

Essay Self: East and West HPPH7013

Module Tutor Prof. David Cockburn

Department of Philosophy, University of Wales Lampeter

November 2020

By Emmeke Vierhout

What does the word “I” refer to?

Introduction

What does the word “I” refer to seems like a very straightforward and simple question, however when trying to give a straightforward and simple answer, you find out that this question contains a manifold of sub questions. And in my search for finding an answer I was drawn into this swamp of questions concerning concepts like “Personal Identity”, “Self” and “Soul”, “Mind and Body” and so on. The way to get out of this swamp, I will argue, is to stop seeing “I” as a referring expression. Even though it might seem in some uses to *be* a referring expression, this is not really the case. The meaning of the word “I” lies in its use, its use indicates the presence of a Self ; a Self being a biological organism which has self-consciousness, can consider itself as ‘itself’ *and* has agency.

The use of the word "I"

In her essay “The First Person” Elizabeth Anscombe (1975) says the word “I” does not refer to anything because

'The suggestion of getting the object right collapses into absurdity when we work it out and try to describe how getting hold of the object may be excluded.'...

and

'...there is no getting hold of an object at all. With names, or denoting expressions (in Russell's sense) there are two things to grasp: the kind of use and what to apply them to from time to time. With "I" there is only the use.'

So according to Anscombe we need to get rid of the assumption that "I" is a referring expression.

'When a man say does not know his identity, has, as we say, "lost his memory", what he doesn't know is usually that that person he'd point to in pointing to himself ... is say, Smith, a man of such-and-such a background. He has neither lost the use of "I", nor would he point to an unexpected body, to a stone, a horse, or another man...'

What Anscombe is saying is that it is perfectly possible for someone who has lost his memory and thus has no idea who he is, to use the word "I" in a meaningful way. I believe it is in the same way as someone who is completely lost in the dark can say that he is "here" without knowing where that is. Someone using the word "here" would not point to some place other than where he is, he has not lost the use of the word "here".

But if it is really this simple, why do we tend to think of "I" as a referring expression? Sometimes it obviously seems like one. The reason for this I believe is *the idea* (suggested by Wittgenstein Ref) that there is a distinction between the use of the word "I" 'as object' and its use 'as subject'. Human beings, are self-conscious beings, meaning that they have the capacity to experience themselves 'as themselves', which is an experience of themselves in an immediate way, not as an object (like when we have the experience of seeing a rock) but as the subject of experience. This experience as subject is immune to error through misidentification, simply because there is no identification taking place.

But what about this idea of the use of "I" as object? Consider sentences like 'I am the daughter of H.V', or 'I am in the kitchen'. In this type of sentences the word "I" does seem to refer to some 'object', however they don't really. In this type of sentences there is identification involved, but it works like a two-stage missile because the use of "I" presupposes identification before or after the use. To understand what I mean, let us look again at the analogy with the word "here". Let's

say that I am in the kitchen and my daughter shouts ‘Mummy where are you?’ When I answer ‘I am here’ the use of here is not really a referring expression about my location. If I would give this answer, I presuppose that she either already knows where “here” is, because she is also in the kitchen or that she can hear from the direction my voice is coming from, that I am in the kitchen. If she still has no idea where I am, it makes perfect sense for her to ask ‘where is here?’ The word “here” is not used as an indication of where I really am, it is not a referring expression. Now say that I come home and say ‘Hello! I am home!’ it would make perfect sense for someone who hears me but cannot see me, to ask ‘who is I?’. In using the word “I”, I presuppose that whoever hears me knows the identity of the user of “I” or is able to identify who I am through for example the sound of my voice. As with the word “here”, the word “I” is not used as a referring expression. In cases where the use of “I” creates the suggestion of being a referring expression what is really happening is that the use of “I” is connected with an identity proposition (who this user of “I” is) but it isn't an identity proposition itself. The word “I” is still being used as ‘subject’; in saying “I am in the kitchen” the user of the word might be wrong in thinking that she is in the kitchen, and might even be wrong about “who she is”, however she cannot be wrong in her use of the word “I”, her identity is not implied by the use of the word “I”. This is what Anscombe is talking about when she says that a man could lose his memory without losing the use of “I”.

But what about when I say ‘I have a broken arm’ while I mistakenly take my neighbors arm to be mine? How can “I” not be an identity proposition (me, this body) in this case? Wittgenstein (REF) uses these kinds of examples to argue for different uses of the word “I” mentioned above: the use of the word “I”, as “object” and use as of “I” as “subject”.

‘[T]he cases of the first category involve the recognition of a particular person, and there is in these cases the possibility of an error’ (REF).

While we already know that in the second category there is no room for error. Wittgenstein argues that because most people fail to recognize this distinction, most people believe that mind and body are two separate substances. I think that Wittgenstein is absolutely right in this conclusion, however I believe that Wittgenstein is more or less making the same mistake when he takes “I” to be a referring expression, either as “subject” or as “object”.

When Wittgenstein says

'The word "I" does not mean the same as "L. W." even if I am L. W., nor does it mean the same as the expression "the person who is now speaking". But that doesn't mean: that "L. W." and "I" mean different things. All it means is that these words are different instruments in our language.'
(REF)

and

'When we feel that we wish to abolish the "I" in "I have pain", one may say that we tend to make the verbal expression of pain similar to the expression by moaning.--We are inclined to forget that it is the particular use of a word only which gives the word its meaning.'

he fails to take the final step, which is the step Anscombe does take which is that "I" is in its use never is a referring expression, there is *only* the use of "I" as subject.

To understand what I mean we might look at the most famous philosophical argument for the separation of mind and body. Descartes argues that the "I" that is thinking 'I think' is a thinking substance ontologically different and independent from anything material. He comes to this conclusion through a method of doubt. Descartes claims that he can doubt everything that exists even his own body, but that he can't doubt his own existence when he can think 'I think':

'Accordingly this "I" - that is, the soul by which I am what I am - is entirely distinct from the body, and indeed is easier to know than the body, and would not fail to be whatever it is, even if the body did not exist.' (Rene Descartes, Discourse on Method, Part IV, paragraphs 1 and 2)

Wittgenstein would say that Descartes is confusing the use of the word "I" as subject ('I think') with the use of the word "I" as object ('I am this body'). The existence of the latter he can doubt, while the existence of the former he can't, but we need not mistakenly assume that "I" therefore cannot refer to both body and mind at the same time. I don't think that this is the mistake Descartes is really making in his Cogito argument. If "I" really was a referring expression then this is indeed confusing, how could "I" refer to something whose existence I can doubt and cannot doubt at the same time? But if you take "I" not as a referring expression than you get: 'I think, therefore there is a thinker, which I am'. In trying to find out what "this thinker" is, Descartes identifies this thinker with a purely immaterial thinking thing. He is identifying an object and in identifying this object one could be wrong. What Wittgenstein sees as the use of "I" as object is as I see it the sum of 'the use of "I" as subject' + 'identification of the user'. The

meaning of “I” lies not in its reference, but in its user; the word “I” can only meaningfully be used in the subjective sense; it can only be used in this experience of ‘the user’ in an immediate way.

Descartes identifies this user of “I” with thinking substance united in a soul. For Descartes, only thinking belongs to the user and nothing else. I think Descartes is wrong and that the user of “I” could potentially be any system that has self-consciousness and that can consider itself as ‘itself’. As I will argue below, any physical system with the capacity for self-consciousness is in potential what we call a Self. And only Selves can be a user of the word “I”. In practice however, so far we only know human beings to use the word “I”, so to stick to human beings, a *competent* user of the word “I” is the entire biological organism that a human being is which thinks, walks, yawns and so on, but crucially, which has the capacity for self-consciousness and can consider itself as ‘itself’. These human beings are what we refer to as Selves.

In answer to the question of this essay: the use of the word “I” does not refer to anything, it indicates the presence of self-consciousness, it indicates a Self as user. But what is the nature of a Self? I realize that saying that it is the entire biological organism, which is the human being, cannot go without argument.

Part II

The Self as biological organism

The biological organism

A biological or living system is an autopoietic entity which is constituted by dynamical self-organization and self-production (Maturana and Varela, 1991). We define dogs, cats, trees, spiders and so on as living systems that possess relatively stable borders with the environment, which are dependent on systematic exchange with the environment. In the natural sciences human beings are defined as living systems as well. Through the process of evolution, some biological systems have developed into cognitive conscious systems so that these can sense and perceive its environment actively process the information and in respond to that act on its environment. If such a cognitive conscious system develops into such a cognitive system that has the ability for self-representation and to combine such representations with conscious experiences, then we have an organism that has self-consciousness (Newen 2018). As I already

argued above a self-conscious system has self-representation about itself 'as itself', which is a representation of itself in an immediate way, not as an object but as the subject of perception and action or thinking. Such a system, in this case a human being, has the capacity to not only combine self-representations with consciousness at a point in time, but can combine them over time uniting them into a transtemporal Self.

From an evolutionary perspective the capacity for being a transtemporal Self was adaptive for human beings; when we look at our evolutionary close relatives, we can find in them a proto-capacity for the full-fledged human self-consciousness, primates like chimpanzees are found to have the capacity for self-awareness, they can recognize themselves in mirrors (Gallup 1979). I think it makes sense to suggest that self-awareness enhances social intelligence, 'an organism that can become the object of its own attention, in the sense of being aware of being aware, finds itself in the novel position of being able to make inferences about comparable states of awareness in others' (Gallup 1996). And indeed, primates like chimpanzees and bonobos are not only found to have the capacity for self-awareness but we know now that they also have theory of mind concerning other chimpanzees and show moral behavior without obvious short-term self-interest. (De Waal 2006) It is not farfetched to assume from an evolutionary perspective that these capacities already found in our evolutionary cousins have through evolutionary processes developed further in human beings (and probably also in other subspecies of the genus *homo* like the Neanderthals)

The embodied Self

Everything I have said so far about how certain types of biological organisms have, through evolution, come to be self-conscious human beings could be agreed both by those who argue that even though we call these beings *self*-conscious beings there is no such thing as a Self, for them the existence of a Self is an illusion (Metzinger 2009), as by those who argue that Selves exist and are *separate* entities from the biological organism, they are not embodied. Because although we might agree that the biological organism plays a crucial role in the development of a Self, this does not mean that this Self, once developed, is then necessarily part of this biological organism. I however will argue that Selves do exist *and* that they are in the human case, necessarily part of the human biological organism, in fact they *are* the human organism.

In a typical human being, self-consciousness results when a continuous flow of bodily information is combined with the cognitive, conscious system in the right way. It is an integration of all activated features in a situation, which involve a registration of my own bodily or affective processes (Newen 2018). Some of these registrations are typically unconscious like homeostatic features, but registration in self-consciousness always involves some conscious experience, some basic self-directed bodily or affective experience. The cognitive aspect of the human system combines and unifies these registrations into a Self, in the same way as the cognitive system combines conscious experience of a dog into a dog. The crucial element in self-consciousness is that in combining information about oneself into a perception of one-self, there the knowledge that the registrar *is* one-self. It is this *cognitive* consciousness of oneself *as* oneself.

Then what is this Self? The Self is nothing above and the self-conscious human being. This being is able to create a self-model as a unity of self-relating information. This model is formed by the brain but firmly anchored in the self-representing biological system which realizes this self-model. (Newen 2018) I would argue that any system that has self-consciousness is by definition a Self. The minimal condition for being a Self is that the system has self-representation about itself 'as itself', the latter being the creation of a self-model; a self-model can be realized in a single situation, which for example might be the case for a severely demented human being. However, when a healthy human being thinks about themselves as themselves, they typically take into account long-term features of themselves. A typical human Self is a consciously experienced, multi-feature-involving and transtemporal Self (Newen 2018). One important aspect of such a transtemporal human Self is the idea of being a person.

The notion of “person” is highly ambiguous and confusing because we can come to this concept from different perspectives and for different purposes. One angle of thinking about persons stems from the ideas of John Locke. His legacy can be found in the idea that continuity of memory and psychological features like character accounts for the sameness of person over time. One can be such a person from a first-person perspective (1PP) and from a third person perspective (3PP). To take the example of Amanda featuring in a documentary of Louis Theroux (2016). She fell of a horse and suffered serious brain damage. While the husband and children would sadly conclude that there was not much of Amanda or mummy 'left', Amanda would claim that she was still the same Amanda only they had stopped liking her. While her family thought she wasn't the same

person anymore, she herself claimed that she was the same person as before the fall. I guess we could agree that everyone from their own perspective could be right. An obvious clash of 1PP and 3PP is when someone who is declared braindead is not capable of seeing herself in 1PP as a person anymore, while her loved ones still see her as a person for example their daughter. But there is another angle to approach the notion of “person” from, which is the notion of “person” seen as a moral concept; in this sense the notion of a ‘person’ highlights a practical dimension concerned with rights and duties. It is in this way that we can attribute personhood to four-week-old fetuses who therewith gains special rights. At the same time, we can stop seeing human beings with minimal self-consciousness after brain damage for example as persons who are therewith refrained from having duties and should no longer be held accountable for their deeds. The moral notion of “person” is mostly used from a 3PP. For a theory of an embodied Self it is the notion of “person” from the 1PP in the Lockean sense that is relevant, it is this notion of ‘person’ that accounts for the long-term features of a Self. The notion of “person” is often used as distinct from a notion of “self”, however used like that they characterize different aspects of self-consciousness. The notion of “self” is often being used to refer to the cognitive features of the subject’s self-consciousness, while the notion of “person” is used in the Lockean sense. Both “self” and “person” are two aspects of a Self; they attribute both short-term and long-term mental properties to the same entity. But they are not the only aspects of the Self.

Other aspects being: the notion of ‘individuality’, seeing oneself as distinct from others; the notion of ‘ownership’, seeing oneself as the owner of one’s body parts, one’s thoughts and one’s feelings; the notion of ‘appearance’, having a sense of who you are on a picture or recognizing yourself in a mirror and so on; the notion of a ‘narrative’, we see ourselves as having a history in and with our environment. All these aspects of the Embodied Self are the result of a unification of various inputs through a system of dynamical self-organization which takes place in the biological system. We know that this process is extremely complicated and multifaceted but also very resilient and adaptive to changes in the environment. But this doesn’t mean that things can’t go wrong. We know they can and in fact we all have experiences in which one of the aspects of our self-model does not match reality, does not match the Self, when we think that our legs are crossed while they aren’t, or when we have memories of things that never happened. However, in most cases the consequences are not severe. But we all know about more serious cases where the model does not match the Self, like in cases of multiple personality, or cases where model

considers certain body-parts as not belonging to the Self. Arguably, transgender issues fall under such a mismatch of self-model and the Self as it is. Interestingly cases like transgender or the lack of feeling of ownership of certain body-parts are examples where the Self can be adjusted to match the self-model.

Research has also shown that some aspects of the self-model can easily be manipulated. There is the famous experiment with the rubber hand, in which someone can be manipulated in taking a rubber hand to be one's own hand. Even more extreme experiments show that someone can be tricked in thinking that a VR-body is its own body (Lenggenhager et al. 2007). It is important to realize however that these kinds of experiments are being done with human beings who actually are embodied beings, they are only fooled in taking a VR-body to be their own. It *isn't* a real change of body. Metzinger (2011) however uses these kinds of experiments to show that the Self is an illusion. These experiments show that there are only constructive brain processes which implement the body scheme and *the sense of agency*, and this is enough to deny the existence of a Self. Metzinger here is referring to another aspect of the Self that I haven't yet mentioned, which is the *sense of agency* or what we might refer to as free will. There is many scientific evidence that shows that is is not just a sense of body ownership that can easily be modified but many experiments have shown also that our sense of agency can rather easily be modified and these are reason for Metzinger to argue that the sense of agency are constructs of the brain and nothing more. I would argue against Metzinger that the most important aspects of a Self *is* this sense of agency and that this agency is real; it is what sets us apart from beings that aren't Selves, it sets us apart from beings that are entirely subject to the laws of nature.

Metzinger (2009) is a determinist in that in denying the existence of a Self, he denies that there is such a thing as true agency. For him there can only be causation from matter (in the human case the body and more specifically brains) to mind (consciousness), while free will in its true sense I believe is about mind having causal relationships with matter. And if you would want to account in any way for mental downward causation (from mind to matter), to account for a Self with agency, this would require the acknowledgment that physical determinism cannot describe everything. What such an account could look like lies beyond the scope of this essay, however, I hold that any credible theory of consciousness and its relation to matter needs to include the existence of true agency. The clue in such a theory may be found in the nature of self-

consciousness itself, because as Gallup remarks “Ultimately, an organism capable of contemplating its own existence may find itself in the peculiar and certainly precarious position of being able to take steps to modify that existence. (1979) This is the crucial element in being a Self, a Self has agency in this world and can alter the world. It may be a rearguard battle against the deterministic forces of evolution, but it is possible, and we see it happening every day. And it is only a Self in the true sense of the word that can use the word “I” in a meaningful way.

Bibliography:

Anscombe, E. (1977) "The First Person" chapter 2 in Guttenplan, S. D. Mind and language. Oxford: Clarendon Press.

Gallup, G. (1979). Self-Awareness in Primates: The sense of identity distinguishes man from most but perhaps not all other forms of life. *American Scientist*, 67(4), 417-421.

Gallup GG Jr. (1998) Self-awareness and the evolution of social intelligence. *Behav Processes*. 1998 Feb;42(2-3):239-47. doi: 10.1016/s0376-6357(97)00079-x. PMID: 24897465.

Lenggenhager, B., Tadi, T., Metzinger, T., and Blanke, O. (2007). Video ergo sum: manipulating bodily self-consciousness. *Science* 317, 1096–1099. doi: 10.1126/science.1143439

Maturana, H. R., and Varela, F. J. (1991). *Autopoiesis and Cognition: The Realization of the Living*. Dordrecht: Springer.

Metzinger, T. (2003). *Being no one*. Cambridge, MA: MIT Press.

Metzinger, T. (2009). *The Ego-Tunnel. The Science of the Mind and the Myth of the Self*. New York, NY: Basis Books.

Metzinger, T. (2011). "The no-self alternative," in *The Oxford Handbook of the Self*, ed. S. Gallagher (Oxford: Oxford University Press), 279–296.

Newen A. (2018) The Embodied Self. The Pattern Theory of Self, and the Predictive Mind. *Front. Psychology* 9:2270 doi: 10.3389/fpsyg.2018.02270

Theroux, L (2016) Documentary 'A Different Brain', BBC UK

Waal, D. F., Macedo, S., & Ober, J. (2016). *Primates and Philosophers: How Morality Evolved* (The University Center for Human Values Series, 30). Princeton University Press.

Wittgenstein, L., & Wittgenstein, L. (1969). *Preliminary studies for the 'Philosophical investigations': Generally known as the Blue and Brown books*. Oxford: Blackwell.