Philosophica 72 (2003) pp. 131-152

PERSONAL IDENTITY AND ITS BOUNDARIES: PHILOSOPHICAL THOUGHT EXPERIMENTS

Farah Focquaert[1]

ABSTRACT

The philosophical writings on personal identity through time are known for their extensive use of thought experiments. In the following article I will not ask myself the question if philosophical thought experiments in general are a useful tool or not. What I will do is examine if thought experiments within the philosophy of personal identity through time fulfil their job. Brain swaps, brain state transfers, teletransportation, fission and fusion are common among them. Their purpose is to bring us closer to a definition of personal identity through time, by giving us more insight in the sufficient and necessary conditions of the concept. I will examine some of the most formulated and reformulated thought experiments used in the debate and try to come to terms with some widespread misinterpretations that join them. It's my opinion that those misinterpretations result from a lack of scientific background knowledge. Without considering the current results in neuroscience, biology and other relevant sciences it's very difficult to get a firm and justified grip on the subject. Without this backing scientific knowledge a lot of thought experiments on personal identity go astray. I will argue that only those thought experiments are legitimate, that take into account the neurological and biological facts that join them.

1. Introduction

Most philosophical thought experiments have been put under great pressure concerning their usefulness. These doubts also reached the

[1] The author is Research Assistant of the Fund for Scientific Research - Flanders (Belgium).

debate on personal identity through time[2]. Kathleen Wilkes (Real People, 1988) especially reasons against the use of thought experiments in this domain. She wants us to focus on real-life cases, thus leaving - what she believes most thought experiments on personal identity (through time) to be - science fiction and fairy tales aside. Perhaps it would indeed be better to dismiss those thought experiments and stick to the actual world. However, Wilkes's own arguments aren't strong enough to support this. Her argumentation doesn't support her conclusion. On the contrary, I believe it indirectly guides us into making philosophical thought experiments a lot less science fiction and a lot more decisive than she claims them to be.

In general Wilkes wants to defend the view that thought experiments are, as she puts it, highly misleading as a philosophical tool. She wants us to focus on actual cases, which she claims to be often stranger than fiction. According to Wilkes we don't need thought experiments to explore the limits of our personal identity, and especially not those that exceed known theoretical boundaries. (1988: 1) The latter is indeed the most basic problem we're confronted with when we start to study thought experiments on personal identity through time. Philosophers taking part in the debate often don't know the nomological boundaries of the issues they're concerned with. And if they do know them, they often use them in a mistaken or misleading way.

A thought experiment refers to an imaginary phenomenon established in a possible world. A possible world in which that phenomenon is held to exist. How does Wilkes define a good thought experiment? When can we consider a thought experiment to be useful, instructive or illuminating? Sören Häggqvist takes the factor *relevance* to be the most central theme of Wilkes's criticisms. (Thought Experiments in Philosophy, 1996: 21-22) To know and define the *relevant* background conditions, impossibilities and possibilities of a particular thought experiment is indeed what Wilkes seems to hold crucial for deciding upon

---

[2] There's an important difference between personal identity and personal identity *through time* that has to be taken into account. I believe a cognitive neuroscientific theory on personal identity through time to be a reasonable goal for the near future. I cannot however account for the same thing to be possible with respect to personal identity as such. So I believe Wilkes' criticism holds for personal identity as such but not for personal identity through time.

its usefulness and illuminative force. According to her a thought experiment can only be useful when it entails nomological possibilities[3]. Knowledge about the relevant background conditions, possibilities and impossibilities makes us aware of the nomological boundaries a thought experiment is dealing with. Impossibilities only pose a threat if they're considered to be *relevant* impossibilities. Irrelevant impossibilities do not matter.[4] What exactly does Wilkes mean by all of this?

2. Background Conditions, Nomological Boundaries and Natural Kinds

Thought experiments count as experiments that can't be realised in the real world. According to Wilkes, they nevertheless have to follow a lot of rules applicable to experiments that can be. First of all, we have to know what is altered in the real world and what isn't. According to Wilkes, this constraint is especially important for their success or failure. When we construct a thought experiment, we have to make sure that all background conditions are in place. The background conditions refer to the *relevant* elements of the possible world against which our phenomenon is being described. They have to remain constant in order for us to determine the impact of the imagined phenomenon. How do we know which conditions are *relevant*? According to Wilkes we often follow our common sense as to what conditions are relevant. Wilkes agrees that our commons sense can indeed be very useful. She also refers to the importance of the *function* or *purpose* of thought experiments to describe their relevant background conditions. It depends on the phenomenon you want to establish, she claims, which background elements are in place and which aren't. Some 'impossibilities' within

[3] I'm aware of the fact that Wilkes - as Häggqvist mentions - uses the notion of theoretical possibility somewhat unstable. I believe however, that Wilkes always intended to use the notion *theoretical possibility* as meaning *nomological possibility*. Nomological possibility is expressed by scientific theories and laws. In this paper, both terms have the same meaning. This also leads to a more consistent interpretation of Wilkes' writings.

[4] Wilkes adds that some impossibilities don't give rise to problems, because their no threat to the situation we want to examine. The impossibility doesn't affect the factor that we want to investigate, so it doesn't pose any problems for the conclusions we draw.

thought experiments give rise to inconsistencies and affect the plausibility of your conclusions, others don't have any affect at all on the conclusions you want to draw. The purpose of your thought experiment helps you to establish the relevance of most background conditions, especially those that can't easily be mastered by our common sense.

Wilkes claims that experiments in science often have a background that is supported by scientific theory. That way it's easier to determine which factors are relevant for your conclusion and which aren't. This follows from the fact that science deals with 'natural kinds'. Science uses natural kind terms to formulate its laws and to make systematic generalizations. It aims to have a tight grip on the concepts it uses and by doing so makes it possible to draw very sharp and secure conclusions. Scientific theories work with natural kinds as their building blocks. Natural kinds are concepts that are well defined within scientific research, because we have laws and generalizations about them. So we know what kind of implications and entailments we can derive from them. Using those terms in thought experiments gives us a firm grip on their relevance or irrelevance within the postulated framework. The problems arise when we start looking at the concepts that are being used in philosophical thought experiments. According to Wilkes, they hardly ever work with natural kinds. Their concepts are often embedded in common sense thinking and consequently no clear implications and entailments can be drawn from them. We don't know any scientific theories or laws that back them. It's more than possible that we ascribe false implications to these concepts and by doing so we're putting the usefulness and illuminative force of thought experiments on a hold. As Wilkes points out, the implications we draw from common sense thinking are irreducibly context dependent, nuance ridden, purpose, speaker and audience dependent. Because of that, it's difficult to determine which background conditions are relevant and which impossibilities of the postulated framework jeopardize the conclusions drawn. So philosophical thought experiments often go astray because they're not using natural-kind terms, terms of which we more or less know the implications and entailments that derive from them. Thought experiments on the concept of personal identity (through time) face, according to Wilkes, just that problem. (1988: 15)

3. 'Agreed' Intuitions and Justified Conclusions

Thought experiments work with imagination and intuition whereas real-life experiments work with observation and measurement. In stead of relying on measurements, we have to trust our intuitions to make sound conclusions about the imaginative world we're dealing with. Wilkes talks about relying on 'agreed intuitions'. There aren't a lot of problems with thought experiments in science she says, for instance in biology or physics, but when we consider philosophical thought experiments again the problems begin. Thought experiments in science usually have their well defined background situation and theories against which the imagined phenomenon is measured. So they experimenter only has to make the right inferences from the scientific theories that support the situation. The 'agreed intuitions' in these cases are the scientific laws and theories that accompany the imagined situation and can be used to draw useful and instructive conclusions about the imagined phenomenon. When we use common sense concepts to describe our imagined world, we don't have those backing scientific theories, so we're not able to draw such unproblematic conclusions as is the case in most thought experiments in science. Common sense concepts and common sense thinking can give rise to very different outcomes, that don't coincide at any given point. It's not very rewarding to rely on 'agreed intuitions', when these intuitions are found to be extremely diverse. And indeed, that is the case with most contemporary thought experiments in the philosophy on personal identity through time. The conclusions drawn from the very same thought experiments give rise to very different positions within the debate. So it's indeed a big problem that at this very point those experimenters don't have backing scientific theories to turn to. Or maybe they do have them, but aren't interested in them.[5] As Wilkes points out, we're not interested in the bare logical possibility of some or other phenomenon. A mere logical possibility doesn't tell us anything about what could or might happen to your or my identity. I follow Wilkes when she says that we should look for theoretical or 'in principle' possibilities.

[5] It's a matter of fact that there are a lot of scientific facts available that can be helpful for the philosopher that conducts thought experiments within the domain of personal identity through time, but most philosophers don't take the biological and neurological facts into account. And if they do, they are often used in a misleading or mistaken way.

We should ask ourselves if our background conditions are in agreement with our scientific knowledge about such concepts and situations, because that's what makes our thought experiment valid, especially when we consider such a concept as our personal identity.[6] When the background of a thought experiment isn't sufficiently described, we risk overlooking some fundamental elements that would count against the conclusions we reached. As I already mentioned, not all impossibilities break down our line of argument. Only those that are relevant for the phenomenon we want to establish, should be taken in due account. Only when we have an adequate background description, can we hold our phenomenon to be *established* and is it possible to draw useful and illuminating conclusions.[7]

Wilkes addresses a difficulty in taking existing scientific knowledge as our parameter in separating theoretic possibilities from theoretic impossibilities. Because scientific knowledge is fallible and theories go through numerous transformations and are sometimes overthrown by

---

[6] Häggqvist argues that 'the purpose of many philosophical thought experiments appears, at least prima facie, to be to investigate "conceptual" theses by studying counterfactual situations' (1996: 30). He then asks himself the question if it is necessary that those situations be theoretically possible. Of course the answer to that question would be 'no' given their goal to investigate 'conceptual theses', but I believe that the current literature on personal identity through time is trying to establish something more than a conceptual framework. As Parfit mentions, the different views about personal identity make claims about actual people with actual lives. (Reasons and Persons, 1987) A lot of the thought experiments draw conclusions that are supposed to be applicable to actual people, and ordinary lives. They have to be theoretically or nomologically possible, or they'll no longer be talking about actual people and ordinary lives. If you want your conclusions to be applicable to actual people, you have to make sure that your thought experiment doesn't transcend the laws of human beings and persons. That's why it's necessary that thought experiments on personal identity through time obey Wilkes' demands. The most prominent philosophers taking part in the debate, use their writings on personal identity through time to delineate and defend their ethical position. They're trying to grasp the necessary and sufficient conditions of personal identity through time in order to defend their moral theories applicable to real people and real situations.

[7] I would like to add that mere technological impossibilities also don't jeopardize the conclusions we draw from a particular thought experiment. As long as we stick to nomological possibilities we don't undermine the establishment of our phenomenon. So any theoretic possibility that doesn't contradict our scientific knowledge, counts as an established possibility from which we are able to draw justified conclusions.

newly discovered facts. According to Wilkes we have to take the five following situations into account when we believe scientific theories to be the hallmark of justifying our thought experiments. (1988: 19)

(1) In the first situation we have a scientific theory that is generally accepted and provides an adequate background for our imagined phenomenon.
(2) In the second situation, this theory is contested by a rival theory that is equally supported and influential.
(3) In the third situation we only have one theory that also is generally accepted by the scientific community. The only problem with this theory is that it's not known outside the circle of scientists. Only the members of the scientific community will know about its existence.
(4) We have two rival theories, but not one of them is generally accepted. There is still a lot of research to be done, before one or the other will be a valid scientific theory.
(5) In the last situation, we again have one scientific theory that is generally backed by the community, but which in reality is false. At that particular moment no scientist is aware of the invalidity of the theory.

Wilkes claims that situation (3) is especially applicable to thought experiments on personal identity through time, because philosophers often lack and aren't that concerned with the biological and neurological knowledge that's available. Also situation (4), says Wilkes, is likely to be the case for those thought experiments. (1988: 19-20)

4. Personal Identity through Time as a Natural Kind Term

I fully agree with Häggqvist when he claims that Wilkes's arguments 'are not strong enough to motivate her broad rejection of philosophical thought experiments' (1996: 27). Indeed, I don't find any decisive arguments in Wilkes dissertation that should bring us to reject philosophical thought experiments. Her argumentation is mainly that philosophical thought experiments work with common sense terms that don't refer to any known scientific theories. However, as in the domain of personal identity through time, there are a lot of scientific facts we can

turn to. And maybe it's true that most scientific theorizing about personal identity through time at this moment reflects situation (3) or (4) in Wilkes discussion, but the fact remains that there is knowledge out there that's relevant to thought experiments on personal identity through time and more importantly it keeps accumulating. So maybe we don't have the backing theories yet, but we will surely have them in the future. Then we'll be able to draw trustworthy conclusions on personal identity through time. Wilkes herself implicitly mentions that a lot of research is to be done when we're talking about the domain of personal identity through time.[8] So she can't claim that personal identity through time will never be discernable as a natural kind term. When it will be discernable as a natural-kind term, it will no longer be possible to dismiss thought experiments on personal identity through time because of the poorly defined background those experiments are facing. And at the moment a lot of scientific knowledge on split-brains, individuals with brain-damage and and ongoing research on the functional mapping of the brain, already provides us with information that makes us able to discern what conclusions can and can't be drawn and also which impossibilities are or aren't relevant. In the present debate, more and more attention is drawn to facts and examples arising from scientific sources as the cognitive neurosciences and biology in general. However, it's my opinion that this line of research isn't fully exploited in the philodophical debate on personal identity through time. As a result, a lot of misinterpretations and maybe even deliberate misreading dominate the debate. I believe that no useful philosophical research on the necessary and sufficient conditions of personal identity through time can be done, without taking into account the neurological, biological and evolutionary boundaries that a concept like our personal identity entails. Most philosophers don't ask themselves the question what it is like for *me* to have a personal identity, instead they ask themselves the question what it takes for me to be the same person tomorrow as I am today. What facts make me the same person today as I will be tomorrow. The only valid answer to be given, has to entail knowledge accumulated within the neurological, medical and evolutionary sciences. At present, there is not enough information available to define

[8] This is, when she takes thought experiments on personal identity through time to reflect situation (4).

personal identity through time as a natural kind. We only have scattered pieces of information available to us, like scientific research on split brains and people with different kinds of brain damage. Some findings in neuro-psychology and recent findings in the field of neuro-imaging are also very helpful.[9] At present, no coherent and fully detailed account can be given of the neuro-cognitive structures that are responsible for our personal identity trough time. (Miller, 2001) I am strongly inclined to define our personal identity as an adaptive function or system and not merely as a side-effect of the way our brains are built. I believe our personal identity is an evolved system that makes it possible to present ourselves to others and makes us able to consider others as coherent entities. At present there is no consensus on what kind of phenomenon our personal identity addresses, whether it is adaptive or not. It would be interesting to find out. (Tooby & Cosmides, 2000: 1163-1165)

5. Meeting Wilkes's Demands

I believe that thought experiments on personal identity through time can meet Wilkes's demands concerning justified and useful thought experiments in general. Her argumentation doesn't force us to dismiss the thought experiments on personal identity as a philosophical tool. It only asks us to put some restrictions on their use. The relevant possibilities have to be nomologically possible. Technical impossibilities, as well as irrelevant impossibilities don't jeopardize the value of the thought experiments. Most recent thought experiments on personal identity through time already implicitly use the notion of nomological possibility. So, in a way they're already responding to Wilkes's demands. I think this notion should be made explicit. I believe it would make an end to a lot of unnecessary discussion and give the debate the progress it needs. If it's the general idea to talk about real people and ordinary lives, we have to stick to what's nomologically possible. Some 'impossibilities' are relevant and others aren't. I believe this much is true. When we consider the case of a person splitting amoeba like into two new people, we can

[9] Wheeler M. et al. (1997), Klein S. et al. (1996), Craik F. et al. (1999), Klein S. (2001), Llinás R. (2001), Miller B. et al. (2001), Tulving E. (2002).

indeed postulate that the impossibility of that operation destroys the validity of the thought experiment. This is because we're asking ourselves what would happen with the personal identity of a person when he or she splits into two new people. We can't draw any legitimate conclusions about that kind of situation, because we don't know and we'll never know what will happen to that person. It will never be possible to perform such an operation, because our biological and neurological features don't allow it. It's much more then a mere technical impossibility. The concept of a person and even of a human being doesn't reach that far and you can't stretch it that far either. That kind of situation doesn't obey 'the rules of the game'. It oversteps the boundaries of the concept it wants to delineate and as a consequence it can no longer appeal to it. When you're talking about amoeba like splitting of persons, you're not really talking about persons. Maybe you're talking about Martians or some other kind of unknown species, surely not about real persons like you and me. When you take yourself to be talking about actual people and ordinary lives, you can't fail to respect the boundaries those actual people and ordinary lives bring with them. In order to use thought experiments as a philosophical tool on personal identity through time you have to be aware of the neurological boundaries such a concept brings. I will now look a bit more closely at some thought experiments on personal identity through time and see if they can meet the demands required.

5.1 Brain State Transfer: Nomological Impossibility

Bernard Williams admits that bodily continuity isn't enough to guarantee our personal identity through time, but he does consider it a necessary condition. So without bodily continuity we aren't entitled to refer to personal identity through time. In 'The Self and The Future' Williams (1970) describes a thought experiment about two persons A and B that could be described as a body swap. After the experiment person A acts surprisingly B-ish and person B surprisingly A-ish. It seems that person A received all the characteristics of person B and vice versa.[10] Williams

[10] We have to take into account, according to Williams, that the persons before the experiment should be sufficiently alike physically and psychologically, otherwise it would be very difficult to think about person B, when confronted with person A. We couldn't imagine something like that.

refers to Shoemaker's (1963) 'Self-Knowledge and Self-Identity'. Shoemaker describes the thought experiment as two persons A and B undergoing brain-transplants and ending up with each others brain. Williams acknowledges that this is a legitimate way to describe the imagined situation, but he has something else in mind. Williams asks us to imagine some kind of device that enables us to retract and store the information of someone's brain. The experiment would then amount to running the information of A's brain on B's and vice versa. He calls the resulting persons the A-body person and the B-body person. The A-body person would have A's body and would be referred to as person A by someone who wasn't aware of what just happened. The same would count for the B-body person that has B's body. Williams asks himself if we can really describe this thought experiment as 'two persons changing bodies'. Are we entitled to call it a 'body swap'? That would imply that the A-body person actually was B and the B-body person actually A. He claims that a non-question-begging description would leave it open which person the A-body or the B-body person was.

What would happen if two persons A and B take part in the experiment? Suppose they we're told that the A-body person will get $ 250 000 after the experiment and that the B-body person will be tortured. We then ask both person A and B to choose on selfish grounds which outcome they wish the most. Should person A choose that the B-body person gets the money and that the A-body person be tortured and vice versa, we should indeed be entitled to call the experiment a body swap. Because person A en B both believe that they'll receive each others body. Confronted with such an experience we're indeed very strongly inclined to call it a body swap, this because we hold the identity of a person to be situated and secured in a person's brain. When we ask ourselves what would happen when person A chooses that the A-body person gets the torture and the B-body person the money and in reality the experimenter performs the experiment the other way around; we are entitled to conclude that the B-body person wouldn't be satisfied with the result. It wasn't what he asked for before the experiment.[11]

However Williams describes a situation in which it's not that straightforward to conclude that the body doesn't matter for personal

[11] This is also the conclusion S. Shoemaker makes in 'Self-Knowledge and Self-Identity'.

identity through time and conversely that only the brain does. Suppose that someone tells you that you will be tortured in the nearby future. Of course you will be angry and afraid. However the person also tells you that you will undergo some changes before the torturing actually takes place. First of all, I won't remember being told about it. Of course I'm still scared, this doesn't change a thing. Second, I won't have any memories left of my life before the torturing. This still doesn't cheer me up, I'll still be able to feel the pain. Third, I will have a whole new set of memories about my past, as if I was a different person. This also doesn't make me feel better. I can imagine myself going crazy at a certain point in my life and believing myself to be Napoleon, but nevertheless I will still feel the pain. Fourth, what if my new memories and impressions of my past belonged to a real person? The information of someone else's brain would then be transferred to mine. Even when this happens, Williams still believes I would be afraid of the torture awaiting me. Ok, this is really strange. If this line of thought adds up, why should I be afraid in this situation, but not in the former 'body swapping' situation? Exactly the same things are happening to me. Williams holds the latter situation to be an argument against the idea that my bodily continuity doesn't matter, and only my brain continuity does. We have to keep in mind that in the second description Williams didn't refer to the other person that would receive my brain information. He keeps referring to what would happen to *me*, while we should be careful to describe the situation as such. In the second situation, won't the other human being that receives my brain information be *me*? If this is the actual situation, I wouldn't have to be afraid of the torturing happening to *me*.

According to Williams it doesn't matter that he doesn't mention the other person who's receiving my memories, values, characteristics, etc. It's not a problem that his description is or isn't neutral, the real worry is that we can perfectly follow each step it brings us to. After each step I can still ask myself if I should be scared of the future and that's what matters according to Williams. He believes it supports the principle 'that my undergoing physical pain in the future is not excluded by any psychological state I may be in at the time, with the platitudinous exception of those psychological states that in themselves exclude experiencing pain, notably (if it is a psychological state) unconsciousness' (1970: 11).

(i) A is subjected to an operation which produces total amnesia;
(ii) Amnesia is produced in A, and other interference leads to certain changes in his character;
(iii) Changes in his character are produced, and at the same time certain illusory "memory" beliefs are induced in him; these are of a quite fictitious kind and do not fit the life of any actual person;
(iv) The same as in (iii), except that both the character traits and the "memory" impressions are designed to be appropriate to another actual person, B;
(v) The same as in (vi), except that the result is produced by putting the information into A from the brain of B, by a method which leaves B the same as he was before;
(vi) The same happens to A as in (v), but B is not left the same, since a similar operation, is conducted in the reverse direction. (1970: 14)

Williams points out that the only difference between the two situations amounts to step (vi). According to Williams's opponents this is a very crucial element that's missing. But Williams doesn't really see how it can make all that difference. Why is it appropriate to be scared until (v) and no longer when (vi) is introduced? The situation A faces doesn't change from (v) to (vi), the only change that occurs affects someone else, namely B. The only way in which (vi) could change the situation for A, is when we assume that in (vi) A survives as the B-body person. A should be one and the same in (v) and (vi), but when we assume that in (vi) A survives as the B-body person, this isn't the case. If the A-body person isn't A in (vi), then A also can't be the A-body person in (v). Who then, is person A in (v)? A simply doesn't seem to exist in (v). Next we can ask ourselves the question if A still exists in (iv), or (iii), (ii), (i)? Where do you draw the line?

What should we conclude? That the thought experiment really amounts to two persons changing bodies or should we conclude that Williams's principle stands stronger? Maybe not, but then Williams dares us to show what is wrong with it. Unless this challenge has been answered, Williams believes it's better to choose that the B-body person receives the torturing. However, this is not without risk.

What can we say about this thought experiment? Does it meet Wilkes's demands? It's quite obvious that it doesn't. A brain-state

transfer isn't "nomologically possible" and as a consequence can't be part of a justified philosophical thought experiment on personal identity through time. This nomological impossibility is essential to the outcome of the thought experiment. It's only because of the fact that we could take the information of someone's brain and slice it up, that Williams was able to mislead us with his second description of the brain state transfer. When you take the whole brain into account, Williams wouldn't be able to formulate his second description. A person's brain is essentially its information. You can't separate both in nomological terms. As a figure of speech: you wouldn't be able to run the software on a different hardware. The structure of a person's brain and the information that goes with it is unique. You can't find two brains exactly alike. So maybe we should restrict ourselves to using the idea of 'brain swaps', in stead of 'brain state transfers'. We can turn to Sydney Shoemaker's description in his 'Self-Knowledge and Self-Identity'. Here goes.

5.2 Brain Transplantation: Nomological Possibility

Shoemaker refers to brain transplantations as, at least, logical possibilities. It might not be possible according to physiologists, but we can conceive of a brain transplantation occurring. Suppose next, that it is possible to elevate a person's brain from his skull, leaving the brain intact. It would enable surgeons to perform operations on it and help curing brain diseases or removing tumours. Suppose one day the surgeon's assistant made a big mistake. Instead of putting Mr. Robinson's brain in his head, he puts it in Mr. Brown's head and vice versa. Suppose that one of them immediately dies. After a few hours the other person, with Robinson's body and Brown's head, awakes. Shoemaker calls him 'Brownson'. When Brownson looks in the mirror, he freaks out. Who's body is this? And where is mine? Brownson appears to have all the characteristics, memories, desires, goals, principles and so fourth of the person Brown before the operation. Shoemaker concludes that, after the experiment, we're strongly inclined to call Brownson Brown, despite the fact that he has Robinson's body. However, we could just as well reach the opposite conclusion. We can't know when we consider the argument from a third person position, because then our answer already relies on the kind of personal identity view we have. The value we address to psychological factors or bodily factors will be reflected in the way we

evaluate the imagined situation. You can't be sure of the result when you reflect on it from a third person perspective. Both answers have equal weight. This results in some kind of ambiguity towards our concept of personal identity through time. Shoemaker believes that there can't be any ambiguity about the concept of a person. A person 'is' or he 'isn't', we can't allow any uncertainty to be the case. So according to Shoemaker we have to look at the thought experiment from a different angle. We should ask 'Brownson' who he believes himself to be and describe the outcome from the first person perspective. Brownson will of course have some memories of which person he is. It won't be ambiguous for Brownson to decide who he is. His memories will tell him.

> It seems to me that I can imagine being in the position of the Brownson of my example. I can imagine waking up after an operation and being surprised by the appearance of my body (e.g., as seen in a mirror). I can imagine seeing some other body, which I recognize (or seem to recognize) as my body of the previous day, and being told that the brain from that body had been placed in the skull of my present one. (1963: 32)

When it really is the case that only Brownson himself can tell us unambiguously who he is, we should conclude that our memories of whom we are, are essential for our personal identity through time. Shoemaker points out that our memories aren't based on the kind of body we posses. Our memories about who we are, aren't in any direct way linked to our bodily identity. We have those memories or we don't, whatever body we posses. Shoemaker claims that the first person perspective supports the memory criterion for personal identity through time.

Shoemaker also refers to Williams's argument about brain state transfers. He believes Williams's dissertation to be mistaken and reveals how we should view the thought experiment. Shoemaker refers to the thought experiment as the *Brain State Transfer* - procedure or BST-procedure. (Shoemaker & Swinburne, 1984) As I pointed out before, we have to refute this kind of thought experiments because they contravene the laws of nature and thus have to be regarded as nomologically impossible. They are not immune to Wilkes's criticisms and therefore can't be regarded as valid thought experiments. They don't bring us any factual knowledge about actual persons and ordinary lives and they never

will.

What is wrong with the 'brain swap' thought experiment? Following Wilkes we have to have knowledge about the background conditions of a particular thought experiment to be able to formulate justified conclusions about it. The relevant background conditions should be 'in principle' possible. Our backing scientific theories should allow for those conditions to be possible. We have to ask ourselves the question what could or couldn't happen given our backing scientific theories. Of course the main problem with those thought experiments is that they hardly ever refer to or hold in mind the existing scientific knowledge available. We first have to ask ourselves the question if, giving our scientific theories about human beings, it's possible that someone's brain is transplanted to someone else's body. At the moment this surely is a technical impossibility, but is it also a theoretical or 'in principal' possibility? Can we believe it to be nomologically possible that my brain gets transplanted into your skull?[12] Does it contradict the physiological and neurological laws regarding human beings? I believe a 'brain swap' to be only a technological impossibility. As Wilkes points out, investigators in the USA, already switched monkey heads successfully. It's possible to (re)wire a monkey's head to a different monkey's body. Nerve cells automatically rewire when they're brought back together within a certain distance and a certain time span. So we can conclude that with the right technological equipment it probably would be possible to perform such a 'brain swap'. Of course we aren't entirely sure, but if we have to choose between it being nomologically impossible or nomologically possible, I would choose the latter. According to the knowledge we now have, a 'brain swap' doesn't contravene the laws of nature. Of course, it could remain a technical impossibility forever. However, being nomologically possible, does our scientific knowledge grant us the conclusions Williams or Shoemaker draw from the thought experiment in question? Can we believe it to support either a bodily criterion or a psychological criterion of personal identity through time? I believe our scientific knowledge, this would be knowledge coming from the cognitive neurosciences, doesn't allow us to reach either of these conclusions. It doesn't seem so obvious that the thought experiment supports the memory

[12] I think we have to bear in mind here, that we can't switch between sexes.

criterion of personal identity through time. It's not because our memories make us able to know who we are, that they are constitutive of our personal identity through time. We know for a fact, that we need so much more. Even if they could, how can we be sure that my personal identity won't change when I receive a new body? Couldn't it be possible that I receive a whole new identity? Our brain may be the carrier of our personal identity, but it's not a sufficient condition for it. Our bodies make us able to act and have experiences in this world. Those experiences are essential for who we are.

> As we grow up each of us is exposed to different combinations of stimuli and develops motor skills in different ways. Thus all brains - even the brains of identical twins who share some genes - are uniquely modified by experience. This distinctive modification of brain architecture, along with a unique genetic makeup, constitutes a biological basis for individuality. (Kandel et al., 2000: 1274-75)

In Nicholas Humphrey's (1993) *A History of The Mind* we also come across the importance of our body for our 'self'. Humphrey tries to come to terms with what it means to be conscious. According to Humphrey to be conscious is essentially to have *sensations*. *Sensations* are subjective experiences, like the pain we feel after accidentally biting our tongue, or the experience of redness we have when looking at a red rose. Humphrey describes *sensations* as affect-laden mental representations of something happening here and now to me. What does this mean? The most important feature of *sensations* is that they characteristically belong to the subject that's having them. My *sensations* are undoubtedly my own, they can't be anybody else's. A *sensation* represents what's happening to me and thus what's happening to my embodied self. So it depends on the body I have, what my *sensations* will be like. Associated with another body, they would be different *sensations*. The pain I feel when biting my tongue, can never be the pain you feel when biting your tongue. The body I own is thus essential for the conscious experiences I have. Humphrey holds that both bodily movements and *sensations* on the one hand refer to a particular place on the body and on the other hand create a physical disturbance at that place. So *sensations* should be some kind of bodily movements in their own right? They have to create a physical disturbance at a certain place of my body, like the pain I feel in my

tongue when biting it. What about phantom limbs? A lot of cases are known of people who still feel the presence of an amputated limb, even years after the amputation took place. We can't say these *sensations* create an active disturbance at that specific place of their body, because it no longer exists. Phantom pains teach us that the physical disturbance a *sensation* creates takes place at a surrogate location. Our brains are equipped with a 'model' of our bodies, a kind of neurological map of our body space. In the case of human beings the physical disturbance no longer occurs at our body surface. It occurs at a surrogate place in the brain, our sensory cortex. So what was a local bodily reaction for primitive organisms, as for instance an amoeba, became a cerebral reaction with the same result for humans. According to Humphrey, consciousness can only come in to place as *sensations* or *reminders of sensations*. All mental representations have to be or refer to some kind of *sensation*, they have to have a sensory component. This is a short introduction to the ideas of Humphrey that allows us to draw a very important conclusion. Namely, that our body does matter in the kind of person we are. Following Humphrey, I believe we can say that: 'to have a different body is to be a different person with a different life' (1993: 115).

The individuality of our experiences and our mental lives depends in part on the kind of body we have. To believe that the body doesn't matter at all for the kind of person we are, is to be ignorant of these facts. Any justified thought experiment on 'brain swaps' should take them into account. We don't know what the result would be of such an operation, but what we do know is that it probably wouldn't leave the personal identity of the patient intact. A lot of discussion could be avoided, because the whole enterprise of trying to find out which criterion holds in the 'brain swap' case is overruled by our scientific knowledge about it. You can't isolate one of them and still be talking about actual persons with ordinary lives. Isolating a person's brain from a person's body is something you just can't do without putting the existence of that person on a hold. A lot of factors suggest that a *person* wouldn't survive his brain being transplanted in a totally different body. Should you of course transplant the brains of two identical twins, this becomes a different story. This is what Williams and Parfit also have in mind when they

require the bodies to be sufficiently alike.[13]

5.3 Teletransportation: Irrelevant Impossibility

Derek Parfit (Reasons and Persons, 1987) describes the following thought experiment. Suppose that I, inhabitant of the earth, want to travel to Mars. I've already been to Mars before, but this time something changed. Normally I have to take a space ship to get there, which takes me a few weeks. At present however, it's possible to enter a Teletransportator that takes me there in what appears to me as a few seconds, a blink of an eye. In reality I'll loose consciousness for one hour, while a scanning device destroys my brain and body on earth and in the meantime records the exact states of all of my cells. That information will be sent at the speed of light to a Replictor on Mars. The Replicator will create my exact brain and body out of new matter. I wake up on Mars in this newly created body. Will the person on Mars be *me*? Most philosophers will be inclined to say yes. However we can change the story a little bit, leaving us with doubts as to who wakes up on Mars. What if the Teletransportator malfunctions, sending information about the exact state of my cells to Mars, while leaving my body on earth intact. After the Replicator on Mars did its job, I can talk through the Intercom with my Replica on Mars. Here it seems that my Replica isn't *me*, but somebody else who's exactly like *me*. What if the Scanner not only left my brain and body intact, but above all damaged my heart? As a result, I will die within a few hours. I can talk to my Replica on Mars about this. Suppose he consoles me and promises me he will take up my life on earth. He will love my wife and children, finish the book I was writing and so forth. According to Parfit this consoles me a little:

> Dying when I know that I shall have a Replica is not quite as bad as, simply, dying. (1987: 201)

I believe Wilkes's view on the importance of 'the impossibility of teletransportation' to be mistaken. According to me this impossibility

---

[13] Unfortunately, in his definition of the necessary and sufficient conditions of personal identity through time, Parfit doesn't mention this demand. By doing so, he puts his own argumentation at risk.

doesn't bring any problems for the thought experiment considered. What Parfit wants to show is that, supposing it would be possible, the person on Mars would be *me* or just like *me*. It would be as good as ordinary survival because I am qualitatively identical to the person on Mars. So if I'm destroyed at a certain point in time and after a few minutes created again in every way qualitatively the same as before, I can believe this newly created person to be *me* or at least exactly like *me*. This point remains, even if Teletransportation isn't and never will be possible. So the impossibility of Teletransportation doesn't affect the outcome of the thought experiment. It's not relevant for our conclusion.

6. Conclusion

At present, our scientific knowledge is too limited to construct 'personal identity through time' as a natural kind term. But I strongly believe this will be possible in the future. Wilkes's conclusion that thought experiments on personal identity through time should be dismissed because of there lack of scientific background theories doesn't add up. If those thought experiments are in agreement with the existing scientific facts, we can take them to be at least nomologically possible. This is exactly what Wilkes's demands amount to. The examples mentioned above show us the possibility of confronting those experiments with Wilkes's demands, giving rise to a more progressive and realistic way of dealing with philosophical theorizing on personal identity through time. Neurological research already pointed out a certain system or function in the frontal lobes that's essential for the 'personalization' of our memories and thus for people being able to have a personal identity through time. Beside this, we do have a lot of neurological findings that at this very moment can be helpful tools in assessing the relevance and nomological (im)possibility of certain thought experiments on personal identity through time. They can guide us in deciding which conclusions can be rightfully drawn. Our intuitions about personal identity through time can be 'agreed intuitions', based on backing scientific knowledge. Of course, there's still a lot of uncertainty and up to now, science didn't provide us with a sound answer to all questions arising from thought experiments on personal identity through time. At present, often all we can do is acknowledge which path is the better one to follow. It's one step forward towards a

more justified use of philosophical thought experiments. However, there's still a long way to go.

Universiteit Gent

REFERENCES

Cosmides L. & Tooby J. (2000), 'Introduction', in M. Gazzaniga (ed.-in-chief) *The New Cognitive Neurosciences*. London, Cambridge: MIT Press.

Cosmides L., Tooby J. (2000), 'Toward Mapping the Evolved Functional Organization of Mind and Brain', in M. Gazzaniga (ed.-in-chief) *The New Cognitive Neurosciences*. London, Cambridge: MIT Press.

Craik F.I.M., Moroz T.M., Moscovitch M., Stuss D.T., Winocur G., Tulving E., Kapur S. (1999), 'In Search Of the Self: A Positron Emission Tomography Study', *Psychological Science* **10**, pp. 26-34.

Dancy J. (ed.) (1997), *Reading Parfit*. Oxford, Malden: Blackwell Publishers.

Gazzaniga M.S. (1998), *The Mind's Past*. Berkeley, Los Angeles & London: University of California Press.

Häggqvist S. (1996), *Thought Experiments in Philosophy*. Stockholm: Almqvist & Wiksell International.

Halligan P.W., Oakley D.A. (2000), 'What do you mean when you talk about "yourself"? Leading Neuroscientists Peter Halligan and David Oakley are rewriting the rules on consciousness.', *New Scientists* **18**, pp. 35-39.

Humphrey N. (1993), *A History of the Mind*. London, Sydney, Glenfield & Bergvlei: Vintage.

Kandel E.R., Scwhartz J.H., Jessel T.M. (eds.) (2000). *Principles of Neural Science*. New York: McGraw-Hill.

Klein S.B., Loftus J., Kihlstrom J.F. (1996), 'Self-Knowledge of an Amnesic Patient: Toward a Neuropsychology of Personality and Social Psychology.', *Journal of Experimental Psychology: General* **125**, pp. 250-60.

Klein S.B. (2001), 'A Self to Remember: A Cognitive Neuropsychological Perspective on How Self Creates Memory and Memory Creates Self', in C. Sedikides & M.B. Brewer (eds.) *Individual Self, Relational Self, Collective Self*. Philadelphia: Psychology Press, Taylor and Francis Group.

Klein S.B., Rozendal K., Cosmides L. (2002), 'A social-cognitive neuroscience analysis of the self', *Social Cognition* **20**, pp. 105-35.

Llinás R.R. (2001), *I of the Vortex. From Neurons to Self*. Cambridge, London: MIT Press.

Miller B.L., Seeley W.W., Mychack P., Rosen H.J., Mena I., Boone K.

(2001), 'Neuroanatomy of the self. Evidence from patients with frontotemporal dementia.', *Neurology* **57**, pp. 817-21.

Parfit D. (1987), *Reasons and Persons*. Oxford: Clarendon Press.

Sacks O. (1995), *An Anthropologist on Mars. Seven Paradoxical Tales*. Great Britain: Picador.

Shoemaker S. (1963), *Self-Knowledge and Self-Identity*. Ithaca, London: Cornell University Press.

Shoemaker S., Swinburne R. (1984). *Personal Identity*. Oxford: Basil Blackwell.

Tulving E. (2002). 'Episodic Memory: From Mind to Brain'; *Annu. Rev. Psychol.* **53**, pp. 1-25.

Wheeler M.A., Stuss D.T., Tulving E. (1997), 'Toward a Theory of Episodic Memory: The Frontal Lobes and Autonoetic Consciousness', *Psychological Bulletin* **121**, pp. 331-54.

Wilkes K.V. (1988), *Real People. Personal Identity without Thought Experiments*. Oxford: Oxford University Press.

Williams B. (1970). 'The Self and the Future', *The Psychological Review* **79**, pp. 161-80. Reprinted in Noonan H (ed.) 1993. *Personal Identity*. Aldershot, Brookfield USA, Hong Kong, Singapore, Sydney: Darthmouth.