

Why is personhood conceptually difficult?

The concept of a person is a vexing one.

There is ample evidence for this claim, both in time-honoured works and in recent publications. Before I concentrate on some of the old stuff, let me briefly turn to recent examples. The following sample of quotations from a Nobel Laureate, a leading neuroscientist and a German professor of ‘neuro-didactics’ may illustrate how deep the confusion about what a person is can go among the educated, even today. Francis Crick stated his *Astonishing Hypothesis* as follows:

“You” [...] are in fact no more than the behaviour of a vast assembly of nerve cells and their associated molecules. As Lewis Carroll’s Alice might have phrased it: “You’re nothing but a bunch of neurons.” This idea is so alien to the ideas of most people alive today that it can truly be called astonishing.¹

A few years later, this ‘idea’ seemed not anymore astonishing to Michael Gazzaniga who prefers to put it this way: Some simple facts make it

... *clear* that you are your brain. The neurons interconnecting in its vast network [...] -- that is you.²

It required the brilliancy of a German professor to take it to a further extreme. He found a way to expand Crick’s and Gazzaniga’s point by enriching it with a homespun piece of congenial ludicrousness. In a German radio-broadcast in November 2006, Manfred Spitzer declared:

You don’t *have* your brain, you *are* your brain.

Maybe this is a world record. Is it humanly possible to display, more fundamental confusion in less than ten syllables? (Well, in fairness to Spitzer, in German, the saying doesn’t take less than ten.) One is almost inclined, with respect to someone who says such a thing, to believe at least the first part of his *dictum*.

¹ Francis Crick, *The Astonishing Hypothesis – The Scientific Search for the Soul*, New York: Charles Scribner’s Sons, Macmillan Publishing Company 1994, p. 3.

² Michael Gazzaniga, *The Ethical Brain*, Chicago: Chicago UP 2000, p. 31 (italics mine).

Note that in these three quotations we are addressed directly, by use of the word “you”. As who or what might we consider ourselves so addressed (given that we are, in the same breath, straightforwardly identified with our brains)? Clearly not as human beings. Human beings aren’t just brains. Almost all of them have one.³ And some of them use it, before they make grand claims. Let’s assume that this much is known even to those who would make, or agree to, such claims as the ones I quoted. It’s unlikely that even they simply confuse a human being with one of his or her organs.

So assuming that we are not addressed, in the statements quoted above, as members of the species *homo sapiens sapiens*, the question remains: As whom or what do Crick, Gazzaniga and Spitzer presume to address us, when they say “you”? Well, I guess, we are meant to be addressed as *persons*. What the two neuro-scientists and the professor of ‘neuro-didactics’ want to tell us seems to be this:

You, the *person* you are, are your brain.

A human person nothing but his or her brain? The negative answer is obvious again. You, as person, are you altogether. When considered as a person, you are considered, so to say, as the completeness of what you are. You are not just an assemblage of certain parts, facets or aspects of yours, however interesting or prominent each of them may be. You’re *not* what or how you feel. You are *not* how you came to be what you are. You are *not* what you did or may accomplish. You are not your looks, moods, skills, genes, memories, sentimentalities, failures, hobbies, hopes or sexual obsessions. You are not your intelligence, deftness, body, body/mass index, charm, career, musicality, brain, character, hormonal state, social behaviour, or innermost thinking. All the items just mentioned, and indefinitely many more of those, contribute, or may contribute, to you as a person. But they aren’t you. Obviously, none of them, taken separately, is you. Arguably, even all of them together, taken collectively in their (impossible) summation, isn’t you either. - In brief, “You are your brain” is to be taken as seriously as “You are what you eat”. It may sound nice as an advertisement jingle, but taken literally, it’s just rubbish.

³ The pitiable exceptions include anencephalics, microcephalics, hydrocephalics, and some brainless adult human beings, occasionally mentioned in the literature, for whom there seems to be no scientific label yet.

I shall not go into this once more.⁴ Instead I shall address, in what follows, a different, an aetiological, kind of question: How can it happen that some people get so confused as to identify persons (and for that matter themselves) with their brains? Part of the explanation seems to me to be this: Our very idea, or concept, of a person is utterly baffling. And I shall investigate some of the reasons why this is so.

*

Given that the concept of a person is a vexing one, what is it that makes it so?

There are various ways in which a concept may perplex us. First, there are concepts which may strike one as inherently unthinkable – or, to put it less sloppily: concepts such that the items which they purport to be concepts of seem unthinkable. Infinity may serve as an example. (Ask a theologian or a philosopher, if you are keen on more examples of this sort.) Second, there are concepts which are, or seem, analysis-proof in a very peculiar way. They are, or seem to be, innocent, well-functioning non-primitive concepts which we, as normal speakers, have fully mastered; and, moreover, we are perfectly in the clear about what we consider as their most important ingredients. Nevertheless there is at least one further conceptual ingredient which consistently resists our attempts to make it explicit. Knowledge is an example. It is fairly uncontroversial that knowledge entails truth, belief and justification, and it is also clear that knowledge is not merely justified true belief – but nobody has been able to pinpoint what else is required for knowledge. The concept of knowledge contains at least one component, that vexing 'last bit', which seems inexplicable. Third, there are concepts which are, or at least seem to be, paradoxical, although they appear to be well-functioning, some of them even indispensable, concepts. Take the concept of being uninteresting. It lends itself to the comparative and the superlative form. But isn't the most uninteresting event of all times *ipso facto* an interesting one? I, for one, would be anxious to be informed about it. Or take the concept of a belief. One holds each of one's beliefs to be true (this is what believing is, after all), but at the same time, a sane person believes that some of his beliefs are false. Or take truth itself. The so-called Liar-paradox is known and unsolved since ancient times:

⁴ For arguments against the 'thesis' of person/brain-identity, cf. my "Ich, mein Gehirn und mein Geist – Echte Unterschiede oder falsche Begriffe?", in: N. Elsner/G. Lüer (eds.), *Das Gehirn und sein Geist*, Göttingen 2000, 221-243. – But let me warn you. You'll probably find nothing in this paper which you do not know anyway. Trying to point out the obviously obvious almost inevitably results in dull papers. What excuse is there for a philosopher to engage in this sort of business nevertheless? Well, as J. L. Austin once put it: "Besides, there is nothing so plain boring as the constant repetition of assertions that are not true, and sometimes not even faintly sensible" (*Sense and Sensibilia*, Oxford: Oxford UP 1962, p. 5).

"What I hereby say is not true". Or, for that matter, take any of those countless concepts for which a paradox of the *Sorites* type can be construed, like famously for the concept of a heap itself.

The conceptual difficulties concerning personhood seem to be of an altogether different kind. *Prima facie*, personhood is nothing inherently unthinkable; there's no problem with a deeply hidden conceptual 'last bit' (we'd be happy to get hold only of the uncontroversial first bits); and we have no compelling reason to think that the very concept itself is paradox-ridden.⁵

On the one hand, the word "person", as it is commonly used, seems to be not much more than a singular form of the word "people"; it serves to denote human beings like you and me. In normal conditions, as soon as we have recognized an adult human being, we have recognized a person; we don't need any extra information about special features of this particular human being in order to draw the 'further' conclusion that he or she is a person. In the absence of very weighty counter-evidence or of compelling reasons to withdraw judgment, the presumption, concerning any human being, that he or she is a person, is not just epistemically admissible or reasonable, it is morally obligatory.⁶ - The *application* of the concept of a person, in familiar standard cases, does not appear to involve problems which are harder than those involved in recognizing people: normal members of the human race.

But *the concept itself* is problematic. At least it is difficult to say, in plain words or, for that matter, more refined ones, what a person is – even given the most basic and austere sense of the word "person".

Person as an ontological category concept

⁵ One may think that the *person* clearly is a vague concept (*i.e.*, allows for borderline cases) and that therefore at least a paradox of the *Sorites* type can be construed. But I am not sure about it. The sad fact is, I think, that *person* is a concept so extremely indeterminate that we cannot even definitely say whether it is vague or not.

⁶ Note that a presumption is not just an assumption, however plausible. As Whately once put it magisterially: "According to the most correct use of the term, a 'Presumption' in favour of any supposition, means, not (as has been sometimes erroneously imagined) a preponderance of probability in its favour, but, such a *preoccupation* of the ground, as implies that it must stand good till some sufficient reason is adduced against it; in short, that the *Burden of proof* lies on the side of him who would dispute it." (Richard Whately, *Elements of Rhetoric*, 1828, ⁶1841, p. 120; Whately's italics). For an attempt at an outline of a theory of presumption, see Oliver Scholz, *Verstehen und Rationalität*, Frankfurt a.M. 1999, part II, pp. 148 – 159.

Two attempts at clarification. The first one concerns the question how much psychology comes with the concept of a person. Addressing this question seems necessary in the light of the best recent discussions concerning personhood I am aware of.⁷ When I talk in the following, interchangeably, of "the concept of a person", of "*person*", or of "(the concept of) personhood", I do not have a psychological concept in mind. *Person*, as I shall consider it, is an ontological concept. For it is meant, by me here, to pick out a special category of entities – a category which is worth considering when the question is raised: "What sorts of particulars are part of the ultimate furniture of the world as we know it?" As an answer I'd mention, with no attempt at originality: physical bodies, fields of gravitation, events, abstract particulars (sets, numbers, propositions, and maybe others), and ... persons.

I don't mean to be making a big claim here. I am not saying that persons are particulars which *do*, in the final analysis, belong to the ultimate furniture of the world as we know it, *i.e.*, particulars which cannot be reduced to (combinations of) more basic particulars. I would simply like to rank them among those entities which should be considered carefully as candidates. (Descartes for example, as we shall see, considered them as candidates, but decided not to assign to them the ontological status of basic entities.) – Now, and that's what I'd like to emphasise at this point, the ontological concept of a person should be kept as pure and austere as possible. In particular it should be kept distinct from any psychological notion, however seemingly close, like, *e.g.*, the concept of a personality. A personality, I take it, is something a person *has* (and presumably it is not a particular, but some universal which, at least in principle, different persons may share; but even if personalities would have to be accepted as particulars, they'd be particulars different in kind from persons). What I'm trying to draw your attention to is not that *person* and *personality* are distinct concepts (this is banal). Rather it is the less obvious point that the tight and rigid connections between these concepts run only in one direction. Personality conceptually requires personhood; but not *vice versa*.⁸

The sparse ontological concept of a person I shall consider in the following is psychologically neutral, or noncommittal, in a thoroughgoing way: It does not exclude, for example, the conceptual possibility of one and the same person's changing his or her personality abruptly

⁷ I am thinking here of authors like, *e.g.*, Bernard Williams, Robert Nozick, Derek Parfit, David Lewis and Martine Nida-Rümelin.

⁸ One may be tempted to assume that at least the more specific concept of a *human* person involves having a personality. But I am not so sure about this either; maybe only our concept of a *normal* (or a non-deficient) human person contains personality as a feature.

and completely. Psychological similarity, continuity, or conscious self-accessibility over time is not a *conceptual* ingredient in personal identity. It is, indeed, a *factual* ingredient in the human persons-over-time we are acquainted with. And, indeed again, the absence of this ingredient may make us wonder whether we are really dealing with the same person. But, and that's what I am trying to bring to the fore, there is a basic ontological concept of personhood which does not by itself compel us to deny personal identity in cases of abrupt and vast psychological discontinuity. That's what I mean by calling the concept psychologically noncommittal: it is as if it were silent about these cases. - In focussing on this basic concept, I don't mean to deny that there are other legitimate concepts of personhood (*e.g.*, the concept of a *human* person) – concepts which may be 'more psychological', in the sense just adumbrated. And it may well be that our most familiar concept of a person is not the ontological one. But I think that the ontological one is fundamental, and a powerful source of our conceptual bewilderment.

The second clarification concerns the realism/anti-realism issue. An important question in this context is whether *person* is an ascriber-relative (or recognition-dependent) concept. I shall call the whole family of such concepts *CAC*-concepts, because in their case the mere counting as a so-&-so is constitutive of being a so-&-so. The mark of a *CAC*-concept can be roughly characterized as follows: It applies to the items to which it applies in virtue of the fact that these items count as falling under the concept. That *x* counts as a *C* may be spelled out in various ways, for example as “Given appropriate information about *x*, the vast majority of normal people who have mastered concept *C* ascribe --or would ascribe, if they encountered *x*-- to *x* the property of being a *C*”, or as “A sufficient majority of relevant experts or authorities⁹ accept, or would accept, *x* as a *C*”.¹⁰

It is fairly uncontroversial that many common concepts are of the *CAC* variety: *piece of art*, *fruit*, *disease*, *car*, *jail*, etc. (*Fashion* is, I think, a particularly clear example of a *CAC*-concept: If *x* counts as fashionable --*i.e.*, if a sufficient majority of the relevant *magistri elegantiarum* accept, or would accept, *x* as fashionable--, then *x* is fashionable.) Many concepts of philosophical interest, however, are highly controversial in this respect. If one

⁹ “Relevant majority” may sound inappropriate whenever there are very few experts or authorities. And this happens often enough. (In a soccer match, *e.g.*, there's only one ultimate 'authority' on fouls, goal, etc.) But let that go.

¹⁰ However crucial they may turn out under closer inspection, I am not going to care about differences between various sorts of counting-as here.

considers the concepts of, *e.g.*, beauty, goodness, truth, happiness and justice to be CAC-concepts, this almost inevitably makes one an anti-realist about beauty, goodness, truth, happiness and justice. That is to say, whoever takes concept *C* to be of the CAC-kind, is strongly susceptible to the assumption that, concerning *C*-issues, there is no fact of the matter – no fact, that is, beyond those facts which are about what is, or would be, the considered judgment of a certain range of people about such issues. Whereas a realist about *C*-ness, who deserves this denomination, holds that, maybe subject to some sophisticated qualification, facts about *C*s are ‘genuine’ facts. Genuine facts are not just states of affairs corresponding to beliefs which have been formed by a ‘relevant’ bunch of people, however impeccable the conditions of forming these beliefs. - I shall say a little more about this presently.

My above remarks, about the particular ontological concept I have in mind, may have already made it clear that my metaphysical sentiment about personhood is downright realistic: Facts expressed by sentences, as used in an ontological discourse, of the type “*x* is a person” (or “So-&-sos are persons”) are genuine facts. The concept of a person, at least the ontological one, does not function as a CAC-concept.

Let me try to explain what I mean. Let’s assume for a moment that I’m a person and that you’re a person. So far, no commitment to realism is implied. Here is what *any* realist about personhood is prepared to add: If this assumption is true, then our personhood is a genuine fact. But realism comes in various stripes, among them are disappointingly soft ones. (As it happens, I am a soft realist about fashion. I’m prepared to accept it as a fact that, for example, a certain sort of belts are fashionable. But for me, this fact is merely a CAC-fact: a fact constituted by such belts’ counting as fashionable.) Now that’s not the attitude of a downright, or hard-boiled, realist. He is eager to up the metaphysical ante. As a downright realist about personhood, I am prepared to strengthen the soft realist claim above considerably: “Given that our personhood is a fact, it’s an objective fact as hard as they can get. They don’t come any harder anywhere - not in physics, not in mathematics, not in logic.”

With regard to the epistemological position, accompanying such a strong metaphysical tenet, even a hard-core realist has various options. My own option is this: Don’t confuse facts which are metaphysically first-rate with those which are our epistemological darlings: with plain, obvious, undeniable facts, facts which (if need be) can even be proven, in some widely accepted logic calculus, from premises whose *a priori* truth can be recognized by way of

intuition. To put it differently: The sheer hardness of a fact doesn't entail a corresponding degree of its obviousness; even some of the hardest facts may be rationally put in doubt. - Harking back to the issue at hand, that is to say: By our above assumption, you and I are persons, and, by my hard-core realism, this is an adamant fact, but it is not beyond intelligible doubt.

How could this be? For one thing, others may have their doubts about us and take you and me for zombies, aliens, cleverly designed robots – all of which they presume not to be persons. But more than this, each of us may have doubts about his or her own personhood. If you have such doubts, and if you assume that doubts can be had by persons only, you should, in all consistency, also doubt that whatever it may be that you're having are genuine doubts. So you should be prepared to consider it as possible that what you have are merely your “non-personal” substitutes for genuine doubts, let's call them oubts. Oubts feel (or 'eel') to, and function for, non-persons just like doubts do to real persons.

This may sound more and more crazy, but I tend to think that one isn't immune to such doubt. To illustrate: Assume that Mr Deckard (Harrison Ford), in the movie *Blade Runner*, is a genuine person (in this fiction) and that he himself assumes androids, or replicants, not to be persons. Even if we assume this, as the movie invites us to do, then nevertheless, from a certain point on, Deckard begins, and he does so for understandable reasons, to doubt his very own personhood. Deckard is part of a fiction, but in the story told by the movie, he, a person by assumption, really doubts his own personhood.¹¹ He does not merely oubt it. And his reasons for doubting are fairly good ones, and not just airly ood easons to oubt, whatever this may be. The example is from Hollywood and may therefore seem just foolish. I chose it, because the movie is fairly well known and because it may convince you of the coherency, in principle, of genuinely doubting one's own personhood. Moreover, I speculate that there are various mental diseases of actual people (*i.e.*, human persons) which establish the same point: Human persons may in all consistency begin to doubt that they are persons and may begin even to fear that they don't suffer but instead uffer, whatever this may be, and however much it urts.

¹¹ In Philip K. Dick's novel (*Do androids dream of electric sheep*, New York 1968) on which the movie is based, Deckard's doubt comes out much more clearly. Moreover, at the end of chapter 20, Dick describes Deckard as being aware of the fact that ,reasons' of non-persons have to be distinguished from reasons proper.

So the kind of realism about personhood I think to be an ingredient of the sparse ontological concept of a person has nothing to do with ‘metaphysical’ or ‘absolute’ certainty about personhood. I tend to think that, painful as this may be, there is no such certainty to be had. Not even the supposedly special sort of unshakable ‘subjective’ evidence of self-consciousness, in virtue of which we could prove, at least each one of us for himself, his own personhood. But what is crucial for our purposes, is this: Even if we had ways to gain absolute certainty in this respect, such certainty would be nothing in virtue of which we are persons – would not be constitutive of our being persons.

The kind of hard-core realism I take to be appropriate with regard to persons, ontologically conceived, is therefore not wedded to any sort of epistemological fundamentalism. But neither is it hostile towards any such a position: A realist may well believe that his personhood is a fact of which he has, or can have, most certain knowledge *a priori*. – This is to say, realism about persons is an ingredient of the ontological conception of personhood, but such realism is epistemologically neutral, or at least is compatible with a wide range of positions concerning the question whether and how facts of personhood could be established.

Let me try to summarize, as straightforwardly as I can, the combination, which I’ve just tried to rough out, of hard-core realism and epistemological permissiveness:

Personhood is independent of what one takes oneself to be, or what others take one to be. Even if it should be somehow rationally inevitable for a human being to assume that he himself, or she herself, is a person (or, over and above that, that all of his or her fellow creatures are persons), this itself wouldn’t be what makes any of us persons. And *vice versa*, even if all of us came to believe that we aren’t persons but merely brains, cleverly designed automata, robots or androids, this would not affect the fact, granted that it is one, that we are persons.

The crucial ‘realistic’ point in all this seems to me to be simply this. *Person* is not a CAC-concept. Granted, it may well be true that this concept is man-made. Forming it, or eventually hitting upon it, may have been one of the great cultural achievements in human history. Moreover, the ability to responsibly ascribe it, if only fallibly, to others --and, on reflection, to oneself-- may be a specifically human talent. But nevertheless, what the concept applies to is

not up to us. The fact that an entity falls, or doesn't fall, under this concept is itself fully independent of what humans have achieved, what they are gifted to do, and what they believe.

There's both a sunny and a dark side to such realism. Sunny side: Nobody's personhood depends on anyone's ascription or 'recognition' of it, not even his own. Therein lurks something dignifying and consoling. – Dark side: No-one is immune to doubts, however far-fetched, about his or her own personhood. Why call it a dark side? Because if we ever went so far as to engage seriously in such a doubt, we wouldn't have the slightest idea of what (not anymore: "who") we were. We'd be, as it were, hopelessly lost in the vast ontological zoo. In this lurks something sombrely disconcerting.¹²

The two early modern philosophers whose thoughts about our topic we shall consider later are clearly realists about personhood. But, interestingly (I think), the dark side of being a realist seems not to have occurred to them.

The bewildering conceptual plenitude of *person*

One thing that is deeply vexing about personhood is this: Even our fundamental and utterly austere ontological concept of a person seems inexhaustibly rich. And it is quite unclear which of its features are core components and which are peripheral. Which of its aspects should be considered as being fundamental and which as being derived?

It is this deplorable fact that I shall be mainly concerned with in what follows. I shall try to present, very briefly, some evidence that it *is* a fact. I shall venture an explanation for it, which I shall try to support, less briefly, by two examples from the history of philosophy. And I shall air, very briefly again, a suggestion about what to do when the aim is to "regain a complex concept of human personhood".¹³

Consider the following random list of characteristics of personhood which have been emphasized by various thinkers as they have employed that concept in their theorizing.

A person, it is said,

¹² What I find so especially outrageous about recent pseudo-scientific stupidities, a few examples of which I mentioned above, is that their authors do not seem to be sensitive at all to the depth of such disconcertion.

¹³ I am borrowing this phrase from Michael Welker's title of the first meeting of our group.

- 1 is an individual capable of rationality
- 2 is responsible for what it does
- 3 has dignity
- 4 is not a something ("*quid*") but a someone ("*quis*")
- 5 is free
- 6 is a unity of a body and a mind (soul)
- 7 is anything to which words and actions of human beings are attributed
- 8 is an intelligent agent, capable of a law, and happiness, and misery
- 9 is an end in itself and an object of respect
- 10 is an entity to which both mental and physical properties can be ascribed
- 11 is capable of treating others as persons
- 12 is capable of verbal communication
- 13 is conscious and self-conscious
- 14 is capable of second-order intentionality (in particular, is capable of second-order volitions which are a precondition of having a free will)

Many of these features themselves do not seem conceptually less demanding than personhood; many of them are somewhat vague. Some of them may appear controversial. (As to #11, for example, there are forms of autism, or so I am told, which disable people from treating others as persons. But we would not be ready to accept without reservation, I presume, that anyone who suffers from such a disease is *ipso facto* not a person.¹⁴) Arguably, not all of these features go together. (For example, #10 is so wide that it seems to allow for persons who don't exemplify several of the other features.) Clearly, several of these features seem to be dependent on others and so, maybe, this list needs to be reduced. But even more clearly and most importantly for our purposes, there is nothing about this list which gives us reason to assume that it is complete. The list is heterogeneous and it is open; and for all we know, it is essentially open. That is, we have no idea about how, by what kind of argument, we could possibly convince ourselves that it, or some improved variant of it, is complete.

¹⁴ This is not to say that *person* is not a QS-concept, *i.e.*, a concept which essentially involves a certain standard of quality in the following sense: It is part of the mastery of such a concept that one acknowledges, concerning the items subsumable under it, that they can be classified according to how good they are as items falling under this concept. Roughly speaking, if *C* is an QS-concept, then it is fully mastered only by someone who has also mastered a family of concepts such as "an excellent *C*", "a good *C*", "a middling *C*", "a lousy *C*", etc. – An example of such a concept would be *argument*; you don't really know what an argument is, as long as you have no idea of how to classify arguments according to their quality as arguments. But you may very well know what a logical proof is, without even being willing to classify such proofs as good or bad ones. So *proof* is not a QS-concept. Three more remarks on QS-concepts: First, they do not need to be evaluative themselves, although their mastery essentially requires the ability to draw value distinctions concerning the members of their extensions. Secondly, it is characteristic of the natural sciences (at least of the more fundamental ones, and clearly of physics) that their theoretical terms do not express QS-concepts. Third, QS-concepts are not reducible to concepts which do not involve standards of quality.

I am not sure what to say about *person*. But I think it is an interesting question whether it is a QS-concept or not. If it is, or were, one, then it may be difficult to stick to the view (which I have taken here) that there is a 'psychologically non-committal' concept of personhood. That's why I am inclined to assume that *person* is not a QS-concept.

So, on the one hand, personhood appears to be a straightforward matter: As a matter of fact, we can, in normal circumstances, tell a person from anything else with remarkable ease. On the other hand we do not have a clear idea of what the crucial marks of personhood are. The features which come to mind when we think about it are too many and too motley, to elucidate what we really mean by "person"; and we are prepared to admit that ever more features may turn out to be conceptually relevant, as we keep on thinking about it. Moreover, there is no reason to think that the word "person" is ambiguous. It would be absurd to claim that the features listed above specify distinct meanings of the word. "Person" clearly is not like "bank" ("ground beside a river"/"institution offering financial services"). It is exactly the fact that "person" is *not* a homonym which makes the essential openness of any collection of its conceptual features an embarrassing richness.

*

How is this richness of the concept of personhood to be explained? One answer to this question is historical. Over the centuries, the concept has been used by many thinkers as a conceptual tool for answering quite different questions: metaphysical, theological, and moral. In reaction to these problems, quite different features have been introduced as characteristics of a person. So the word "person", for a very long time, has been a technical, or semi-technical, term in various quite distinct theoretical frameworks, and it has been used in these frameworks for the solution of various quite distinct theoretical problems.

I shall try to illustrate this by two examples from the history of philosophy which I take to be quite telling. I hope that they reveal some aspects of the complexity and heterogeneity of our inherited concept of a person which has been partly formed (reformed and, arguably, deformed) by thinkers like Descartes and Locke.

Descartes on personhood

Let us consider, as a first example, the use Descartes makes of the concept of a person. Although he acknowledges the existence of God and angels, his doctrine is exclusively about human personhood. According to his metaphysics, any human being consists of two entities which are really distinct: the body and the soul, or mind. They are really distinct, because the

body is a physical substance and the soul (or mind) is an immaterial substance, and these two substances could exist without each other. It should be noticed that what Descartes calls a *real* distinction between substances is not a factual separateness, but a possible one: two substances are really distinct if they are capable of being separated, "at least by God" [AT VII 78]. In the *Sixth Meditation* Descartes presents (the definitive version of) his famous proof that his body and his soul are really distinct. The crucial point of the proof is this. Descartes claims that one can clearly-and-distinctly think of oneself insofar as one is only a thinking thing and not a material (or extended) thing; and one can clearly-and-distinctly think of one's body insofar as it is merely an extended thing and not a thinking thing.¹⁵ Whenever anyone can clearly-and- distinctly understand one thing apart from another, God could have created these things in that way. And this is to say: his or her soul and his or her body are really distinct things. One can, in principle, exist without the other.

Nowadays, Descartes' mind/body dualism is considered by many as an unscientific folly of a clueless and pious philosopher, especially by those who don't know his work. But we should not forget that he was a sober man and an accomplished scientist. In fact, (much more than to metaphysics, *prima philosophia*) he was devoted to natural science, *philosophia naturalis*, and specifically to the project of explaining the totality of phenomena in the world as we know it in terms of a mathematico-physical theory. In such a theory matter is hypothesized to consist of nothing but micro-elements (too small to be humanly perceivable), and these elements are ascribed nothing but decent corporeal properties like shape, size, motion and position. Descartes was a naturalist and reductionist, pretty much in the way which is common among scientists today. But the human mind (specifically *human* cognition, in contrast to animal cognition) he considered as a phenomenon which defies any naturalist-reductive account. In fact, he seemed to have held that no science whatsoever of the human mind is possible. The doctrine that the human mind is scientifically impenetrable wasn't due to some unfounded defeatism on his part. For Descartes, it rather results from a certain speciality of the human mind: the pure intellect (and maybe also from another one: the absolute freedom of will). The *intellectus purus* is a mental capacity completely independent of the others we have (like, *e.g.*, sense-perception, memory or imagination; these are

¹⁵ The hyphens in "clear-and-distinct" are meant to remind you that this is a technical term of Descartes'. An idea, or an perception, is *clear*, if it is vivid (like the idea of pain when you suffer from one); it is *distinct*, if it is sharply separated from all other ideas (as the idea of pain is not, according to Descartes, since we have a tendency to mix up the sensation of pain itself with something painful in the cause of the pain). – But the term "clear-and-distinct" has a very special meaning for Descartes: it is reserved for those ideas of which it cannot be assumed, on pain of manifest absurdity, that they are misrepresentations.

importantly body-bound and mental, at least in human beings, only in virtue of the fact that they are connected with the intellect). The intellect is our capacity to genuinely understand things and to conceive their essence - our capacity to theorize by employing concepts which have been purified from any sensory, pictorial or other admixtures. When Descartes claims that animals (most probably) do not have a mind, he does not deny them sense-perception, pains, desires, etc. They have all this, he agrees, but not as genuinely mental phenomena, *i.e.*, not as something which informs an intellect. The right way to understand Descartes' disbelief in animal minds is to understand him as precisely believing that they do not have a pure intellect. For him, the pure intellect is the mind proper, the original and true mind; nothing else is intrinsically mental. Several other human capacities, acts and processes are mental only in virtue of being appropriately connected with the intellect.

The real distinction between human mind and body is, for Descartes, a fact of metaphysics. But metaphysics is not everything there is in life. Not even for Descartes. As he says during a conversation with the theologian Frans Burman: "A point to note is that one should not devote so much effort to the *Meditations* and to metaphysical questions, or give them elaborate treatments in commentaries and the like. Still less should one ... dig more deeply into these questions than the author [*i.e.*, Descartes himself] did; he has dealt with them quite deeply enough. It is sufficient to have grasped them once in a general way, and then to remember the conclusion. Otherwise they draw the mind too far away from physical and observable things, and make it unfit for studying them. Yet it is just these physical studies that it is most desirable for people to pursue, since they would yield abundant benefits for life" [AT V 165]. And in a letter to Princess Elizabeth he puts this point as follows: "I believe that it is very necessary to have properly understood, once in a lifetime, the principles of metaphysics, since they are what gives us the knowledge of God and of our soul. But I think also that it would be harmful to occupy one's intellect frequently in meditating upon them" [AT III 695].

The metaphysical conclusion that our minds and our bodies are distinct entities is hard to bring into unison with how we experience ourselves. This conclusion is true, it is even shown to be absolutely certain by a metaphysical proof, Descartes insists. But he recommends leaving it at that. The way we experience ourselves, he concedes, is not as consisting of two distinct entities; rather we experience ourselves as the union of our soul and our body. But this union is, in reality, not anything which exists *sui generis*. There is no third entity, over and

above our body and our soul.¹⁶ When it comes to taking stock of the basic really existing entities, then, strictly speaking, there are only the two substances of body and soul which are distinct however intimately they may be interrelated. So, in a sense, when we experience ourselves as a mind/body-union, the way we experience ourselves is not mirrored in the basic metaphysical facts.

It is exactly this union of body and soul which Descartes denotes by the concept of a human person. "Everyone feels that he is a single person [*une seule personne*] with both body and thought [*i.e.*, soul] so related by nature that the thought can move the body and feel the things which happen to it" [AT III 694].

But, as he makes it clear, particularly in his correspondence with Elizabeth, Descartes is prepared to concede that this way of experiencing ourselves as persons is not just due to some sort of negligence or other kind of avoidable mistake. He says, surprisingly, that among our primitive notions which are innate and "can only be understood through themselves", there is not only the notion of body and the notion of mind, but also the notion of their union [AT III 665]. This is surprising, since –in the final analysis-- there is, as we have just seen, no thing to which this notion applies in reality, and therefore the notion of a person is, metaphysically speaking, at least a misleading one. Whereas both the soul and the body can be conceived by the pure intellect, their union, Descartes says, "is known only obscurely by the intellect alone ... but it is known very clearly by the senses" [AT III 692]. This means for Descartes: Although we have very strong and vivid ideas of the senses concerning the union of the body and the soul, these ideas never amount to genuine knowledge, since our senses can *never* give us ideas which constitute knowledge, not even when they are clear (*i.e.*, strong and vivid). Genuine knowledge consists in the intellect's perceiving clear-*and-distinct* ideas. It is only such clear-*and-distinct* ideas of the intellect which God guarantees to be true. But, to repeat, the ideas we have of the mind/body-union, Descartes insists, are not clearly-*and-distinctly* perceived by the intellect. So when Descartes says: The union of mind and body is "known very clearly by the senses", we must not forget that the knowledge in question is at best

¹⁶ There are some attempts at terminological appeasement. In a letter from January 1642 to his follower Regius, a professor of medicine at the university of Utrecht who later caused severe trouble for him, Descartes recommended, as Regius' ghost-writer in his dispute with the Dutch theologian Voetius, the following formulations: "... human beings are made up of a body and a soul ... by a true substantial union [*per veram unionem substantialem*] ... If a human being is considered in himself as a whole [*homo in se totus*] ... he is a single *Ens per se*, and not *per accidens*; because the union which joins a human body and a soul to each other is not accidental to a human being, but essential, since a being without it is not a human being" [AT III 508]. – This is intended to sound soothing, but the plain fact remains: mind and body are distinct substances, and their union, even if a "true substantial" one, is not a substance.

second class knowledge, or strictly speaking: not knowledge at all. What we do have, when we experience ourselves as *persons*, is nothing but vivid ideas of the senses, but no clear-*and-distinct* ideas of the intellect.

As soon as the intellect, in a metaphysical effort, has brought the ideas both of body and of soul to clearness-and-distinctness, and has achieved the insight that body and soul are really distinct, it faces what we nowadays call Descartes' mind/body-problem: How can there be a causal interaction between these entities, one of them material, the other immaterial? When Frans Burman asked him, in 1648: "But how can this be, and how can the soul be affected by the body and *vice versa*, when their natures are completely different?", Descartes replied: "This is very difficult to explain; but here our experience is sufficient, since it is so clear on this point that it just cannot be gainsaid" [AT V 163].

So here is why the use Descartes makes of his concept of a person is important for him: Although we have no clear-and-distinct idea of a person, this idea is a primitive innate notion which cannot be reduced to notions which are clear-and-distinct. It is as persons that we experience ourselves quite naturally, as long as we do not philosophize about our nature. And as long as we experience ourselves in this natural way, the mind/body-problem simply does not arise.

That is why people who never philosophize and use only their senses have no doubt that the soul moves the body and that the body acts on the soul. They regard them as a single thing, that is to say, they conceive their union; because to conceive the union between two things is to conceive them as one single thing. Metaphysical thoughts, which exercise the pure intellect, help to familiarize us with the notion of the soul; and the study of mathematics ... accustoms us to form very distinct notions of body. But it is the ordinary course of life and conversation, and abstention from meditation ... that teaches us how to conceive the union of the soul and the body. [AT III 692]

Descartes seems to suggest here, and in other passages,¹⁷ that metaphysics (pure thinking, performed by employing clear-and-distinct notions of the intellect) does not and cannot give us the solution to the mind/body-problem. The way to deal with this problem is rather to *dissolve* it, by recognizing that it simply does not arise as long as we experience ourselves in the way which is most natural for us: as persons. So the concept of a person is used by Descartes as a philosophical tool for the dissolution of a problem –indeed, a mystery-- arising in his metaphysics.

The dissolution he hints at seems to be along the following lines: The so-called mind/problem cannot be solved theoretically, because we have no clear-and-distinct ideas in terms of which we could explain how an immaterial mind and a material body form an interactive union. It cannot be solved, because no conceptual tools required for a theoretical solution are available. Human cognition is such that it has no access to any category of entities beneath, or beyond, the categories *mind* and *body*. For us, these two categories are rock-bottom. Even our conception of God is within these two basic categories (we have to conceive Him as a mind, albeit an infinite one). Don't ask why God didn't give us the conceptual resources to solve this problem theoretically.¹⁸ There is no such problem, except when you philosophise. Apart from this special case, you are dead sure that you are a person. Your assuredness about this is not metaphysical certainty, but it is good enough for all matters of human concern. You enjoy it, except when doing metaphysics, because God was kind enough to let you constantly *feel* that you are a person. Don't complain that you are not capable of reaching full-blown ("distinct") understanding *how* there can be such unions of mind and body. Be grateful for clarity about this issue, for vividly feeling *that* you are a person. – This may sound pious, or like a cheap escape. But as far as we can tell, Descartes was fully serious about it. He accepted it as obvious that there are lots of things about which we, as finite minds, aren't capable of reaching *scientia*, knowledge in the strictest sense.

According to the Cartesian account, the concept of a person is not a 'theoretical' concept which could help us to gain metaphysical insights into the ultimate structure of reality. It is not clear-and-distinct; it is not one of those concepts by which we can reach genuine

¹⁷ E.g., in a letter to Arnauld (July 29, 1648), where he writes: "That the mind, which is incorporeal, can set the body in motion is something which is shown to us not by any reasoning or comparison with other matters, but each and every day by the surest and most evident experience [*certissima & evidentissima experientia*]. It is one of those things which are known by themselves and which we only make obscure when we try to explain them" [AT V 222].

¹⁸ In the *Fourth Meditation*, Descartes argues that asking such questions betrays a fundamental misunderstanding.

knowledge. In the letter to Elizabeth from which I have quoted extensively, Descartes says: "It does not seem to me that the human mind is capable of forming a very distinct conception of the distinction between the soul and the body and, at the same time, of their union; for to do this it is necessary to conceive them as a single thing and at the same time to conceive them as two things; and *this is absurd*" [AT III 693, my italics].

Taking all this together, I suggest that Descartes' thought is this: When you do metaphysics, when you inquire into the ultimate structure of what there is, you are bound to accept that your soul and your body are really distinct; and then, as long as you are engaged in nothing but pure metaphysics, you cannot conceive of yourself as a person (*i.e.*, of the union of your body and your soul). Strictly metaphysically speaking, this is not just too difficult, it would be simply absurd. At the end of the day, the concept of a person is not just confused, but it is in principle so and for a simple reason: the very concept is in tension with an irrefutable metaphysical fact. Nevertheless, this concept (which God was kind enough to put into our souls) is of enormous value. It captures an important aspect of our worldly existence, "which everyone invariably experiences in himself without philosophizing" [AT III 694].

Let me list a few salient features which are characteristic of Descartes' concept of a person as I have just sketched it:

- (1) The concept of a person is the concept of the mind/body-union.
- (2) This concept is innate and a primitive, *i.e.* unanalysable, concept.
- (3) It is not clear-and-distinct, and since it is primitive, it cannot be reduced to clear-and-distinct concepts. So we may say that it is essentially not clear-and-distinct.
- (4) Nevertheless, it is of enormous value. Not because it helps us to solve the mind/body-problem, but because it helps us to dissolve it.

Descartes' silence on transtemporal personal identity

Assuming for a moment that this sketch of Descartes' doctrine, concededly an unorthodox one (this is a concession, not an apology), is on the right track, there is little wonder that he never cared to raise questions of transtemporal personal identity. I have wondered for many years, if you allow me to intersperse a personal (*sit venia verbo*) remark, why Descartes, otherwise a

most subtle thinker on topics concerning the metaphysics of the mind, was apparently never puzzled by the problems about fission and fusion, the body-hopping of minds (or the mind-hopping of bodies, if that makes a difference – I think it does) and all that kind of weird stuff which seems to spring immediately from his substantial mind/body-dualism.

So why was Descartes, of all thinkers, never puzzled by these questions which have occupied metaphysicians ever since Locke's *Essay*, and which seem to be taken bitterly seriously in recent metaphysics - indeed, today seem to be considered more urgent and important than ever? A tempting answer goes as follows: Because, for him, these are all pseudo-problems. A problem which wears its insolvability- *in-principle* on its sleeves is a pseudo-problem. To put it in a bunch of slogans: "There's no *puzzle* of transtemporal personal identity. If the relevant questions could be framed at all, they could be framed clearly-and-distinctly; and then they could be answered. But they can't be framed clearly-and-distinctly, since they essentially involve the concept of a person.¹⁹ A question which *in principle* cannot be phrased clearly-and-distinctly is a pseudo-problem; it simply has no answer."

This, I gather, was not Descartes' reason for avoiding issues of transtemporal personal identity. The problems in question would be pseudo-problems for him only if the concept of a person were a ("materially") false idea, *i.e.* one which is "such as to provide subject-matter for error" (AT VII 231) by not representing anything real, but representing what they represent as something real (AT VII 44). But *person* is not a false idea. What it represents is something real (the mind and the body as a union), so whatever is wrong with it is not that it represents something as real which is not real. What is cognitively inferior about it, in comparison to concepts like *mind* and *body*, is that it essentially represents its *repraesentatum* indistinctly (or as Descartes would put it: "*confuse*", which is his technical term for the opposite of "*distincte*"). Yet this, by itself, is not a stain on its conceptual credentials. For its rationale is exactly to represent two-things-considered-as-one. Its appropriate realm of application is *outside* metaphysics. (Within metaphysics, mind and body demonstrably are to be considered as two distinct things. But as I said: Metaphysics is not all there is in life, not even for Descartes.)

¹⁹ All these puzzling questions (*e.g.* "Would somebody, let's call him E.P., who enters, on Earth, a Parfitian Teletransporter be the same *person* as the one who, on Mars, leaves the teletransporter, given that the brain and the body in the cubicle of the Earthian Teletransporter were destroyed in due time?", "If E.P. were teletransported twice over and subsequently destroyed, would any of the two duplicates be the same *person* as E.P.?",) involve the concept of a person *essentially* -- *i.e.*, they could not be rephrased without this concept.

For Descartes, the concept of a person is a fine concept, for *the conduct of life*. It is of utmost importance within this realm. It is a concept which captures an important aspect of the human condition.²⁰ And it would betray a grave intellectual misunderstanding to sneer at it because of its lack of distinctness. A concept's lack of distinctness is not, *per se*, a conceptual deficiency. This sort of lack is the hallmark of many perfectly good concepts. In fact, the vast majority of the concepts on which we have to depend in order to lead our humble human lives are indistinct in not separating their bodily and their mental components: hunger, thirst, love, pain, sweet, soft, red – to mention but a few.

Nevertheless, *person* is merely a second-class concept when it comes to *the contemplation of truth*.²¹ The contemplation of truth is to the conduct of life like a move in a game of Blitz chess is to its analysis without time-limit. A perfectly good move in the one context may not live up to the standards of the second.

Therefore, given that the concept of a person can, in principle, not be brought to distinctness, questions about transtemporal personal identity, for Descartes, are fated to imperfect answers (all of them, not only those bizarre cases which are characteristic of our contemporary debate). No answer could possibly possess genuine certainty. True knowledge, *scientia* in the emphatic Cartesian sense, is restricted to the realm of our most clear-and-distinct thoughts. A crucially important philosophical fact about transtemporal personal identity is that no knowledge *sensu stricto* is to be had on the topic – and that therefore, in a sense, personal identity is not a metaphysical topic at all. The only 'knowledge' that could be hoped for would be epistemically second-class, knowledge merely "in the moral sense [*moralis sciendi modus*] which suffices for the conduct of life" (AT VII 475).²²

In a nutshell, Descartes' view might well have been that the questions of transtemporal identity aren't pseudo-problems, but neither are they questions to which a philosophical

²⁰ Many things which, as it happens, only human beings can do are things which, for conceptual reasons, only persons could do. Michael could take a stroll, but his dog Carl could do so only in a non-literal sense of this phrase. Why? I guess it's not because members of the biological species *homo sapiens sapiens* have this ability, and as a matter of empirical fact, members of the species *canis canis* happen not to have it (so that one day a new breed of dogs might turn up whose members literally could take a stroll).

²¹ For the Cartesian distinction between the conduct of life and the (metaphysical) contemplation of truth see AT VII 149.

²² In this passage of his *Seventh Replies* to (Bourdin's) objections, Descartes adds: "I frequently stressed that there is a *very great difference* between this type of knowledge and the metaphysical knowledge ..." (*ibid.*, my italics). The very great difference lies in the following: Only metaphysical knowledge has God's truth-guarantee; He would, *per impossibile*, have to be a deceiver, if our alleged metaphysical knowledge turned out to be false belief. But concerning our alleged knowledge in the mere moral sense, God's benevolence does not guarantee the truth of what we believe. Moral certainty inextricably contains an element of epistemic risk.

answer could be given. We would have to try to find answers (or rather: practical decisions about how to deal with the situation), if we were confronted, in practice, with a problem-case. For such cases, we could not bring to bear moral certainty and not even practical knowledge [*connaissance en pratique*], since the latter would at least require a firm habit of belief (AT IV 296), which we could not have acquired concerning novel extravagant situations (body-hopping of souls, etc.). Our guidance would have to be good common sense [*sens commun bon*, AT XI 386; *sensus communis*, in the non-technical sense, AT X 518, 527], which Descartes mentions occasionally, but does not theorize about.

Now suppose we were to actually confront such a case, *e.g.* one in which "the Soul of a Prince, carrying with it the consciousness of the Prince's past Life, enter[s] and inform[s] the Body of a Cobbler as soon as deserted by his own Soul"²³, and had to face the question whether the cobbler now is the same person as the prince. From a Cartesian point of view, no answer could be given with certainty, not even with moral certainty.

A narrow-mindedly straightforward application of the criterion for transtemporal personal identity suggested by Descartes' concept of a person would yield the negative answer: No, the cobbler-now is not the same person as the prince-then. For personal identity, according to Descartes, obviously would have to be identity of the mind/body-union; and the prince's mind and the cobbler's body clearly constitute a union very different from the prince's original mind/body-union. But the strategy of, *first*, concluding that the cobbler is not the ex-prince and *then* drawing whatever consequences from this result as if it were a theorem proven, presumably would not be what our good common sense recommends.

It would display more common sense to take into account what concrete practical consequences are at issue. (For example, is there a large sum of money the prince-then owes to somebody, and are we facing now the question whether the cobbler-now or the prince's wife should pay the debt? Or is the question whether the cobbler-now ought to be hanged for a crime, committed by the prince-then? etc.). Get clear about what, *in concreto*, is at issue in this particular situation, and in the light of this and of all that you know, if only with moral certainty, discern the best solution²⁴ to this concrete problem with all of its contingent features. – This may sound convoluted, as a piece of advice delivered by common sense. But then again, common sense may be more refined than the scoffers would concede. Its maxims

²³ Locke, *Essay concerning Human Understanding*, II.27.15.

²⁴ Or rather: "one of the best solutions"; for there may be more than one optimal solution.

may not be confined to what can be expressed in six-word sentences without hypotaxis. Descartes thought very highly about common sense – where it belongs. And, for him, it indispensably belongs to all the matters, where problems of personhood are concerned.

*

Let's turn to something else. It is worth emphasising that for Descartes mind-identity is not sufficient for personal identity. He has not explicitly formulated a criterion of transtemporal personal identity, but it is quite clear that, given his concept of personhood, his doctrine would yield the following criterion:

Person *A*, at *t*, is the same person as *B*, at *t'*, if and only if (i) the mind of *A* at *t* is the same mind as the mind of *B* at *t'* **and** (ii) the body of *A* at *t* is the same body as the body of *B* at *t'*.²⁵

It is a common mistake to assume that Descartes is implicitly committed to a purely mental criterion of personal identity. The reason for this mistake, presumably, is this: According to the Cartesian doctrine, I could exist without the body I happen to have; I could even exist without a body; but I could not exist without my mind; and this is to say, my essence is my mind and nothing physical is part of my essence. Therefore: if one's mind is one's total essence, then mind-identity is that which (completely) constitutes personal identity.

But this last step is a *non-sequitur*: more specifically, it is a fallacy of equivocation. For Descartes, there are two ways of using the word "I". If it is used, as almost always, in the common way, it refers to the speaker (or thinker) as a person, *i.e.*, as a mind/body-union. But it can also be used in a special, technical sense, in which it refers to some particular aspect of what it usually refers to. In the *Meditations*, Descartes' thinker is for quite awhile not in a position to refer to himself as a person, because he cannot yet exclude the possibility that there are no bodies at all in the world, not even his own. In order to make sure that he nevertheless refers successfully, when he uses the word "I", he uses it in a specially narrow sense (roughly, in the sense of "the entity whose existence has been proven with utmost certainty in the *Existo*-argument"). When he uses the word in such an exceptionally restricted way, Descartes speaks of using it *praecise*. It is important to notice that "precisely" here does

²⁵ Note that transtemporal body-identity need not be strict "atom-to-atom" identity.

not mean “in the word’s exact (proper, real, strict or genuine) sense”. What it rather signifies is that the word is used *in a technically restricted sense*. Descartes sometimes cares to distinguish between these two uses by applying phrases like “*ego totus*”²⁶, in contrast to “*ego quem novi*”.²⁷ The metaphysical result that my mind is my complete essence is a truth exclusively in the second, technical sense of “my”. From this, nothing can be inferred to the effect that my mind is my complete *personal* essence.

Descartes is committed to the criterion for human personal identity just mentioned (same mind & same body). But how this criterion would have to be applied to the enormous variety of bizarre possibilities discussed as problems of transtemporal personal identity, is a matter about which he, at least in published writing, simply remained silent. And for this, as we have seen, he may have had very good reasons: first, these problems do not have a strictly philosophical or otherwise *a priori* justifiable answer; second, as long as we do not encounter these problems, there is no practical reason for dealing with them; and third, as long as we do not know the practical consequences of our answers, there is not much which could guide our good common sense when we attempt to come up with an answer. And common sense is all we could rely on in such cases.

Locke on transtemporal personal identity and personhood

For Locke, the concept of personal identity is an important one because the justice of all reward and punishment, whether performed by us or performed by God, depends on whether the one who did it is the same person as the one who is rewarded or punished. Our best clue of what we *really* consider personal identity to consist in does not come from metaphysics (“the same immaterial thinking substance”), physics (“the same material body”), or biology (“the same human being”) but rather from how we proceed in applying our laws. The fact that we do not punish (and would not consider it just to punish) “the *Mad Man* for the *Sober Man*’s actions, nor the *Sober Man* for what the *Mad Man* did”²⁸ is of utmost importance. For Locke, this shows that when serious practical decisions need to be made, we treat the sober man and the mad man as different persons (his actual wording is “thereby making them two Persons”). If they are two persons, this is so in spite of the fact that, physically speaking, they are

²⁶ E.g. AT VII 81, where he adds “insofar as I am composed of a body and a mind”.

²⁷ See for example AT VII 27.

²⁸ *Essay concerning Human Understanding* II.27.20.

(approximately) the same body, in spite of the fact that they are, biologically speaking, the same human being and in spite of the metaphysical presumption that their immaterial thinking substance is one and the same.

This fact about what we consider just (namely: not punishing somebody, *e.g.* the sober man, who is physically, biologically and mind-substantially identical with the wrong-doer) is, I suggest, Locke's *ur*-observation about transtemporal personal identity. Certainly, his conclusion (concerning personal non-identity) is not inevitable. (We may, instead of jumping to Locke's conclusion, prefer to say that in certain circumstances we do not punish the very person who did the deed.) But I am not concerned here with the feasibility of Locke's theory, but with trying to bring to notice what sort of problem he is actually dealing with, how the concept of personhood is meant to be serviceable to a solution, and what is supposed to motivate the proposed solution.

If I am right, Locke's primary target-concept was transtemporal personal identity, not personhood. Where do we actually, and in a serious and responsible manner (not just in the context of idle metaphysical speculation), apply this concept? What is characteristic of this particular sort of application? The answers to these questions pave the way to our best understanding of what personal identity is. Once we have reached such an understanding, the subsequent clarification of a fitting concept of personhood will be light work. Two equations have to be solved, in the right order. So I suggest the following as the Lockean agenda:

1. Personal identity = that relation, whatever it is, that makes it just to reward/punish someone for something that was done in the past
2. Person = that entity, whatever it is, which is a proper relatum of this relation

Famously, Locke offers consciousness (or more specifically, conscious memory) as the solution for the first equation. The rough idea is this: Person *A*, at *t*, is the same person as *B*, at *t'*, if and only if *B*'s consciousness at *t'* could contain a memory of an action consciously performed, or a thought (consciously) had, by *A* at *t*. (If *A* committed crime *c*, then the relation between *A* and *B* which makes it just to punish *B* for *c* is *B*'s (potential) memory of having done *c* – or somewhat more complicated: *B*'s being able to remember a thought *θ* such that *θ* was a thought of *A* at *t* in virtue of which *A* was conscious of committing *c*. The crucial point

is that the relation in question is a psychological relation obtaining between conscious states: one particular conscious state θ of A at t , e.g. A 's awareness of doing c , and another particular conscious state of B , θ' , which is B 's memory of θ . Given this psychological relation, A and B are "by the same consciousness ... united into one Person" (*E* II.27.10).

Equally famously, in solving the first equation, Locke starts with what he presents as an uncontroversial specification of personhood:

... what *Person* stands for ..., I think, is a thinking intelligent Being, that has reason and reflection, and can consider it self as it self, the same thinking thing in different times and places ... (II.27.9)

What is needed, for a solution of the second equation which is satisfactory in the light of the proposed solution of the first equation, is a close connection between this concept of a person and the concept of consciousness (which is all that constitutes transtemporal personal identity). Locke makes the desired connection as close as possible: as consciousness "unites" persons over time, it unites simultaneous mental states into the same person's mental states.

[consider it self as it self] ... which it does only by that consciousness, which is inseparable from thinking, and as it seems to me, essential to it: It being impossible for any one to perceive, without perceiving, that he does perceive. When we see, hear, smell, taste, feel, meditate, or will anything, we know that we do so. ... For since consciousness always accompanies thinking, and 'tis that, that makes every one to be, what he calls *self*; and thereby distinguishes himself from all other thinking things, in this alone consists *personal Identity*, i.e. the sameness of rational being. (II.27.9)

Beneath the surface of Locke's account of personal identity, both of the momentary and the transtemporal sort, something is at work which deserves our attention. Locke had a strong dislike for the concept of a substance. He scolds traditional philosophy for "the promiscuous use of so doubtful a term" (II.13.18); in using the word "substance", he says,

... we talk like Children; who, being questioned, what such a thing is, which they know not, readily give this satisfactory answer, That it is *something*; which

in truth signifies no more, when so used, either by Children or Men, but that they know not what; and that the thing they pretend to know, and talk of, is what they have no distinct *Idea* of at all, and so are perfectly ignorant of it, and in the dark. (II.23.2)

So there is at least one completely different sort of *desideratum* for Locke's account: A person should not turn out to be a substance of some sort. It is this, I presume, which makes consciousness so irresistible to Locke: The concept of consciousness, for him, is not the concept of a substance, neither a material nor an immaterial substance which, allegedly, is the permanent, indivisible underlying *substratum* of all mental activities. (Whereas Descartes took the concept of a substance to be metaphysically inevitable and crystal clear, but the concept of personhood to be essentially obscure and merely practically helpful, Locke took the concept of substance to be metaphysically inevitable and hopelessly obscure, but the concept of personhood to be perfectly faultless.)

Let's take stock of some of our findings in Locke; for the sake of perspicuity I arrange them in an order which gives us the Lockean echo to the four Cartesian tenets listed above:

- (5) The concept of a person is the concept of an entity which is justly rewarded/punished for its doings (including mental doings).
- (6) This concept is neither innate (there are no such concepts, according to Locke) nor primitive, but rather a complex idea which is reducible to the concepts of consciousness and memory.
- (7) It is clear and distinct, since it is made of clear and distinct simple concepts. (Whereas our use of the *word* "person" creates obscurity.)²⁹
- (8) It is of enormous importance. Not because it helps us to solve, or dissolve for that matter, the mind/body-problem, but because it is central to our conceptions of justice and self-care.

Two ways of accounting for the bewildering plenitude of *person*

²⁹ See *E* II.27.28. – It is exactly this alleged linguistic obscurity ("ill use of Names") which makes it necessary for Locke, following a suggestion of Molyneux', to include a separate chapter on these topics into the second edition (1694).

It would be rash to explain the remarkable clash between Descartes' and Locke's tenets as a manifestation of the fact that the two thinkers are not really addressing the same topic (*i.e.*, the same concept of personhood). What I'd rather suggest is something else, namely that the concept of a person, in virtue of its indeterminate richness, lends itself to wildly different accounts; and that the accounts which have been developed by various influential thinkers, sometimes within incommensurable theoretical frameworks and inspired by disparate philosophical motivations, additionally have left discordant marks on what we, today, dubiously consider as our 'intuitions' about personhood.

As I said, and as the two examples in the excursion are meant to demonstrate, the word "person", for a very long time, has been a technical, or semi-technical, term in various quite distinct theoretical frameworks, and it has been used in these frameworks for the solution of various quite distinct theoretical problems. Moreover, again for a very long time, the word "person" has been in common use as a non-technical term which is not connected to any particular theory or problem, but which has nevertheless surreptitiously incorporated in its meaning an indefinite amount of the semantical complexity just indicated. Maybe what manifests itself as abundant richness inherent in our concept of personhood is only a reflection of the fact that we do not have a shared intuitive grasp of it, but only a common learned tradition, which has bequeathed to us a blend of quite diverse conceptual features that were never meant to go together.

Even if this is true, there may be another explanation for the conceptual richness. It has to do with a certain tension right at the core of our concept of a person:

- (1) The concept of a person is **anthropocentric** in its actual application. Leaving God (and angels) aside, the only clear cases of persons we are *familiar* with are human beings.
- (2) It is **not at all anthropocentric** in its intension. The concept of a person is not supposed to be the same as the concept of a human being.

There is conceptual leeway both for the possibility of non-personal human beings (members of our species lacking exactly those features, whatever they are, which are constitutive of personhood) and for the possibility of non-human persons. Any kind of creature could be, or could turn out to be, a person, if it only had that special something, whatever it is, that makes

us persons. Fairy tales, novels and movies keep reminding us of this non-anthropocentric aspect: Hauff's stork is a person, Shelley's monster is a person, Mathison&Spielberg's E.T. is a person, we are pretty sure that some of the androids or replicants in Dick&Scott's *Bladerunner* are persons, and we are supposed to wonder whether Clarke&Kubrick's computer HAL is a person. - If we try to specify what this conceptual leeway comes to, presupposing as we should that any normal human being is a person, we might look at the following two identifications:

Personhood = that, whatever it is, *without* which a common human being would be only
biologically speaking a human being

Personhood = that, whatever it is, *with* which any being whatever would be, at least, of the
same standing as a common human being

These equations may look funny at first sight, but I gather that they capture an important aspect of our concept of a person. And they may explain the embarrassing conceptual richness we found vexing: The list of features by which these two equations can be "solved" may be essentially open.

The first equation makes it quite clear that "common human being" is not to be taken in a biological sense. It is an honorific term for *us* (who happen to be common human beings) and for every possible being which is of the same standing. It is built into the very concept of a person that there is something valuable about common human beings (an accidental feature which each of them may lack) in virtue of which they are, as it were, not *merely* members of the human race. So we should consider the following as another conceptual core fact about personhood:

(3) It is part of the concept of a person that persons are distinctively **valuable**.

And since there is so much about us which can be considered specifically and distinctively valuable, this again may explain why the concept of a person is inexhaustibly rich.

So much about my rather sketchy attempt at a diagnosis of what is vexing about the concept of a person and what may account for its characteristic inexhaustibility. – Now what would have to be done in order to tidy up a bit the conceptual mess? I call it a mess, because (as a result of its richness) we have too many 'intuitions' about personhood and almost nothing to give them structure. There are too few universally accepted constraints on this concept in order to make it possible to accept some of our (allegedly) *a priori* assumptions as valid and central and others as questionable or peripheral. I suspect that something like conceptual *analysis* is not what we need in order "to regain a complex concept of human personhood". If we just stare at the concept and brood over its richness, we will drown in a bottomless pit. Rather something like conceptual construction, or re-construction, is needed. And for this purpose it is necessary to get clear about what theoretical work we want the concept of a person to do. (Think back to the two examples given above: Descartes knew what theoretical aim he was after. He tried to solve the problem: Given that in reality mind and body are categorically distinct, how come we do not experience ourselves as consisting of two separate entities? He employed the concept of a person in his attempt to answer this specific question. - Locke tried to solve a different problem: What is the appropriate subject of punishment and reward? He used the concept of a person for this particular theoretical purpose. Both thinkers had quite determinate ideas of what the concept of a person was supposed to effect within their theories; and this allowed them to attach a determinate sense to it.) So my suggestion is this: Only if we get clear about what kind of theory we are striving for, and what role the concept of a person is supposed to play in it, we can get, or regain, a less vexing concept of personhood.

A final *caveat*. In this theoretical, constructive endeavour of getting clearer about personhood, we ought not to expect much help from the natural sciences. The best we can hope for is corrective cooperation. The natural scientist may warn us, for example, that given a certain conception of personhood, persons so conceived could not be members of the natural world. But we must not forget that from a strictly naturalist point of view, *person* is just not a category. (*Nota bene*, this is not to say that personhood cannot be accounted for in a naturalist way. David Lewis, for example, has presented an ingenious naturalist account of personhood and transtemporal personal identity, which is based on a psychological concept of person.³⁰) The crucial point here is this: We would have to have reached, independently, considerable conceptual clarity about personhood, before we could reasonably hope for a naturalist

³⁰ David Lewis, "Survival and Identity", (1976), reprinted with a postscript in: D. Lewis, *Philosophical Papers*, vol.1, Oxford UP (1983), 55 – 77.

characterization of the entities which exemplify it. We cannot ask the natural scientist "What is a person?", in the same state of almost complete conceptual ignorance and with the same hope for conceptual elucidation, in which we may ask: "What is a magnetic moment?".

Natural science can teach us what human beings (considered exclusively as members of a certain biological species) are, what storks, computers and, if there are or were any, of extraterrestrials and replicants - science can inform us about their physical and functional similarities and differences. But we must not hope that among the distinctions drawable in naturalist terms, there is one --already drawn, as it were-- between those human beings, storks, computers, extraterrestrials and replicants which (or who) are persons and those which (or who) are not. This would be silly. The natural sciences, with good reason, attach importance to providing no methodological space for value-concepts. This, of course, is not a frivolous narrow-mindedness on their part but a well-considered delimitation of what does, and what does not, fall within their cognitive realm.³¹

³¹ Thanks to the members of the group, especially to Maria Antonaccio, Philip Clayton, Malcolm Jeeves, Eiichi Katayanagi, John Polkinghorne, William Schweiker, Günter Thomas and Michael Welker. In developing my ideas about these issues, I have profited a lot from discussing and conversing with them. Many thanks also to my old friends Mark Helme, Rolf-Peter Horstmann, Rainer von Savigny and Hans-Peter Schütt for support and encouragement. They were kind enough to read earlier versions and spotted several things they found flatly mistaken or just cranky, some of which I have tried to correct or tone down.