arguments like the one in the preceding paragraph can be used in various topics in metaphysics.)

5. The view that I am not identical with my body may appear to contravene materialism. In fact, however, it is a view that materialist philosophers can accept. Physicalism is the main form of materialism today. It is the proposition that everything that exists is a

physical entity. An entity is characterized as *physical* iff the question what kind of entity it is and how it differs from other entities of the same kind admits of an answer that contains only terms of the natural sciences (irrespective of whether we know that answer). Of course, physical entities are governed by the laws of nature. Physicalists usually say that thoughts, decisions, etc., are physical entities, e.g. exchanges of electrical signals in the brain. They can accept that a person is not identical with their body, but add that both entities are physical.

This combination may sound odd. However, a distinction similar to that between me and my body can also be made about many lifeless entities, such as statues and tables. Let's say that S is a clay statue of a man and C is the corresponding piece of clay. If C was first cut out of a bigger clay mass, and then the artist gave it the desirable shape in order to create the statue, C pre-existed S. Again, if we cut an arm from the statue, C (the particular piece of clay) will cease to exist, but S will continue to exist amputated. So if, for one reason or another, S and C do not have the same duration in time, then they are not identical.

Even if they have the same duration, we can apply a type of argument that we saw earlier. We now consider that in fact S and C begin existing simultaneously and also stop existing simultaneously. It is, nevertheless, logically possible that someone should have cut a small piece out of the statue and, indeed, should dissolve that piece completely in some way. It is therefore logically possible that S should continue to exist without C existing. Hence, in reality too, S has a property that C lacks: it is something for which it is logically possible to exist at a time when C does not exist. Consequently, S and C are not identical.

There are of course difficulties with the view that S and C are not one and the same object. If, for example, they are not identical, and each one has a weight of 50 kilos, why isn't their total weight 100 kilos? Still, owing to the arguments of the preceding paragraphs, many philosophers agree that S and C are not identical. But at the same time they can accept that they are both physical entities. (And if that doesn't sound plausible for statues, it is more plausible for tables.) They usually say that C *constitutes* S.

6. B Williams conjured up two imaginary scenarios and argued that their intuitive treatment leads to opposite conclusions about the problem of personal identity.

In the first scenario, we start with two persons, A and B, and subject them to the following experiment. We erase all the information that existed in A's brain and transfer it to B's brain, while simultaneously we erase all the information that existed in B's brain and transfer it to A's brain. After the changes, we have a person with A's body and a person with B's body. As a result of the experiment, the impressions that the person with B's body has about her past coincide with the memories that A had before the experiment, and the actions of the person with B's body fit in with the character that A had before the experiment. Analogous points are true of B and the person that now has A's body.

Intuitively, the person with B's body is A, and the person with A's body is B. This would become particularly clear if, after the experiment, we asked them various questions about what they remember, whether their expectations about the experiment came true, etc. The words of the person with B's body would be those we expected from A, and the words of the person with A's body would be those we expected from B.

In the second scenario, now, someone is told that his body is going to be subjected to torture. He of course feels fear. In a short while, it is announced to him that, before

the torture, he will suffer amnesia. He fears even more. And, finally, he is told that, after the memories he has are erased, they will be replaced by impressions copied from someone else. His fear reaches a peak.

Does that man rightly fear the torture or should he think, “Since all those interventions will take place in my mind, I shall cease to exist. Another person will have my body. Consequently, I have many reasons for fearing, but at any rate I needn’t feel fear of torture and physical pain”? Intuitively, it seems that this man rightly fears torture too; it seems that he rightly expects to suffer physical pain, which in fact will have been preceded by serious mental derangements.

Is it possible for both the intuitive treatment of the first scenario and the intuitive treatment of the second scenario to be correct? One may hold that it is. One may hold that if the information from B’s brain is transferred to A’s brain and conversely, then A and B exchange bodies, so the person with A’s body is not A; but if the information from B’s brain is transferred to A’s brain without there being any corresponding change in B, then the person with A’s body is A, although he now has a different psyche; the second scenario gives just the impression that there is no corresponding change in the person from whose brain the information is copied. Williams rejects that view. His argument is that once the information is copied from B’s brain into A’s, then, whether or not there is a corresponding transfer of information from A to B, the person with A’s body is similar in the two cases as regards his inner world and the origin of his inner world; hence it cannot be that, in the one case, he is identical with A but, in the other, he is not identical. So Williams concludes that it is not possible for both our intuitions about the first scenario and our intuitions about the second to be correct. He prefers, though, our intuitions about the second, as well as the criterion of personal identity to which they lead, that is, the criterion that x and y are identical iff they have one and the same body. He seems to prefer them because he considers that they provide a reasonable treatment of more imaginary scenarios.

7. D Parfit considers that x (who exists at the time t) and y (who exists at the later time t´) are identical iff the following conditions are satisfied: (i) there is psychological continuity from x to y; and (ii) there is at t no person, other than x, from whom there is psychological continuity to y, and there is at t´ no person, other than y, to whom there is psychological continuity from x. Thus Parfit agrees with the intuitive treatment of the first of Williams’s two scenarios.

We can imagine cases of fission. E.g. my body, together with my brain, is destroyed, but the information stored in it is copied into the brains of two new, artificially made bodies. Or the left hemisphere of my brain is put, together with a copy of the right, in a new body made with no brain, and simultaneously the right hemisphere is put, together with a copy of the left, in another new and brainless body. After such an operation, there are two persons, B and C. According to Parfit, I am not identical with B. For condition (i) is satisfied (there is psychological continuity from me to B), but condition (ii) is not satisfied (because of the presence of C). And for the same reason I am not identical with C either.

We can also imagine cases of fusion. E.g. both my own body, together with the brain, and D’s body, together with its brain, are destroyed, but a mixture of the information in the two brains is copied into a new body. Let’s suppose that this mixture makes up a unified and healthy mental life. According to Parfit, neither I nor D is identical with the person E that exists after the information has been copied. For, even if condition (i) is satisfied, (ii) is not.

Parfit distinguishes between identity and survival. He considers that, in the example of fission, I survive in B and also survive in C. But it's logically impossible for me to be identical with both of them. In the example of fusion, he considers that both D and I survive in E. But, once more, it's logically impossible for both D and me to be identical with E. He adds that survival admits of degrees while identity does not. That is, at a future time, I may survive to a great degree (if the psychological continuity between me now and a person that will exist then is tight) or to a small degree (if the psychological continuity between me now and a person that will exist then is slack — because e.g. there intervene some abrupt psychological changes). On the other hand, it makes no sense to say that x and y are one and the same person to a great degree (or to a small degree).

Here are some difficulties for Parfit's views:

(a) Let's suppose that my body, together with my brain, is destroyed, but the information stored in the brain is copied into the brain of a new body. Or, alternatively, let's suppose that the left hemisphere of my brain, together with a copy of the right, is put in a new body made with no brain. B´ is the person that exists after the operation. According to Parfit, I am identical with B´, although, in the fission scenario, I am not identical with B. Yet B and B´ are, at least right after the respective operations, exactly similar as regards their mental features and body, and they are also exactly similar as regards the origin of those mental features and body. With such similarity, is it justified to say that, in the one story, I am identical with B´ while, in the other, I am not identical with B?

(b) Let's suppose that a machine creates an exact replica of my body and brain, molecule for molecule; this replica behaves in accordance with my character, his impressions about the past coincide with my memories, etc.; but at the same time my body and brain undergo no alteration and continue their activity like before. Intuitively, we would say that I continue to exist, but together with me there is a second, similar person. Yet, according to Parfit, I am not identical with either of the persons that exist after the machine has copied me.

(c) In Parfit's view, it is logically impossible that someone should exterminate me but, simultaneously, all the information stored in my brain should be transferred into one, and only one, new body, so that the person having that body behaves in accordance with my character, possesses impressions about the past that coincide with my memories, etc. It is logically impossible because if all the information were transferred with that effect, I wouldn't have been exterminated, I would continue to exist. Intuitively, though, it's not clear that the setup I described is logically impossible.

(d) What Parfit says about survival is very problematic. Most people consider it conceptually more accurate to say that survival presupposes identity: if, at a future time, I survive in someone x, then x and I are one person (who exists both now and at that time). In Parfit's view, it is logically possible that, at a future time, I should survive but there should be no one with whom I am identical. This clearly seems to be logically impossible.