abstract

Empirically minded and naturalistically inclined post-Cartesian philosophers have refused to accept the idea that we human persons are immaterial, senseless souls. This rejection has led to a fragmentation of the self and eventually to its theoretical disappearence. A way to resist this eliminativist trend is to see the self as an embodied entity, a promising thesis which has assumed prominence in contemporary debates. The paper is a (fairly partisan) reconstruction of this post-Cartesian scenario.

keywords

Body, soul, person, no-self theories
1. From a Senseless Self to a Selfless World

1.1. I Have no Hands

“I will think that the sky, the air, the earth, colours, shapes, sounds, and all the external things are no different from the illusions of our dreams […] I will consider myself as having no hands, no eyes, no flesh, no blood, and no senses, but yet as falsely believing that I have all these”.

This is, of course, a well-known passage from Descartes’ Meditations: an evil spirit is perhaps deceiving the author – and the reader, and all of us: we have no hands, no senses, no body, and ‘our little life is rounded with a sleep’.

But there is no evil spirit, and the sky, the air, and the bodies are there where we think they are, created by a perfectly good God. This is the conclusion that, eventually, the Meditations come to. And yet, for Descartes, there is a clear sense, the strictest sense, according to which we indeed have no body and no senses, neither flesh nor blood. “I think in myself faculties of thinking in various specific ways – namely the faculties of imagination and sensation – without which I can understand myself clearly and distinctly as a whole”: imagination and sensations do not belong to the whole of me, as it were: I have sensations and I see images of various sorts, but these ‘sensuous thoughts’ are generated by faculties that do not exist in myself. I myself do not have the passive faculty of sensation, and I am not a sensible thing at all: I am an ‘intellectual’ substance, a soul or a mind, which is really distinct from the body which I call mine.

This is not to say that we, human persons are not strictly united to a material body and that emotions and sensations are not a pervasive part of our lives. On the contrary, Descartes says, the passions of the soul, such as perceptions, sensual love, hatred, wonder, melancholy, joy, desire and various bodily feelings are everywhere; and for some it can even be difficult to conceive of ourselves as devoid of these kinds of experience which would seem to be the actual constituents of our very humanity.

But not so. A human person is identical with an immaterial soul, and, for example, it is ‘rational love’ alone that properly belongs to it, a ‘pure’ form of love which is not generated by any bodily movement. Descartes’ real self, the mind/soul, is therefore completely outside the physical order of nature and it is strikingly different, as a metaphysical principle, from ‘aristotelian souls’. It is a pure immaterial and rational substance and not the form of the body, which – according to aristotelians – bestows on the body itself not just the power of rational thinking but also its vegetative and animal life.

---

1 Descartes (2008), p. 16.
2 Descartes (2008), p. 56.
Many scholarly caution notwithstanding, it is indeed quite difficult to avoid the notorious image of the ghost in the machine when one thinks of Descartes’ selves.

Now, there have been, in post-Descartes philosophy, many ghost-friendly thinkers, happy to admit immaterial spirits – or perhaps transcendental egos – and to identify us with them; these ‘idealistic’ philosophers will not be my concern here.

I will instead deal with the empiricist and naturalist tradition which felt, and feels, uncomfortable with immaterial souls and the like: Cartesian – or Kantian – selves are too rarefied, too thin, to content down-to-earth temperaments; and yet, if human persons are not simple, senseless souls, what are they? If not souls, and not bodies – Descartes’ ban has been a powerful one – well, perhaps nothing at all. And indeed, in the naturalistic field, so the story has gone.

1.2. Nothing but Consciousness, and Perhaps Nothing at All

According to Descartes, we are immaterial thinking substances and we persist through time exactly when the same immaterial substance exists at different times. Locke disagreed.

That is not to say that according to Locke thinking does not concern human persons; on the contrary, he famously wrote that “a person is a thinking intelligent being that has reason and reflection and can consider itself as itself, the same thinking thing, in different times and places; which it does only by that consciousness which is inseparable from thinking, and, as it seems to me, essential to it”. ³

So, according to this picture, there are thinking things and these are persons. But Locke, the empiricist, is not able to establish the nature of these thinking substances: “We have the ideas of matter and thinking, but possibly shall never be able to know, whether any mere material being thinks or no; it being impossible for us, by the contemplation of our own ideas, without revelation, to discover whether omnipotence has given to some systems of matter, fitly disposed, a power to perceive and think, or else, joined and fixed to matter so disposed, a thinking immaterial substance”. ⁴

So, as far as knowledge is concerned, thought and consciousness are properties of substances whose material or spiritual nature is beyond our reach. And yet, it would seem, we are identical with these mysterious subjects of experience. But, somewhat surprisingly and somewhat incoherently, Locke is not of this opinion: these subjects of experience cannot be identical with human persons because they have different identity conditions. Suppose ‘the bearer’ of my thinking is actually a soul: according to Locke, I could acquire a new soul and my soul could become the soul of someone else: our identity as person does not depend upon the identity of any thinking substance. Sameness of substance, Locke says, does not concern personal identity at all: the same substance without the same consciousness no more makes the same person: “Consciousness alone makes self - Nothing but consciousness can unite remote existence into the same person: the identity of substance will not do it”. ⁵

So we are something like a continuous stream of consciousness which, at least in principle, could be transferred from one thinking substance to another. We belong, as it were, to a pure realm of ideas, and are quite separate from the substances that we happen to inhabit.

According to Locke, one cannot know whether Cartesian souls do exist, but this does not threaten the knowledge we have of ourselves: we are, one could almost say, our own experiences, and these are certainly not beyond our epistemic possibilities. The Cartesian ghost is separate from ourselves, and we are identified with a non-substantial stream of ideas.

But are we really a stream of consciousness? Not exactly, said Hume: what is there consists, more properly, of a flux. “Setting aside some metaphysicians […] I may venture to affirm of the rest of mankind that they are nothing but a bundle or collection of different perceptions, which succeed each other with an inconceivable rapidity and are in perpetual flux and movement”. ⁶

³ Locke (1975), p. 335.
⁵ Locke (1975), p. 344.
It is worth noting that this Humean flux is composed of many particular mental states which seem to be substances on their own, capable of separate existence and in no need of a material body or an immaterial soul as their ‘support’. “All our particular perceptions [...] are different, and distinguishable, and separable from each other, and may be separately considered, and may exist separately, and have no need of any thing to support their existence”.

The mind, Hume says, is constituted by many different perceptions and there is no place for a simple, enduring individual to whom these perception belong - even though there is a clear and inevitable propensity to feign the continued existence of an invariable substance: myself. So one could say that Hume goes further than Locke in denying the very existence of a thinking, though mysterious, substance; but, nonetheless, in the Humean picture we do exist and in a somewhat Lockean form: at every moment of our personal lives there is a certain collection of perception and this collection is nothing but a stage in the sparkling flux of mental events that is to be identified with a person persisting in time. Not quite so, however.

According to Hume it is a mistake to think that an object can gain or lose a part. Gaining a part brings something new into existence, losing a part brings the object under consideration to an end, and one should not confuse the relation of strict identity with the relation of similarity, which holds between successive objects ‘born’ after the loss or the gain of a part by a previously existing object. So the collection of perceptions that exists at one time is never identical with the collection of perceptions that exists one moment later: in a flux, many perceptions change, and so there cannot be strict identity. If a person has to be something capable of persistence over time, then, it seems, one should conclude that there are no persons after all.

And the idea suggests itself that there are not even instantaneous persons. Indeed, in the Humean picture, a person existing at a certain time should probably be viewed as a collection of perceptions existing at that time. But even that is problematic: Hume is quite clear in saying that these collections are bundles of separate items, with no real bond among them. So if an instantaneous person is something endowed with a unitary mental life, it seems that one should conclude that there are not even instantaneous persons.

Hume’s Treatise, I think, directly suggests these radical conclusions, but we should note that Hume himself did not unambiguously state these unsettling ideas. Contemporary Humeans have not been so cautious.

2.1. The Way of Cognitive Science

“The no self alternative” is the telling title of a recent paper by the neuropsychologist and philosopher Thomas Metzinger; and it is no surprise to find in it an approving mention of Hume’s ‘bundle theory of the self’.

Metzinger considers Hume’s position as a prominent example of a non-substantialist approach to the nature of the self, and according to him denying the substantiality of the self is no different from denying its very existence. I think the last point is far from obvious, but let us concede it for present purposes.

Metzinger thinks that Hume’s model of the self, indeed Hume’s elimination of the self is on the right track and that present-day cognitive neuroscience offers an empirically adequate account and a vindication of these old philosophical intuitions. According to him there is a common experience of something like a self-subsistent enduring entity, a self, forming a non-exchangeable and irreducible part of the world, a ‘nugget of reality’. But this ‘phenomenology of substantiality’, by itself, has no metaphysical value, and the principle of parsimony demands a deflationist explanation of such a phenomenology, an explanation that, following Hume, accounts for our tendency to feign a fictional and stable character behind the flux of thoughts. Contemporary cognitive science, Metzinger holds, offers such an explanation and such an account.

8 Metzinger (2011), p. 282. The core ideas of Metzinger’s two books, Being No-one (Metzinger 2004) and The Ego Tunnel (Metzinger 2009), are usefully summarized in this article.
“All technical details aside [...] science offers conceptually clear models of functional mechanisms which could parsimoniously explain the integration of individual property-representations into a unified self-representation. This theoretical model requires no transcendental subject to stand behind the appearance of a ‘self’ as consciously represented, because it gradually emerges out of the self-organizing interaction between a large number of simpler components. This possibility [...] simply was not available to thinkers in the past, it is a novelty in the history of ideas. [...] dynamical self-organization is a new theoretical option for the bundle theorist”.

There is a representation of a substantial self, Metzinger claims, but not the substantial self itself: nothing in the available scientific data obliges us to posit persons as unchangeable bearers of mental states – and Metzinger, a paradigmatic naturaliser, is implicitly suggesting that science has to be the measure of all things.

Such an attitude is by no means rare in the contemporary science of mind, and one of its well-known champions is certainly Daniel Dennett, who, of course, thinks there are no persons at all.

The centrepiece of Dennett’s eliminativist strategy is a theory of the mind based on a sustained and empirically based critique of the so-called ‘myth of the Cartesian theatre’ – empirically minded scholars very much like pick on Descartes.

Neuro-cognitive sciences, Dennett believes, have finally dispelled this ancient myth, namely the idea of an ‘interior boss’, the controller of the body and the privileged, Cartesian viewer of all mental states which play their roles on the stage of our mental theatre. This tenacious myth still has a strong persuasive force, but contemporary cognitive sciences have, at last, allowed us the resources to free ourselves from it. What they teach us is that the brain’s processes are parallel and distributed ones, and there is no place in the brain where it all comes together. What happens is that some processed bits of information sometimes gain a more or less stable ‘cerebral celebrity’: temporary, in-the-limelight goings-on which have no guarantee of keeping their privileged position in the Humean flux of the mind/brain’s activities. These temporary ‘celebrities’ constitute a sort of linear order, a brain narrative whose main character is what Dennett calls “the virtual captain”. And this character is represented by the narrative as a substantive Cartesian boss, but really it is nothing; its illusory existence is just the product of the brain’s impersonal processes. “If asked what a centre of gravity was made of [physicists] would say, ‘nothing’”;

and if one asked what a self is, one should reply as physicists do in the case of physical centres of gravity: “nothing at all”. But if the self is nothing but a centre of narrative gravity – and so, Dennett says, nothing at all –, what about what we think of as our introspected selves? From Dennett’s perspective, clearly, no substantial self is the object of a special interior faculty called “introspection”; there is just a usefully deceptive mental representation of a unitary subject, allegedly responsible for the actions of complex cognitive systems – a misleading and yet convenient illusion, and nothing more.

The main points of the theory just outlined are summed up in this notable passage, which introduces some more powerful metaphors, the very trademark of Dennett’s style of thinking: “In our brains there is a cobbled-together collection of specialist brain circuits, which thanks to a family of habits inculcated partly by culture and partly by individual self-exploration, conspire together to produce a more or less orderly, more or less effective, more or less well-designed virtual machine, the Joycean machine. [...] this virtual machine, this software of the brain, performs a sort of internal political miracle: it creates a virtual captain of the crew”.

2.2. The Way of Philosophy

Contemporary cognitive science, we have just seen, has lent us empirical arguments for the theoretical elimination of subjects from the book of the world. But the Humean attitude has been alive, and indeed

dominant, even in the strictly philosophical field, or at least in the philosophical tradition which has been closer to empirical sciences, namely analytic philosophy. Indeed, in the last forty years or so, Derek Parfit’s ideas – in which there are distinct echoes of Hume – have been a true landmark for every analytic philosopher trying to make sense of the puzzles of personal identity.

As is well-known, Parfit is a very sophisticated thinker and the intricacy of his discussions is not always open to a straightforward interpretation. Nevertheless an eliminativist reading of his ideas is quite easy, and I think quite correct.

According to Parfit we use the language of personal identity when we think there is an appropriate psychological continuity: we say, for example, that the person who is now writing this paper is the same person who was writing it ten minutes ago because between the mental lives of these ‘two’ persons there is a relation of psychological continuity.

Now consider the following case, concerning three identical twins. "My Division. My body is fatally injured, as are the brains of my two brothers. My brain is divided, and each half is successfully transplanted into the body of one of my brothers. Each of the resulting people believes that he is me, seems to remember living my life, has my character, and is in every other way psychologically continuous with me. And he has a body that is very like mine", 12

What happens to me after the transplant? There seem to be only three possibilities.

(1) Given that the criteria of personal identity we actually use are grounded in psychological continuity, it seems that I should be identical with at least one of the two persons – let us say A and B – existing after the surgery. But my psychological continuity with A is the same as that I have with B. Why should I be one of the two, and not the other?

(2) One could then think that I am identical with both A and B; but if so, since A and B are different, A and B would be at the same time identical and different persons, which is of course impossible.

So (3) perhaps I have not survived the transplant. And yet, if half of my brain had been successfully transplanted and the other half destroyed, I would have survived, I would have been A for example. But, if this is so, how is it possible that the mere existence of B prevents me from existing? A double success cannot be such a failure: if creating A grants my survival, why on earth should the creation of A and B cause my death?

There is no way out, it seems, even though the question concerning my survival and identity after the transplant is perfectly clear and, plausibly, it should have a perfectly determinate answer. Parfit disagrees on the last point: "we are naturally inclined to believe that our identity must always be determinate. We are inclined to believe, strongly, that this must be so. I [...] argue that this natural belief cannot be true unless we are separately existing entities" 13 namely unless we are something like a Cartesian soul.

Therefore, if there are no Cartesian souls – as Parfit thinks –, then there will be cases in which questions about our persistence over time do not have a determinate answer. The ‘my division situation’ is one of these cases. "We know what this outcome [i.e.: the transplant’s outcome] is. There will be two future people, each of whom will have the body of one of my brothers, and will be fully psychologically continuous with me, because he has half of my brain. Knowing this, we know everything. I may ask: ‘But shall I be one of these two people, or the other, or neither?’ But I should regard this as an empty question" 14

So, according to Parfit, even an omniscient being could not give a definite answer to the question concerning my survival in the transplant case, and this is so because there is no fact of the matter to be known. This is, indeed, a quite bewildering idea because it suggests that the world itself could be indeterminate with regard to my existence. But things are not exactly like this, I think.

---

As I understand what Parfit is saying, the world cannot be indeterminate with regard to my existence and this is because, in a strict and literal sense, there is no such thing as myself. The whole of reality is completely describable in an impersonal way, and nothing in this description can be identified with a person. To be sure, one talks about persons and their identity over time, and there are certain criteria of correctness for such talk. In many cases it is possible, at least in principle, to talk about persons and their persistence through time in a correct and determinate way; but there are cases – such as the ‘my division situation’ – that are “not covered by the criteria of personal identity that we actually use”.\footnote{Parfit (1971), p. 3.}

In such cases, Parfit suggests, questions concerning the continued existence of people are empty and so unanswerable, and we should simply abandon the language of identity. Yet the world, a selfless world, is perfectly determined. “Buddha would have agreed”, Parfit writes,\footnote{Parfit (1984), p. 272.} and referring to a classical buddhist source he approvingly quotes: “O Brethren, actions do exist, and also their consequences, but the person that acts does not”\footnote{Parfit (1984), p. 272.}.\footnote{This assumption is indeed explicit in Metzinger (2011), p. 281.}

At this point one noteworthy thing has to be underlined. The contemporary eliminativist theories we have discussed so far seem to share an implicit assumption: a human person, if it exists, has a ‘purely mental’ nature and, in particular, it cannot be identified with a living body.\footnote{Among the ‘constitutionalists’ one may recall Baker (2000, 2007), Corcoran (2006), Johnston (1987, 1997), and Shoemaker (1999).} In this respect, Hume’s heirs agree with a fundamental tenet of Cartesian philosophy: the self is not a bodily self, if it exists at all - and Metzinger, Dennett and Parfit say it does not. But perhaps one could avoid the disappearpance of people, of ourselves, simply by restoring our forgotten material dimension. Let us look at two recent ways of doing exactly this.

### 3. Embodying the Self: Two Notable Attempts

#### 3.1. Mind and Body: The Quasi-Identity Theory

A human person cannot be identical with a living body, or so many philosophers think, offering an argument along the following lines. Suppose I am indeed identical with my living body; so I should exist exactly where and when my body exists as a living organism. (But if my brain is seriously and irreparably damaged, my body can still exist although there is no longer a person: I myself have disappeared. A body does not essentially have a mental life; I do. So I, the person, am not identical with my body. And yet I am intimately related to a body: indeed, when my brain is irreparably damaged I disappear. So what relationship might there be between me and my body?

Constitution theorists have a surprising answer: a human person is constituted by a human body to which the person itself is not identical. Here is the usual – and useful – analogy: a piece of marble constitutes a statue but the statue is not identical with it; in fact, if they were identical, they would have the same persistence conditions: neither could exist without the other; but they have different persistence conditions: the piece of marble can change its form and can exist without constituting a statue, and the statue can change many parts and can exist without being constituted by the original piece of marble. So a statue and the piece of marble that constitutes it are not identical. And the same holds for my body and myself. When a piece of marble is suitably shaped, and perhaps is suitably related to an artworld, a new thing, a statue, comes into existence. When a human organism comes to a suitable degree of biological development it constitutes a new thing, a person, which is essentially endowed with a sophisticated mental life. And it can happen that, when the constituting organism is seriously damaged, the person disappears, no longer constituted by a suitable organism.

The picture that emerges is therefore quite peculiar. On the one hand, persons are essentially thinking beings distinct from any material body. On the other hand, however, human organisms constitute persons who are intimately related to material bodies, so intimately that one can correctly assign to
a human person a spatial location, just as we do with statues. Constitution theorists look for a duality without dualism, as it were, a fascinating and seemingly impossible compromise between our rational nature and our bodily roots. This sounds like a miracle and indeed it has attracted its share of (rational) disbelievers.

Consider, to begin with, the mental life of a human person. This is the mental life of an entity which is distinct from the organism that constitutes it and so, it seems, the organism itself, in the strictest sense, has no mental life. But why should this be so? A fully developed human organism seems able to think in every respect: it appears to have all the thought-enabling features one would expect; and yet, the constitution theory says, these features are just sufficient conditions for constituting a different entity which is the unique owner of a personal mental life. This sounds quite implausible. On the other hand, if one concedes that human organisms are indeed able to think, then we have two distinct thinkers, the organism and the person it constitutes. And this is even worse. So the constitution theorist seems to be forced to defend the odd thesis according to which a fully developed human organism is unable to think.

But now consider non-human animals, such as cats, dogs, cows, and so forth. Everybody agrees that these animals, these living organisms, are indeed endowed with a mental life, although (probably) not the mental life of a person. So there are thinking organisms after all, and excluding fully developed human organisms from thinking seems a rather dubious move.

At this point one could perhaps suppose that human organisms are indeed able to think: they simply do not have personal mental lives, but nevertheless they do have perceptions, feelings, desires, and possibly an elementary ability to reason. If so, however, what would the relationship between these animal thoughts and the mental life of the person constituted by the animal be? Should we say that humans have a ‘broken’ mental life – perceptions and feelings belonging to the animal and, say, theorising belonging to the person? Or should we say that when the animal constitutes a person its mental life is ‘transferred’ in some way to the person? These are not exactly promising ideas.

But even setting aside this metaphysical maze, a fundamental question remains to be answered: when, why and in which cases something constitutes something else?

3.2. Animals

Constitutionalists say I am not an organism, even though I am intimately related with the human animal that constitutes me: a fascinating view, to be sure, but also a perplexing one. So one could maybe prefer an obvious alternative: I am not constituted by a human animal, I am simply identical with it. This is what so-called ‘animalists’ maintain.

Are all persons identical with human animals according to this view? Not necessarily. If God, angels, and a certain kind of alien do exist, they are persons but not human animals. And notice: biologically alive human organisms in a permanent vegetative state, whose mental capacities are permanently destroyed, are certainly human animals but not persons – if a person is to have certain mental qualities such as rationality, self-consciousness, memory and so on. So not every human organism is a person, animalists say.

This point, however, seem very puzzling indeed. The human organism to which I am identical could exist without being a person – if, for example, it enters a permanent vegetative state; so I could exist without being a person; but this, one would think, is certainly absurd: whenever I exist I am a person,

---

20 One can find a sustained critique of the constitution theory in Olson (2007).
I am essentially a person; how could animalists deny this? They can. Animalism is precisely the view according to which people like you and me are identical with thinking organisms and therefore we are not essentially thinking beings; just as a former president of the United States can exist without being a president, so you and I can exist without a personal mental life and without being a person. “But”, so the protest goes, “surely I am not a mere animal!”: Human persons cannot be animals, it is said, because in many respects they are absolutely special beings. They compose quartets, have complicated love affairs, write novels and are rational and morally responsible beings – they can be so, at least. How could an animal be like that? This line of argument I think – and animalists think so as well – is rhetorically quite strong but not exactly a very compelling one. We may simply be very peculiar animals, and nothing more: our being special – an undeniable feature of us – does not call for a special metaphysical status. And indeed in an unreflective mood we seem obviously inclined to admit that we are animals; we ordinarily say that we have hands, we are sitting on the chair, we get fat and so on and so forth; but, of course, human animals too have hands, sit on chairs and get fat: it seems no coincidence. Animalism, after all, seems like a sensible and down-to-earth idea.

And yet consider this case. Your cerebrum is put into another head and the being who gets that organ is mentally continuous with you: she has your memory, your feelings, your tastes, and so forth. She is you, one would say. So you have gone along with your transplanted cerebrum while no animal has moved from one head to another. Therefore you are not an animal.

Many people, and many philosophers have found this argument inescapable: animalism must be false. But animalists have an answer. In ordinary life we are mainly interested in the mental lives of people and, ordinarily, whenever there is the same mental life, there is the same person. But this does not hold on every occasion. In extraordinary cases, such as the cerebrum transplant, a mental life is transferred from a person to another: there are one mental life and two persons. Metaphysically speaking, a kidney transplant is no different, even though, of course, the cerebrum case makes for a dramatic practical difference (our interest will shift from the donor to the recipient).

I think this is a quite convincing reply, and, more generally, I think animalism is quite a convincing idea. If you are skeptical about Cartesian souls and transcendental egos, and if you want to avoid our theoretical disappearance, you’d do best to admit we are robustly embodied things. And the most straightforward and promising thesis concerning our bodily nature is that we are identical with living bodies. Or so I suggest, if you trust an animal like me.

---

22 See Olson (1997).
References
Descartes, R. (2008), Meditations on First Philosophy, edited by M. Moriarty, Oxford University Press, Oxford. (Original work published 1641);