Final version published in George Bealer and Robert Koons (eds.), *The Waning of Materialism,* Oxford University Press, 2010

An Argument from Transtemporal Identity for Subject Body Dualism

Martine Nida-Rümelin[1]

1. Subject Body Dualism

I this paper I argue for a version of dualism that is stronger than property dualism and that may be taken to be a version of so-called substance dualism. But the latter term invites associations that I would like to avoid. Subject body dualism as I will use the term includes the claim that there is an individual that has experiences, thinks and is active and is *neither identical* to any material thing *nor constituted* by any material thing. The view is thus incompatible with the claim that - just like the statue is constituted at any moment of its existence by the piece of bronze that makes it up without being identical with that piece of bronze - the subject is constituted by its body.[2] According to subject body dualism the subject is an individual that exists wholly at any given moment of its existence and persists across time while changing its properties. According to subject body dualism subjects perdure and they do not endure (subjects exist across time but they are not temporally extended).[3] Subject body dualism does not imply that conscious subjects can exist without having a body. If the term 'substance' is reserved to entities that do not depend for their existence on the existence of any other entity, then the view proposed cannot be classified as a version of substance dualism.

Substance dualism is often associated with the view that the person is metaphysically composed of two parts: a body and a soul. According to subject-body dualism the person is not composed. The person has a body and her body is in no sense a part of her. There is no difference to be made, according to subject body dualism between the person and 'her self'. When a person uses the first person pronoun, or is addressed by the second person pronoun or is referred to by her name, then the reference in each of these cases is the same, it is the person (the subject of experience). Subject-body dualism rejects many of those ideas that are traditionally associated with the concept of a soul. Souls are often described as composed of some non-material stuff. Subject

1 I would like to thank Max Drömmer, Gianfranco Soldati, Dominic O'Meara, Fabian Dorsch, Gian-Andri Toendury and Jiri Benovsky for discussions about the topic that helped me a lot to further develop the view here presented. And I would like to thank the participants in my German advanced seminar in the spring semester 2008 at the philosophy department in Fribourg and the participants in my French seminar in the same semester for questions and critical remarks that helped me a lot to see further points to be developed and to be clarified.

2 For a version of the constitution view see Lynne Rudder Baker, 2000.

3 The distinction between enduring and perduring objects has been introduced in these terms by David Lewis, 1986 and Marc Johnston, 1987.

body dualism rejects the idea of a non-material stuff. Souls are often described as the immaterial part of a person. The subject body dualist does not believe in the existence of immaterial parts of people. The experiencing subject is not a part of the person, it is the person itself. Persons are subjects, the term 'subject of experience' is only the more general term covering human and non-human conscious beings. Souls are sometimes thought of as literally leaving the body. The capacity to leave the body presupposes spatial location. According to subject body dualism subjects are located only in a derivative sense: they are where their bodies are.

Substance dualism is often understood as limited to the human domain. All motivations for subject body dualism with respect to a given being are based on the fact that the being considered is endowed with consciousness. The complexity and the sophistication of the being's conscious life does not play any role in the motivation for subject body dualism. It follows that subject body dualism cannot be restricted to the human domain. If it is correct for humans it must be correct for the most simple organism in which consciousness in the broadest sense arises. Consciousness in the broadest sense is present in a simple creature capable of enjoying warmth or feeling pain.

If subject body dualism applies for instance to an elephant then somebody might make the following objection. According to subject body dualism when I see the huge impressive organism making up the body of an elephant I do not see an elephant since the elephant is not its body. According to subject body dualism it is impossible to see elephants just like it is impossible to see my human friends. I can never see them, I can only see their bodies. This is clearly a reductio ad absurdum of subject body dualism. The subject body dualist must answer this objection by pointing out that it is based on a mistaken view about what it is to see something. The subject body dualist will insist that we see the subject by seeing its body or parts of its body.[4] .

Substance body dualism does not imply the eternal existence of the subject. Subject body dualism is compatible with the metaphysical hypothesis that it is impossible for a subject to exist without having a body. Subject body dualism can but need not include the claim that it is metaphysically possible to change one's body.

The dualist view here proposed has no religious motivation. It is well compatible with the idea that subjects of experience are products of nature and come into existence without any intervention of any supernatural being. Subject body dualism is compatible with the plausible assumption that naturally evolved subjects of experience on this and other planets are alone in the

---

4 The objection deserves an elaborated answer that will not be developed in the present paper. The idea that we cannot see the subject according to subject body dualism is in part based on the mistaken presupposition that the subject is somehow hidden within its body and in part on the mistaken idea that the subject cannot be in causal contact with its environment. A positive account of seeing that would imply that we see the subject has to develop an account of seeing-as. When we see a smile (a particular movement in a face), we see it as the smile of someone. The smiling person is directly present in the content of our perceptual experience. Also, the subject is active in its smiling. So we are in causal contact with the subject itself.

universe: no supernatural being causes or knows of or cares about their existence.

2. Motivation for Subject Body Dualism

There are, in my view, three main interdependent motivations for subject body dualism. The first is based on phenomenal consciousness, the second on the phenomenon of being active and the third on identity across time.

Here is a brief sketch of how reflecting about phenomenal consciousness may lead to subject body dualism: a first step consists in the insight that occurrences of phenomenal experience require the existence of a subject who has the experience. One will thereby realize that what is amazing about the emergence of consciousness is not the instantiation of a new kind of properties (phenomenal properties) but rather the beginning of existence of an experiencing subject. One may thereby come to see that no explanation of what it is to be a subject in terms of having experiences is available since any such explanation would be circular. This result can be used in an argument for the claim that no satisfying account of phenomenal consciousness can be given without accepting a specific ontological category for subjects of experience.[5]

A second motivation of subject body dualism starts from the phenomenological insight that we experience ourselves as active in our doings and that we perceive other conscious individuals as active in their activities. Arguably, these experiences cannot be veridical unless the active subject is itself a cause of certain physical events. In a third step one may see that this kind of causation (the subject is itself a cause) could not possibly exist between a material thing and events occurring within its spatial boundaries. If each of these steps is correct, then subject body dualism is phenomenologically supported: to deny subject body dualism would imply that we are constantly the victim of a fundamental illusion in the way we experience our own doings and in the way we perceive the activities of other conscious individuals.[6]

A third motivation of subject body dualism is based on intuitions concerning the identity across time of conscious individuals. One way to gain the relevant intuitive insight is by reflecting upon cases of so-called duplication where we know of a conscious individual A that it will 'split into' two successors B and C who will stand in symmetrical empirical relation to A. Considering these cases it seems obvious that despite the lack of any difference in the relevant empirical relation obtaining between A and B on the one hand and between A and C on the other, we have a clear positive understanding of the difference between a future course of events that makes the

5 This first motivation for subject body dualism is partially developed in Martine Nida-Rümelin, 2008.

6 The argument is based on the idea of subject causation which is similar to Chisholm's thesis of agent or immanent causation (see Roderick M. Chisholm, 1976 and for a new elaborated version of agent causation, see Tim O'Connor, 2000). The phenomenological observation alluded to is closely related to the claims about phenomenology presented in Terence Horgan et al. 2003. For a development of the argument sketched see Martine Nida-Rümelin, 2007.

assumption of identity between A and B true and a future course of events that makes the assumption of identity between A and C true. The argument given in the present paper will be based on this observation. It will be argued in a first step that the clear and positive understanding we have or seem to have of the apparent factual difference between these two possibilities is due to deep conceptual structures present in any thinker who is capable of I-thoughts about the past and the future and who has the cognitive capacities to conceive of others as subjects of experience. On that basis it is argued in a second step that any philosophical theory that denies the factual difference between the two possibilities attributes unavoidable and fundamental illusion to every thinker capable of transtemporal I-thoughts and of conceiving of another being as conscious. This is reason to reject any theory that denies the factual difference. In a third step it will be shown that only subject body dualism can plausibly fulfil that constraint.[7]

3. Grasping the difference between two possible identity facts in a case of duplication

Let us consider a case where a person, Andrea, will be operated and thereby 'duplicated' tomorrow: Her brain will be divided into two halves, each of them will be transplanted into a different body. Let us call the woman waking up after the operation with the left hemisphere L-Andrea and the woman waking up after the operation with the right hemisphere R-Andrea. The human organism containing Andrea's left hemisphere will be called the L-body and the other organism will be called the R-body. L-Andrea and R-Andrea, let us suppose, will stand in normal psychological continuity with Andrea, - each of them feels close to her friends, has Andrea's attitudes towards others, has the same nice humour and the same ideas about what constitutes a good life, etc. Each of them will initially be convinced of being Andrea.

Let us suppose that *there is no relevant empirical difference* in the relation between the original person and her two successors in this sense: there is no difference in these relations that could be responsible for the fact that one of the two successors but not the other is identical to the original person. For the sake of argument let us suppose that the empirical relations of psychological and bodily continuity obtaining directly after the operation between Andrea and L-Andrea on the one hand and Andrea and R-Andrea on the other are perfectly symmetrical. Any such relation either obtains between both successors and the original person or between none of the successors and Andrea.[8]

Let D be a highly complex sentence that describes all the details about the way Andrea

7 The argument is developed in detail in the book Martine Nida-Rümelin, 2006.

8 For discussions of the brain division example see John Perry, 1972, David Wiggins, 1967, p. 50, Derek Parfit, pp. 245, P.F. Snowdon, 1991, F. Doepke, 1996, D. Hershenow, 2004 and Swinburne, 2006.

divides into two successors. D describes all the details about the way her brain splits and is connected with the two new bodies and all the relevant details about the psychological relation between Andrea and her two successors. From the point of view before the operation we may say that D characterizes the future course of events. According to what I take to be our natural understanding, the possibility described by D can be subdivided in at least three possibilities by adding assumptions about Andrea's identity:

P1: D and Andrea is L-Andrea.
P2: D and Andrea is R-Andrea.
P3: D and Andrea is none of the two.

It is sometimes said in discussions about 'duplication cases' that the original person might be identical with *both* successors. There is a way to make sense of this idea: Andrea might be L-Andrea *and* R-Andrea by having both successor bodies. Since this supposed possibility does not play any role in the argument here presented, I will simply put it aside. In what follows I will focus on the difference between P1 and P2.

In the first step of my argument I would like to convince the reader of what I take to be an insight about our cognitive architecture: we have or seem to have a clear positive understanding of the factual difference (or an apparent factual difference) between P1 and P2. If the future is such that P1 will be rendered true, then Andrea will wake up with the L-body, she will see the world from the L-bodies perspective, *she* will be the one who suffers if the L-body is damaged. But if P2 correctly describes what will happen, then Andrea will have quite different visual experiences when waking up (the ones' connected with the R-body), she will act with the R-body, she will live the life of the person who has the R-body.[9]

In the present first step of the argument my claim is merely conceptual. When we seem to grasp the difference between P1 and P2 we might be under an illusion. It might be that there is no objective difference corresponding to the two different descriptions. But this is not how the situation presents itself to us when we reflect about it. We seem to have a clear conception of an at least apparent factual difference between P1 and P2. The difference appears to be factual in this sense: "D and Andrea is L-Andrea" and "D and Andrea is R-Andrea" are not just two legitimate description of one and the same course of events. Rather, there is - according to the way we conceive of the situation - an objective possible feature of the world that makes one of the two descriptions true and the other false.

9 I will use the letters "P1" and "P2" in a systematically ambiguous way, sometimes to refer to the description of the possibility and sometime to refer to the posiibility itself. It will be clear from the context which of two is meant.

The factual difference may be described pointing out that Andrea will have a different future depending on which of the two possible identity facts will obtain. This is a difference *for* Andrea, in some sense, but it is not only a difference for her; it is not a *merely* subjective difference. It is - or so it seems - an objective feature of the world that Andrea has - at the later moment considered - such and such properties and such and such experiences. We conceive of the difference between P1 and P2 by realizing that Andrea's future is different depending on which of the two possibilities will be realized. If we knew that L-Andrea will have a wonderful life and that R-Andrea will have a horrible life, we can refer to Andrea at her presence and say: if P1 is realized, then *she* will live L-Andrea's happy life and if P2 is realized then *she* will live R-Andrea's horrible life. I take it to be an undeniable fact that we at least seem to understand the difference thus described and that we cannot but conceive of that difference as of a real, factual and quite substantial difference.

It has often been pointed out that the intuition that there is factual difference between P1 and P2 gets much clearer when one imagines oneself being in Andrea's situation. Suppose you wish to know whether you will be the person with the L-body who will live a happy life or the person with the R-body who will live a horrible life. A philosopher might tell you that the answer to your question is under-determined: the real course of future events does not make one of the possible answers true and the other false. It has often been pointed out that the person concerned will not be satisfied by that reply. But it is not easy to spell out what that dissatisfaction consists in. One might think that the person concerned is unsatisfied since she does not know what emotional attitudes she should take and she might find it uncomfortable to oscillate between happy expectation and fear. But the dissatisfaction is not just an emotional one. The roots of that dissatisfaction lie deeper. There is a sense in which the person concerned cannot make sense of the idea that her future is under-determined. There must be - or so it seems - an answer to the question of whether I will lead this life or rather that life; there must be an answer to the question which of two future bodies will be mine. Our conceptual make-up is such that each of us has (or seems to have) a clear understanding of the difference between a world where his or her own I-thought "I will be happy after the operation" will be rendered true on the one hand and a world where "I will be unhappy after the operation" is rendered true on the other. Under the presupposition that there is no possibility that I will be between being happy and being unhappy the world must be such that one of the two I-thoughts is true and the other false, or so it seems. Given our conceptual make-up a person who states the underdetermination of the issue seems to be talking nonsense. Our dissatisfaction with her claim is cognitive and not or not just emotional.

In what follows I will propose an explanation of why it may help to consider the case from the first person perspective. However, the insight that taking the first person perspective in one's reflection about the case makes the relevant intuition more salient can invite a number of

misunderstandings. Imagining oneself being in Andrea's situation helps to see that there seems to be factual difference between P1 and P2 and that we seem to be able to grasp that difference. This may invite the idea that empathy is relevant here. One might think that a person who empathizes with Andrea will be able to understand her tendency to insist that there must be an answer to the question of which body will be hers. And one might thus be lead to the hypothesis that we seem to see a difference between P1 and P2 as a result of emotional confusion and that emotion-free reflection about the case will reveal that this as an illusion. But this would be to miss the point. We can realize that there clearly seems to be a substantial factual difference between P1 and P2 even when we think about Andrea's case in a cool and emotionally detached way. We seem to grasp the difference even when we do not care about which of the two possibilities will be realized. To see the difference is not the result of an emotionally coloured conception that might be misled for that reason.

Another misunderstanding may occur when we take the exercise to imagine being in Andrea's situation as a case of taking the perspective of another person with the intention to trigger those cognitive processes that the other person is likely to undergo. One may thus be led to the conclusion that P1 and P2 only seem different when considered from the perspective of the person concerned while there is no such impression when we think about the case from the third person perspective. But this is a mistake. We seem to grasp the difference between P1 and P2 when imagining being the person concerned but we also seem to grasp the difference when thinking about the case as a story concerning somebody else. It is not a case where changing perspectives changes the way things appear. There seems to be a substantial factual difference between P1 and P2 from the first person perspective as well as from the third person perspective. Still, being oneself the victim of future duplication would make it particularly difficult to deny that the difference between P1 and P2 is a factual difference. This psychological fact has a number of convergent explanations. One of these makes reference to the fact that I-thoughts play an important role in our grasp of the difference even when we consider the case 'from the outside'. When thinking about Andrea's future we take Andrea's perspective in the following sense: *we use the conceptual resources of self-attribution in considering the question of her identity across time*. This idea will be elaborated in the following section.[10]

4. Transtemporal self-attribution and transtemporal self-identification

10 The idea that we do understand the difference between the two possibilities is present in many discussions of personal identity and it is referred to by those who defend a non-reductionist view about transtemporal personal identity. Instances of this may be found in Chisholm, 1970, Madell, 1981, Williams, 1970 and Swinburne, 2006 and 2007. Parfit, 1984 repeatedly stresses the intuitive appeal of the claim that there is a such a factual difference but he argues that the intuition must be rejected.

Suppose you are in Andrea's situation and your brain will be divided and transplanted tomorrow. You wonder if you will ever wake up again after the operation. You wonder if you might wake up with the body of the L-person or with the body of the R-person. You have no difficulty to understand what would have to be the case if you wake up with the body of the L-person. In that case, when looking in the mirror after the operation you will see the L-person's face. You will see what is visible from the L-person's perspective. In that case, when you talk after the operation then you will be moving the L-bodies lips. You understand what would have to be the case for you to be the L-person by understanding the consequences of this assumption for your future. By understanding that being the L-person involves for you having all her future properties you gain a clear grasp (or so it seems) of how the world would have to be for the hypothesis "I will be the L-person" to be true.

If this description is correct then we may say the following: You understand the assumption "I will be the L-person" on the basis of understanding thoughts like "I will have property P." In other words and more precisely: you understand what has to be the case for your utterance "I will be the L-person" to be true on the basis of your understanding of what would render your self-attribution "I will have property P in the future moment m" true. We can formulate this claim in a more abstract way: Transtemporal self-attributions (thoughts that can be expressed by sentences like "I will have property P" or "I had property P") are conceptually prior to self-identifications (thoughts that can be expressed by sentences of the form "I will be P at moment m" or "I was P at moment m"). Any person when thinking about her own future has an understanding of what makes thoughts of the second kind true on the basis of her understanding of what makes thoughts of the first kind true and only on that basis (and not vice versa).

A second point about transtemporal self-attribution is relevant in the present context. Your understanding of what has to be the case for your I-thought "I will have property P" to be true in no way depends on the empirical criteria of transtemporal identity of subjects of experience that you implicitly accept. It does not matter if you accept for instance the psychological theory or the bodily theory of transtemporal personal identity. To see this point one may consider Shoemaker's example of the brain state transfer procedure (BST procedure).[11] You live in a society where people regularly get a new body and a new brain cloned from their own genes. They undergo a procedure where the old body is incinerated and the whole brain state of the original brain is transferred to the new brain. Suppose you believe in the psychological theory of personal identity and that you therefore expect that the procedure preserves identity. In that case you believe that for every property P at moment m of the person who will leave the BST machine your corresponding I-thought "I will have property P

11 Shoemaker, 1984, p. 109

at m" is true. Suppose now that one day you change your mind. A brilliant philosopher in debates of several nights has convinced you: the psychological theory of identity cannot be true. The only acceptable theory of personal identity says that your identity across time depends on bodily continuity defined in a sense which is not preserved in the BST procedure. You now think that your existence will come to an end once the BST machine will have destroyed your original brain. You think that your conscious experience will stop for ever during that procedure. You now believe that you will not have the properties of the person who will leave the BST machine. If P is an arbitrary property of which you know that it will be instantiated by that person at moment m, then before your change of mind about the criteria of transtemporal identity you were convinced that your thought "I will have property P at m" is true. Now you are convinced that "I will have property P at m" is false. You have changed your mind about what will be the case. But you have *not* thereby changed your understanding of the content of your own I-thought. Your conceptual grasp of what has to be the case for your I-thought to be true has not changed at all. What has changed is your opinion about the empirical criteria that are necessary and sufficient for your I-thought to be true. - If this is true then we may say the following: our understanding of what makes our own I-thought about future properties (transtemporal self-attributions) true is invariant with respect to changes of the empirical criteria of transtemporal personal identity that we implicitly or explicitly accept. Our conceptual grasp of what has to be the case for the self-attribution "I will smell the odor of basil at m" to be true does not depend on the criteria of transtemporal personal identity that we accept. Empirical criteria of identity across time do not enter the conceptual content of transtemporal self-attribution. They do not enter the way we conceive of what must be the case for a given transtemporal self-attribution to be true.

To see this clearly suppose the opposite were true. Suppose that the criteria of transtemporal identity that we accept did enter our conceptual grasp of what would have to be the case for the relevant I-thoughts to be true. It is quite obvious in what way these criteria would have to enter the conceptual content of the relevant I-thoughts. When thinking about your future properties before your change of mind in the example above, your thought "I will have property P at moment m" would be a thought with roughly the same conceptual content as the thought you might express saying "There will be a person at m standing in the relevant psychological relation to me who will have property P". According to this proposal, to think that you will have P is to think that someone standing in a certain empirical transtemporal relation to you will have P. If this was the correct analysis of what you think then after your change of mind about the empirical criteria of transtemporal identity you would think a different thought (a thought with a different conceptual content) when thinking "I will have property P". What you then would think after your change of mind in having a thought expressible in these words would be roughly given be the different

sentence "There will be a person at m standing in bodily continuity to me who will have property P". We can see the inadequacy of theses glosses of the conceptual content of your I-thoughts in the following way. When you change your mind about the correct criteria of personal identity then you also change your mind about your own future. For any given property P that you believe will be instantiated by the person who leaves BST machine you believe your thought "I will have property P at m" to be true before your change of mind and you believe your thought "I will have property P at m" to be false after your change of mind. You thereby change your mind about your actual future. There is some possible future fact that you believe to be true before your change of mind and that you believe to be false after your change of mind, - or at least this is the way things appear to be according to your understanding of your own I-thoughts. You expect to wake up after the procedure before your change of mind and you do not expect to ever wake up again after your change of mind. You have thereby changed your mind about a specific feature - or so it seems - of the real course of future events. - All this would however be wrong if the empirical criteria for transtemporal personal identity accepted by a person did enter her conceptual grasp of what has to be the case for her transtemporal self attributions to be true. We then would have to say that the thought you expressed saying "I will have property P at m" before your change of mind is not the same thought that you express saying "I will have property P at m" after your change of mind. If this was right then there really would be no conflict between the belief you expressed using these words before your change of mind and the apparently contrary belief you express uttering the negation "I will not have property P at m" after your change of mind. It would be a mistake to think that you changed your mind about your own future. There is, according to that proposal, no fact about the future that you first believe to obtain and then believe not to obtain. If the proposal were correct then in thinking those thoughts it should not even seem to you as if there was such a possible fact that you changed you mind about. The reason is, quite simply, that according to the proposal to reject, in using the same words "I will have property P at m" you really express quite different thoughts. Before your change of mind you express the thought that there will be someone at m standing in a certain psychological relation to you who has P at m and you still believe this to be true after your change of mind. After your change of mind in using the words "I will have property P at m" you express a quite different thought namely that some person existing at m and standing in a certain bodily relation to you will have property P and you still believe *this* to be wrong. So there would be no fact or apparent fact about you that is true according to what you believe before your change of mind and that is false according to what you believe after your change of mind. I take this to be a clear reductio ad absurdum of the hypothesis that the empirical criteria of transtemporal identity for people accepted by a thinker enter the conceptual content of his or her I-thoughts about the future.

We thus have arrived at two related claims about the conceptual status of transtemporal self-

attribution: transtemporal self-attribution is conceptually prior to transtemporal self-identification and transtemporal self-attribution is invariant with respect to changes of accepted empirical criteria of transtemporal identity of people. This makes it easy to see why we grasp or seem to grasp a factual difference between P1 and P2 when the case is considered from the first person perspective. Whenever a person in the situation of Andrea thinks about her own future, she has a clear positive understanding of what would have to be the case for her thought "I will be the L-person" to be true. This clear positive understanding is due to his or her clear positive understanding of what has to be the case for certain I-thoughts of the form "I will have property P" to be true. The latter I-thoughts do not in any way depend in their conceptual content on any empirical criteria of transtemporal identity. No empirical criteria of transtemporal identity enter the conceptual content of these thoughts. This is why we have no difficulty in grasping or at least apparently grasping the difference between the possibilities P1 and P2 when we consider the case from the first person perspective despite the fact that we assume that there is no relevant empirical difference between P1 and P2.

6. Thoughts about identity across time of things without consciousness

What has been said about conceptual priority and invariance with respect to changes in the acceptance of empirical criteria for first person thought is quite clearly false when transferred to thought about transtemporal identity of objects that we do not believe to be endowed with consciousness. This point is particularly clear if one considers the special case of socially constituted objects like swimming clubs or restaurants. Suppose a person, Giovanni, accepts as a necessary and sufficient condition for the identity of restaurants across time that the owner, the cook and the name of the restaurant remain the same. According to his understanding a restaurant may move from Naples to New York. Suppose Giovanni changes his mind later on and now accepts that the surroundings and the interior decoration are essential properties of restaurants. The restaurant "Lucia" in Naples is closed in April 2008. The owner and the cook move to Little Italy and open a restaurant named Lucia in May 2008. Before his change of mind with respect to the appropriate criteria of identity across time for restaurants Giovanni would have judged that the restaurant Lucia has opened its doors in New York in May 2008. After his change of mind Giovanni judges that the restaurant Lucia does not exist any more. Is there any fact of the matter with respect to which Giovanni has changed his mind? Do we have to assume that Giovanni will be under the impression of having changed his mind with respect to a factual issue? Obviously, the answer to both questions is negative. When Giovanni utters "The restaurant Lucia will open its doors in May 2008 in New York" before his change of mind and when he utters the same sentence after this change of mind,

then on these two occasions he expresses thoughts with quite different conceptual content. The conceptual content associated in Giovanni's thought with the sentence at issue changes with the modification of the criteria of identity for restaurants across time that he presupposes. There is no possible future fact with respect to which Giovanni has changed his opinion when he changed his mind about the appropriate criteria of identity for restaurants. Before his change of mind Giovanni's thought that the restaurant Lucia will open its doors in New York in May 2008 could be paraphrased saying that there will be a restaurant in New York named "Lucia" and run by the same owner with the same cook that will open its doors in New York in May 2008. Giovanni has not changed his opinion with respect to this detail about the future. After his change of mind the thought expressed by the same words must be paraphrased quite differently: there will be a restaurant named "Lucia" in the same location in Naples, with the same interior decoration, the same owner and same chef cook which will open its doors in New York. This thought is trivially false and rejected by Giovanni before and after his change of mind. – Reflection on this and further examples motivate the following claims. (1) In the case of thought about non-conscious individuals the conceptual content of transtemporal property attributions and of transtemporal identity statements changes with the accepted criteria of transtemporal identity. The latter can in principle be explained by reference to these criteria. (2) In the case of thought about non-conscious individuals transtemporal identification is conceptually prior to transtemporal property attribution. (1) and (2) explain why, when considering duplication cases for non-conscious individuals, we have no temptation to think that we can grasp a factual difference between the possibilities analogous to P1 and P2 when there is no relevant difference in the relations between the two successors and the original object. When a restaurant splits into two and both successors can be regarded with equal right as the original restaurant (the relevant empirical relations are symmetrical) then there is no apparently open factual question about the original restaurant's identity.

6. First person thought and other-directed thought

I have argued that transtemporal self-attributions and transtemporal self-identifications have a special conceptual status which can be summarized by the following claims:

Claim 1: Transtemporal self-attribution is conceptually prior to transtemporal self-identification.
and

Claim 2: Transtemporal self-attribution is conceptually invariant with respect to changes in the

thinker's accepted criteria of identity of people across time.

We may add a further claim that has not yet been explicitly mentioned. Since our conceptual grasp of our own transtemporal identity is based on our grasp of what has to be the case for certain transtemporal self-attributions to be true and since the latter are invariant with respect to the thinker's accepted criteria of transtemporal identity, self-identifications exhibit the same conceptual independence. We can thus add:

Claim 3: Transtemporal self-identification is conceptually invariant with respect to changes in the thinker's accepted criteria of identity of people across time.

I will now defend the view that these special traits of first person thought carry over to other-directed thought.

I argued above that we understand the difference between P1 and P2 by taking Andrea's perspective. Taking her perspective in the relevant sense is a conceptual exercise that is natural for us or even forced upon us in other-directed thought. We use our conceptual resources given by the specific conceptual status of first person thought in thinking about her. Andrea will have property F just in case her first person thought "I will have property F" will be made true by the future course of events. Given the specific conceptual status of first person thought each of us has a clear understanding (or seems to have a clear understanding) not only of how the world would have to be for one's own first person thought "I will have property F" to be true but also of how the world would have to be for *her* first person thought "I will have property F" to be true. In other-directed thoughts like "Andrea will have property F" we fix reference to a particular subject, Andrea in this case, and then consider how the future would have to be for *that subject's* first person thought "I will have property F" to be true. To make this point one might say a bit paradoxically: in thinking about future properties of others we use first person thought applied to them. (I use the term “first person thought” in the sense of thought in the first person mode.)

If it is correct that we apply the resources of first person thought in the sense just explained in other-directed thought, then of course the special conceptual status of first person thought carries over to other-directed thought. Other-directed thoughts like "A will have property P" and "A had property P" are invariant - just like the corresponding first person thoughts - with respect to changes in the accepted criteria of personal identity or subject identity. Accepted criteria of subject identity do not enter our understanding of what would have to be the case for these other-directed thoughts to be made true by the actual future course of events.

We can and we do use first person thought applied to others independently of whether these

others actually think or can think I-thoughts. It appears obvious that first person thought applied to others is appropriate with respect to any conscious being independently of wether that subject is sufficiently sophisticated to think I-thoughts. In the case of a baby or a non-linguistic animal we may fix reference to that particular subject and wonder if the future is such that *its* first person thought "I will have F" if the subject at issue had that thought would correspond to reality. We then use the conceptual resources of first person thought in application to a being that does not or cannot have the corresponding I-thought. We clearly are capable of thinking in that way and we clearly do think in that way all the time. Whenever we believe of a being that it is conscious we use the conceptual resources of first person thought applied to it in the way described. A still stronger thesis appears adequate: To conceive of another being as a subject of experience partially consists in thinking about its past and future in that specific way. A being that is incapable of applying the conceptual resources of first person thought to others does not have the full concept of a subject and cannot think of another being as a subject of experience.

We can now formulate the claims that state the transfer of the special status of first person thought about the past and the future to other-directed thought. Given that transtemporal self-attribution is invariant with respect to changes of accepted criteria of personal and subject identity, other-directed thought - via the application of first person thought to others - is also invariant with respect to changes of criteria of transtemporal personal and subject identity accepted by the thinker:

Claim 4: Transtemporal attribution of properties to other experiencing subjects is conceptually invariant with respect to changes in the thinker's accepted criteria of subject identity across time.

According to claim 4, if we take A to be conscious, then we conceive of what has to be tha case for the thought "A will have (or had) property P at m" to be true in a way that does not apply or presuppose any empirical criteria of transtemporal personal and subject identity. Since (a) we grasp what has to be the case for a conscious being to have the property F at a future moment m by grasping what has to be the case for its corresponding I-thought to be true and since (b) we grasp what has to be the case for another conscious being to be identical with a future subject on that basis, it follows that the conceptual priority of transtemporal property attribution carries over from first person thought to other-directed thought. For the same reason the invariance of the conceptual content of thoughts about one's own past and future properties carries over to thoughts about transtemporal identity of other conscious beings. The following claims summarize the result of this transfer:

Claim 5: Transtemporal attribution of properties to others is conceptually prior to transtemporal identification with respect to others.

Claim 6: The conceptual content of other-directed transtemporal identification is invariant with respect to possible changes of the accepted criteria of subject identity across time. Transtemporal criteria of subject identity do not enter the conceptual content of other-directed transtemporal identification.

The preceding claims explain why we are under the impression that we have a clear positive understanding of the difference between P1 and P2. The cognitive appearance of grasping a factual difference between P1 and P2 has been explained by features of our conceptual architecture. Often in philosophy the explanation of cognitive appearances is intended as an error theory: the appearance is explained in a way that excludes its veridicality. The present explanation is not intended as an error theory. Nothing about the explanation of the appearance undermines the appearance. The explanation given is perfectly compatible with the claim that the cognitive appearance at issue is veridical and that we do grasp a substantial factual difference when comparing P1 and P2.

7. The illusion theory and why it cannot be accepted

The following claim is hard to deny: it is appropriate to think about another individual's identity across time and about its past and future using the conceptual resources of I-thought just in case the other individual is conscious. It is precisely in that case that it makes sense to consider the individual's future from its perspective in the sense explained. To resist from thinking about another being in that particular way is to resists from conceiving it as a subject of experience. In other words: To think of another being as having its own perspective in the sense of being conscious is incompatible with thinking about its identity across time without asking the question of its past and its future using the conceptual resources of first person thought. Using the conceptual resources of first person thought is part of what it is to be aware in one's thinking of the fact that the other being too has its own 'point of view', that it is a subject of experience. If this is correct then we are justified to make the following general claim:

Claim 7: A thinker who conceives of another individual as conscious (as a subject of experience) necessarily uses the resources of first person thought in his conception of the other individual's

identity across time and in his conception of the other individual's past and future.

It is plausible to assume that the specific features of first person thought pointed out above are essential features of first person thought. When for instance a subject thinks "There will be someone at m in psychological continuity with me who will have property F" then the subject has a thought about itself but its thought is not a first person thought about its future. I propose to assume that fully developed self-conscious beings think about their future and about their past in terms of criterion-free I-thoughts ('criterion-free' in the specific sense given by the conjunction of claim 1, 2 and 3). We may summarize what has just been said in the following way: (a) To be a self-conscious thinker essentially involves thinking about one's own identity and about one's own past and future in a criterion-free manner. (b) To conceive of another individual as a subject of experience essentially involves using the conceptual resources of first person thought in one's thinking about the other's identity across time and in one's thinking about its past and future. It follows from these two claims together with the explanation developed in the preceding section that a self-conscious individual capable of conceiving of others as subjects of experience cannot free itself from the cognitive impression that there is a factual difference between P1 and P2. In other words: any philosophical theory that denies that there is a factual difference between P1 and P2 attributes unavoidable error to every self-conscious thinker capable of conceiving of another being as an experiencing subject. There is then a sense in which a self-conscious thinker capable of conceiving of another as a subject of experience cannot seriously believe a philosophical account that denies the apparent factual difference between P1 and P2: a thinker of that kind - independently of his or her theoretical convictions - will always be under the impression that P1 and P2 are substantially different possibilities. In order to bring ourselves to seriously believe that the apparent capability to grasp what the difference consists in is illusionary we would have to loose the capacity to think about our own past and future in the first person mode and we would have to loose the capacity to conceive of others as subjects of experience. There is reason to reject any theory that can only be seriously believed by beings that are conceptually impoverished in such a dramatical and undesirable manner.

To have a convenient term I will call the claim that there really is no factual difference between P1 and P2 the illusion theory. The illusion theory is appropriately so called since it says that the unavoidable cognitive appearance of grasping a factual difference between P1 and P2 is a cognitive illusion due to our conceptual architecture as self-conscious beings who can conceive of other beings as experiencing subjects. I just argued that there is good reason to reject the illusion theory since it attributes unavoidable cognitive illusion not just to the contingent human mind but to *any* self-conscious thinker capable of conceiving of someone else as conscious. At this point,

however, it might be objected that the error attributed by the illusion theory is rather limited nonetheless. It only occurs when we consider strange science fiction cases. But this objection misses the point.

If the illusion theory is correct then our daily thoughts, perceptions and emotional attitudes towards others are all infected by the illusion at issue. We have a criterion-free notion of identity across time of conscious individuals in the sense explained earlier. If the illusion theory is correct then - contrary to what clearly appears to be the case - we cannot use that notion in order to grasp a fact that makes identity statements true. If the illusion theory is correct then our whole thinking about the past and the future of subjects of experience and about their identity across time including our first person thought is based on an inadequate notion. The corresponding real facts about identity of conscious individuals (if there are any) are then constituted by relations that do not play any role in our understanding of what constitutes our continued existence. In all those cases in which we apply or presuppose that criterion-free notion we only seem to grasp a possible state of affairs.

Once we realize that the criterion-free notion of identity across time for conscious individuals is present not only in our thinking but also in our perceptions and emotions it becomes clear that the illusion at issue is still more general and deeper than one might think at first sight. When a person is touched by meeting a friend that she has not seen since many years then she perceives that friend *as identical* to the younger person she knew so well in the distant past. Perceiving the other person in that way which incorporates the criterion-free notion of identity is an essential component of that emotional experience. Following this line of thought it becomes clear that most of what we value in life would be based on a fundamental cognitive illusion if the illusion theorist was right.

The point that the illusion attributed to human thinking by the illusion theory is not restricted to isolated instances of theoretical reflection but rather concerns our whole cognitive and emotional life which would then be shot through with fundamental error can be made in a different way. It is an essential component of our concept of a subject that the identity of subjects can be grasped in the criterion-free way described. So if the illusion theory is correct then there really are no experiencing subjects in the sense of that notion which is deeply incorporated in our thinking and we are then constantly under a massive illusion when we conceive of ourselves as subjects of experience and when we conceive of the world around us as populated by subjects of experience. It seems clear that this is a fairly radical consequence of a philosophical theory. The evidence in favour of the illusion theory would have to be immense to make it acceptable despite its extreme, unbelievable and counter-intuitive consequences. Quite obviously, however, there is no such massive evidence in favour of the illusion theory. Philosophers who explicitly or implicitly endorse the illusion theory

are motivated by the suspicion that denying the illusion theory will lead quite directly into substance dualism. But the version of substance dualism supported by that denial is subject body dualism in the sense explained earlier. Subject body dualism is freed from many suspicious ideas that are usually associated with substance dualism. Furthermore, as briefly sketched at the beginning of the paper, subject body dualism has independent support by other considerations.

7. How a denial of the illusion theory leads to Subject Body Dualism

It remains to be shown that the acceptance of a factual difference between P1 and P2 has to be combined with subject body dualism. In a first step we can see that the denial of the illusion theory implies that the subject is not identical with its body. We may assume as a premiss that material bodies have empirical criteria of identity across time. According to this premiss, the fact that a given material object at m1 is identical to a given material object at m2 consists in the fact that certain empirical relations obtain between the material object at m1 and the material object at m2. Suppose a situation is realized in which the description D is satisfied. Suppose that Andrea is identical to her body or some part of her body. Then, according to the assumption that material bodies have empirical criteria of identity across time, there could be a factual difference between P1 and P2 only if there was a difference in the empirical relations between Andrea and her two successors. But we had assumed perfect symmetry in these relations between Andrea and her two successors. So, if Andrea was her body or some part of her body, then P1 and P2 could not be factually different.

This result is not sufficient to show that the denial of the illusion theory implies subject body dualism. The result is still compatible with a number of accounts quite different from subject body dualism. For instance, a functionalist analysis of personal identity is compatible with the claim that the person is not identical to her body or any part of her body since the empirical criteria of identity across time are different for material bodies and for people. Still the functionalist does not posit the existence of non-material individuals. But it is easy to see that the functionalist's analysis of personal identity is incompatible with the idea that there is a factual difference between P1 and P2. For the functionalist the identity of Andrea with L-Andrea consists in the instantiation of certain causal relations (L-Andrea has memories caused by Andrea's experiences, L-Andrea's actions causally depend on Andrea's earlier intentions etc.) So if functionalism was correct then there could be a factual difference between P1 and P2 only if there was a difference in the relevant causal relations between Andrea and her two successors. But we had assumed perfect symmetry in all empirical relations between Andrea and her two successors. So, if the functionalist account of personal identity was correct, then P1 and P2 could not be factually different.

A similar argument excludes the view that Andrea is constituted by some material thing just

like a statue is constituted by the piece of bronze making it up at any moment of its existence. We may assume that the proponent of the constitution view accepts the following principle. If B is the body of a person P at a given moment m and there are two human bodies B1 and B2 at a moment m', and if B1 but not B2 constitutes the person P at m', then B1 and B2 must be different with respect to their empirical relations to the body B that originally constituted person P. For Andrea to be L-Andrea it is necessary that the L-body constitutes Andrea after the operation. So according to the constitution view there could be a factual difference between P1 and P2 only if there was a difference in the relation between Andrea's body before the operation and the R-body on the one hand and Andrea's body before the operation and the L-body on the other. But we had assumed perfect symmetry in all empirical relations between Andrea and her two successors. It follows that the constitution view too implies the illusion theory.

Obviously, a parallel argument can be repeated for any view about transtemporal subject identity according to which there are some empirical facts that constitute a subject's persistence. The only plausible alternative to any such reductionist account seems to be the view that there is a subject, distinct from its body whose identity across time cannot be reduced to empirical relations but can be grasped by employing the resources of first person thought in the way described earlier.

8. Why four-dimensionalism does not help

The four-dimensionalist states that people are spatially *and* temporally extended. The four-dimensionalist can make sense of P1 and P2 in the following way. Andrea's use of "I" and our use of her name do not have a definite referent; they can both be interpreted as referring to (a) Andrea's temporal parts before the operation united with L-Andrea's temporal parts after the operation and (b) Andrea's temporal parts before the operation united with R-Andrea's temporal parts after the operation. The four-dimensionalist may propose a third interpretation: Andrea's temporal parts united with the temporal parts of both successors. For simplicity we may confine the discussion to a version of the theory that states an ambiguity of the reference of "I" in Andrea's thought and speech only between (a) and (b). On this view whether we should say that P1 or rather P2 is realized depends on how we interpret the relevant singular terms in their use before the operation. If we interpret both singular terms along the lines of (a), then we may state that P1 is realized. If we interpret them along the lines of (b), then we may state that P2 is realized. Obviously, according to that analysis, there is no conflict between the two descriptions of the real course of events. We may say that P1 is realized when we interpret the name occurring in the corresponding description in one way and we may say that P2 is realized when we interpret the name in another way. So the four-

dimensionalist view about the duplication problem is clearly a version of the illusion theory.[12]

Let me note in passing that the four-dimensionalist proposal is clearly inadequate if it is used to describe the cognitive content of I-thoughts and other-directed thoughts involved in reflections about Andrea's case. Any account of the conceptual content of these thoughts has to explain why P1 and P2 when considered from the first person perspective as well as when considered from the third person perspective clearly appear to be two substantially different possibilities. No such explanation can be given by reference to an ambiguity in the singular terms used before the operation. Suppose Andrea first believes that she will be L-Andrea and then believes that she will be R-Andrea. It clearly appears to her that she has changed her opinion about the future. When we use the four-dimensionalist account to describe the content of Andrea's thoughts, then we must say that there is no common content she has changed her mind about. What appears to her as a change of mind really would have to be an unnoticed change of the object of reference in her thought. But when considering the two possibilities, there is no change in her concept of herself corresponding to the shift of reference. The analogous observations apply to other-directed thought about Andrea's case. I conclude that the four-dimensionalist cannot explain the conceptual facts about I-thought and other-directed thought within his or her framework.

9. Concluding remark

The argument presented in this paper is not a conceivability argument. Conceivability arguments usually start with the claim that a certain scenario is conceivable without hidden contradictions. They then go on to argue that there is a real metaphysical possibility corresponding to the scenario that we can coherently conceive of. In the second step of conceivability arguments certain principles are evoked and applied which specify under what specific conditions the transition from conceivability to possibility is unproblematic.[13]

The first step in the present argument is not a claim about conceivability without hidden contradiction. The first step is an argument for the claim that we seem to grasp a specific factual difference between two possibilities. The second step in the present argument does not appeal to any general principles that allow in specific cases to proceed from conceivability to metaphysical possibility. Rather, the second step is based on the claim that any account that denies the veridicality of the cognitive impression of grasping the difference implies that our self-conception and our conception of others is deeply misguided and that any self-conscious being capable of

12 For a four-dimensionalist treatment of the duplication problem compare Lewis, 1976 and Lewis 1983.
13 Compare David Chalmers, 2002.

conceiving of another being as conscious is necessarily misguided in the same way. The argument proceeds by pointing out that any alternative to subject body dualism would force us to accept that the most valuable aspects of our life are built upon a deep, permanent and unavoidable cognitive illusion.

References

Baker, L.R., 2000, *Persons and Bodies. A Constitution View,* Cambridge University Press.

Chalmers, D., 2002, "Does conceivability imply Possibility?", in T. Gendler & John Hawthorne (eds.), Oxford: Oxford University Press, 2002.

Chisholm, R.M., 1970, "Identity through Time", In Howard E. Kiefer & Miiton K. Munitz (Hrg.) *Language, Belief and Metaphysics*, reprinted in Chisholm, 1989.

Chisholm, R.M., 1989, *On Metaphysics*, University of Minnesota Press, Minneapolis.

Chisholm, R. M., 1976, *Person and Object*. Dordrecht: D. Reidel.

Doepke, F., 1996, *The Kinds of Things: A Theory of Person Identity Based on Transcendental Argument*, Chicago: Open Court.

Hershenow, D., 2004, "Countering the Appeal of the Psychological Approach to Personal Identity", *Philosophy 79*, 447-475.

Horgan, T. with J. Tienson and G. Graham, 2003, "The Phenomenology of First-Person Agency.", nn S. Walter and H. D. Heckmann (eds.), *Physicalism and Mental Causation: The Metaphysics of Mind and Action*. Imprint Academic, 323-40.

Johnston, M., 1987, "Is there a problem about persistence?", *Proceedings of the Aristotelian Society* 61: 107-135.

Lewis, D., 1976, "Survival and Identity", in A. Rorty, *The Identity of Persons,* Berkeley: University of California Press, 17-40.

Lewis, D., 1983, "Postscript to 'Survival and Identity'", *Philosophical Papers,* Oxford: Oxford University Press.

Lewis, D., 1986, *On the Plurality of Worlds*, Oxford: Blackwell.

Madell, G. 1981, *The Identity of the Self,* Edinburgh: University Press.

Nida-Rümelin, M., 2006, *Der Blick von innen. Zur transtemporalen Identität bewusstseinsfähiger Wesen*, Suhrkamp Taschenbuch Wissenschaft.

Nida-Rümelin, M., 2007, "Doings and Subject Causation", in A. Newen, V. Hoffmann, M. Esfeld, "Mental Causation, Externalism, and Self-Knowledge", Special Issue of *Erkenntnis* 67(2), S. 147-372.

Nida-Rümelin, M., 2008, in preparation, Experiencing Subjects, in Antonella Corradini et al (eds.). *Emergence in Science and Philosophy*.

O'Connor, T. , 2000, *Persons and Causes: The Metaphysics of Free Will.* Oxford: Oxford University Press.

Shoemaker, S. (1984), Personal Identity: a Materialist's Account. In Sydney Shoemaker and Richard Swinburne (eds.), *Personal Identity,* Oxford: Basil Blackwell.

Snowdon, P.F., 1991, "Personal Identity and Brain Transplants", in D. Cockburn (ed.), *Human Beings*, Cambridge: Cambridge University Press.

Swinburne, R., 1984, "Personal Identity: the Dualist Theory", in: Shoemaker und Swinburne: *Personal Identity*, Oxford: Basil Blackwell

Swinburne, R. 2006, "Was macht mich zu mir?", in B. Niederbacher (ed.), *Die menschliche ,* Ontos Verlag, 2006.

Swinburne, R., 2007, *, in P. Inwagen and D. Zimmerman (eds.), *Persons,* Oxford University Press

Williams, B., 1970, "The Self and the Future", *Philosophical Review* 79: 161-180, reprinted in Williams, 1973.

Williams, B., 1973, *Problems of the Self*, Cambridge University Press.