Skepticism regarding the self has been widespread among naturalist philosophers. Contrary to this view it is shown here that naturalism can provide a deeper understanding of the self. Starting with a phenomenology of the self it is argued that self-consciousness can be understood as an act of perspective taking. Thus, self-consciousness turns out to be a natural ability, which can be investigated empirically. These studies can further improve our understanding of the self.
The problem of the self has been one of most important issues in occidental philosophy ever since it was discussed in Florentine Neoplatonism. However, many philosophers think that there is a basic incompatibility between naturalism on the one hand and a sufficiently strong idea of the self on the other. Here I would like to show that this is not the case. You can have the cake and eat it, too. In order to show this, I will present a reductive account of the self in naturalist terms. Note that I will use the word “reduction” not in the ordinary sense of the word, according to which reduction means decrease. Rather, I am using it in the original Latin sense of *reducere* which is at issue when we talk about reducing an effect to its cause, say in order to understand how the effect came about. Obviously, such an explanation does not put the discussed phenomena at risk. Rather, it helps us understand them. Accordingly, I will try to reduce the higher level phenomenon of self-consciousness to psychological and neurobiological lower level phenomena – not because I am denying its existence but because I want to understand it. In doing so, I will also discuss the role of language in the ontogeny of self-consciousness. I will refer to empirical evidence which seems to show that language might be not very important, particularly in the earliest stages of the development of self-consciousness. Also, I think that the development of
self-consciousness in ontogeny tells us something about the elements which constitute the underlying abilities. In using the word ‘self’, I do not mean some mysterious Cartesian entity outside the physical world. It is just shorthand for ‘a person who has self-consciousness’. I will explain shortly in more detail what I mean by this.

The existence of the self has been put in doubt since long. Such doubts can already be found 1500 years before Christ in the Indian philosophy of the Vedas, as well – and much later – as in the German tradition, particularly in the philosophy of Schopenhauer and Nietzsche. More recently the idea that the self is just an illusion has been defended by Daniel Dennett: “The self turns out to be a valuable abstraction, a theorist’s fiction rather than an internal observer or boss. If the self is ‘just’ the Center of Narrative Gravity, then, in principle, a suitably ‘programmed’ robot, with a silicon-based computer brain, would be conscious, would have a self (Dennett 1991, p.431).” A similar idea has been brought forward by Thomas Metzinger. In his view, there is no self, but only a self-model: “Metaphysically speaking no such things as selves exist in the world: the conscious experience of selfhood is brought about by the phenomenal transparency of the system model” (Metzinger 2003, p.627).

What Metzinger seems to have in mind is that we have a representation of ourselves, the so-called self-model, which we mistake for a real entity inside ourselves, that is, for something like the notorious homunculus. It seems that Dennett has a similar idea: what is really at issue when we talk about the so called self is the notorious homunculus. I do not think this is true. Obviously the idea of a homunculus is a very bad one, but I think that there is a very rational way of talking about the self that does not leave you with a homunculus. I would like to show what this alternative might look like.

I will start with some classical views, particularly with some skeptical positions that I have already alluded to. Then I will refer to a discussion which was started by Fichte and which still has some followers in Germany particularly in the so-called “Heidelberg School” around Dieter Henrich (Henrich 1967; Henrich et al. 1966) and Manfred Frank (Frank 1991). Fichte pointed to a specific problem of any theory of self-consciousness which concerns recognizing yourself. To recognize an object such as a bottle of water you need certain criteria that enable you to distinguish the water bottle from other objects. The problem with self-recognition is that you already need self-consciousness if you want to find out which criteria can help to distinguish yourself from someone else: In order to determine whether a certain feature can serve as a criterion, you have to already know
that this feature is one of your features. And this means that you need self-
consciousness at this point already. Many philosophers have tried to solve
this problem. I will present a solution proposed by the Heidelberg school,
which I think goes in the right direction, but still leaves some questions
unanswered. Finally, I will suggest an alternative idea.

1. Classical Views

First, however, I will talk about some classical views. Self-consciousness
already played a role in pre-theoretical thinking of humans about
themselves, for example in ancient religions and myths. Most typically the
self was thought of as a kind of observable substance, say as a soul, which
consisted of a particularly fine matter, or some kind of breath. The former
idea is present in the platonic tradition, particularly in the Timaeus, the
latter can be found in the Bible but it also persists in certain Greek and Latin
terms like flatus, spiritus, pneuma or psyche that are used to refer to the self or
the soul.

Another idea was prevalent in ancient Egypt, where every person was
thought to have various different souls each of which having a specific
function. One of them represented a person’s individuality. This soul had the
same look, size and shape as the persons themselves, and left the persons’
odies when they died.

The ancient concept of a soul was much broader than most of the terms we
use today. As a consequence, we need various different concepts in order
to cover the entire meaning of the “soul”, among them “consciousness”,
“perception”, “emotion”, but also the “identity”, “self”, and “self-
consciousness”.

One of the most famous theoretical accounts of the self in classical
philosophy has been called the reflection theory which was endorsed, among
others, by Descartes. The basic idea here is that the self emerges from self-
reflection. So in order to develop self-consciousness, you have to think
about yourself. In this process you are simultaneously the subject and object
of thought, and whatever thoughts you may have, they are immediately
present to your introspection.

However – and this was Fichte’s idea – the problem is that this theory ends
up in a vicious circle. Imagine that you want to recognize your fear as your
fear in such a process of reflection. In order to do so, you already need some
kind of self-consciousness regarding the feeling in question, otherwise you
might end up ascribing your feelings to someone else or someone else’s
feelings to yourself. The same holds when you try to recognize yourself in
a mirror. In order to do so you already need some idea of what you look like
– otherwise you might mistake someone else for you. But this shows that
you already need self-consciousness. As a consequence, the self-reflection theory turns out to be false because it cannot explain what it is supposed to explain.

2. **Skepticism**

As I have already mentioned, there has been quite some skepticism about the self in the history of philosophy. One famous example is David Hume:

> For my part, when I look inward at what I call *myself*, I always stumble on some particular perception of heat or cold, light or shade, love or hatred, pain or pleasure, or the like. I never catch myself without a perception, and never observe anything but the perception. When I am without perceptions for a while, as in sound sleep, for that period I am not aware of myself and can truly be said not to exist. If all my perceptions were removed by death, and I could not think, feel, see, love or hate after my body had decayed, I would be entirely annihilated – I cannot see that anything more would be needed to turn me into nothing. If anyone seriously and thoughtfully claims to have a different notion of himself, I can’t reason with him any longer. I have to admit that he may be right about himself, as I am about myself (Hume 1978, p. 252; Book I, Part IV, Section VI).

Yet, I do not think that Hume really wants to deny the existence of self-consciousness. What he wants to deny is rather the existence of an immaterial substance as it had been postulated by Descartes. So what appears as an attack to the self turns out to be an attack to a specific theory of the soul. This is possible because – as already indicated – there has been no clear distinction between the self and the soul in huge parts of the philosophical tradition. Kant, for example, explicitly states that he uses the term “self” for the term “soul”.

Another strand of skepticism has been put forward by Marvin Minsky (1988). Minsky argues that there is not one self, but instead a multiplicity of agents within a person which cooperate with each other although none of these agents has more than a limited knowledge about what the others know. Today we might say that there are different systems within the brain or within the mind. Minsky claims:

---

1. “Wenn ich von der Seele rede; so rede ich von dem Ich in sensu stricto. Sofern ich mich nun als einen Gegenstand fühle und dessen bewußt bin; so bedeutet dies das Ich in sensu stricto oder die Selbstheit nur allein, die Seele. Diesen Begriff der Seele würden wir nicht haben, wenn wir nicht von dem Object des inneren Sinnes alles Äußere abstrahiren könnten; mithin drückt das Ich in sensu stricto nicht den ganzen Menschen, sondern die Seele allein aus” (Kant 1821, p. 200).
All this suggests that it can make sense to think there exists, inside your brain, a society of different minds. Like members of a family, the different minds can work together to help each other, each still having its own mental experiences that the others never know about. Several such agencies could have many agents in common, yet still have no more sense of each other’s interior activities than do people whose apartments share opposite sides of the same walls (Minsky 1988, p. 290).

What Minsky wants to deny is that there is one single agency, part, or system which exerts control and also has overall knowledge.

3. The Heidelberg School

These were only some famous examples for skepticism regarding the self. I have now mentioned the main points of some skeptical arguments about the self. Now I would like first to refer to one strategy of “saving” the self, which has been done in response to Fichte’s criticism of the reflection theory. Fichte argues that you cannot develop self-consciousness by reflecting on yourself (Fichte 1991, p. 11). The members of the so-called Heidelberg-School, Dieter Henrich and Manfred Frank, have developed an alternative idea. They think that the self is real and that we do have self-consciousness. However, they argue that the reflection model that was introduced above is misguided since it cannot explain the emergence of self-consciousness. Apart from this, it is also misleading to take the self as an internal object, since this would again invoke some kind of homunculus. Like Fichte, Henrich and Frank claim that self-awareness cannot emerge from self-reflection, and the self cannot be a constellation of properties. The reason is simple: We would have to know for each of such properties that it is our own property. And this would require self-consciousness in the first place rather than explaining it. Let us assume that the feature that best identifies myself is the fact that I am the best aluminum welder in Berlin. So this feature would distinguish me from all other residents of Berlin. So why cannot I use this property in order to identify myself in acts of self-consciousness? The reason is that I already need self-consciousness in order to know that this feature belongs to me! So whatever the constellation of features or properties is, and however complex it may be, I cannot recognize that this constellation is my own or that it is the constellation that identifies myself prior to being self-conscious. The reflection theory therefore cannot explain the emergence of self-consciousness.

Henrich and his colleagues have concluded that because reflection does not explain the emergence of self-consciousness, this ability has to start with some sort of pre-reflexive self-awareness. So before you can begin to
reflect on yourself you need some kind of direct, pre-reflexive access to yourself. Frank argues: “We cannot describe self-awareness as an awareness of something, if this ‘something’ represents a single object named ‘self’ (or ‘I’ or ‘person’). Self-awareness is not object-like, its familiarity is not mediated by something else, its original instantiation is irreflexive, without criteria, and it’s not based on observations, either” (Frank 1991, p. 5).
Frank makes it clear what self-consciousness is not, but he does not provide a positive account – which is an endeavor that the Heidelberg school never attempted. His idea of pre-reflexive self-awareness is not easy to grasp: If it is not object-like, irreflexive, without criteria and not based on observation, what, then, is it? We know that the reflection theory will not help us gaining an understanding of the self, but what does help is still an open question, even if we accept the account provided by Henrich and Frank.

As a consequence, the pre-reflexive self has to be made intelligible too. So we need additional explanations of what such a pre-reflexive self might look like, and this is what I would like to present in what follows. I will start with some minimal criteria for self-awareness, and then describe the phenomena involved. A reductive account of self-consciousness, after all, can only be successful if it captures the entire phenomenon.

So what do we mean when we talk about self-consciousness? What is the relevant phenomenon? It seems essential, first, that one is able to recognize one’s body. This, in turn, means that one can distinguish it from things outside the body, including other persons’ bodies. This is a basic ability which can be found in many animals, and it may be that even plants might have some rudiments of it.
However, this ability is certainly not sufficient for self-consciousness. In addition, recognition has to be what I call transparent: you have to recognize your properties as your properties, your body as your body, and yourself as yourself. Many animals are able to perceive themselves in a mirror but most of them fail to understand that they are looking at themselves, so they do not perceive themselves as themselves. Something similar might happen when I talk about the best aluminum welder in Berlin without understanding that I am talking about myself. Given that I am, in fact, the best aluminum welder in Berlin, I am talking about myself. But I am not talking about myself as myself in the sense at issue here, as long as I do not understand that this is so.
So far we have only talked about individual acts of self-consciousness as they occur when I recognize myself in a mirror or fail to recognize my visual perception as my perception. But this is certainly not sufficient for a
description of the phenomenon of self-consciousness, even if we focus only on the most essential features. We would not say that someone has self-consciousness in the full sense of the word unless he would have something like a persistent idea of himself, what he is, what his name is etc. I take it that this also requires autobiographical memories. So what is needed as a third criterion in order to capture the phenomenon of self-consciousness is something like a self-concept that is, a persistent representation of oneself. A person who sincerely claims to be Napoleon or George W. Bush would be considered to suffer from a severe self-disorder; the same would hold if someone told us yesterday that they are a very good aluminum welder but deny this very fact today. I assume that this is also where language comes in, even if language is not a necessary precondition for the first two criteria of self-consciousness.

The most important problem at stake here is transparency, which means recognizing yourself as yourself. This is also where the pre-reflexive self comes in according to the Heidelberg School. Philosophers have argued that this act of recognizing oneself as oneself leads to a paradox: On the one hand, there is no self-consciousness without being able to recognize oneself as oneself. Unfortunately, however, you need self-consciousness in order to recognize yourself as yourself.

In what follows, I would like to suggest a solution for this problem which also contributes to a deeper understanding of the emergence of self-consciousness. Though I think that the body and the “body scheme” play an important role for basic aspects of self-consciousness that have been described as the “core self” by Antonio Damasio (1994), I will focus on higher level cognitive phenomena like perspective-taking and theory of mind because I think that these abilities can give a very important contribution to our understanding of self-consciousness.

Let us first talk shortly about the body scheme. It implies the ability to recognize your body as your body, not on a personal, explicit level; rather, it appears as a direct feeling that your body is yours. For instance, you would react differently depending on whether someone threatens to injure your hand or some external object: Most likely, you would withdraw your hand only in the first case. This seems to show that we have a very deeply rooted, sub-personal access to our own body. Though I will not take a definite position here, it seems that this kind of body scheme is part of the immediate feeling of familiarity that we have with respect to ourselves. A body scheme also appears to be present in non-human beings, as it is non-cognitive, sub-personal and represented as a feeling.

Empirical support for this approach comes from the study of the disorders
of the body scheme. Patients who suffer from somatoparaphrenia, a psychological disorder, really think that their own limbs are not parts of their body at all (Sacks, 1985). Merely telling the patients that this idea is irrational will not help them. It seems as if there is something wrong that is not affected by rational arguments. This is one reason why I think that the sub-rational body scheme may play a role in our immediate feeling of self-familiarity or self-acquaintance.

Additional evidence for these lower level aspects of self-consciousness comes from studies that show how the body scheme can be misled. Ramachandran (Ramachandran & Hirstein, 1997) treated a patient with phantom pain in his left hand, although it had been amputated. Ramachandran treatment consisted in showing the patient the mirror image of his right arm, thereby tricking the patient into believing that he was observing his left arm with a hand attached to it. Apparently the patient included the mirror image into his body scheme, thus allowing him to overcome his phantom pain.

Another example is the so-called “rubber hand illusion”, which also shows that the body scheme depends on sub-personal stimuli, rather than being reactive to rational reasoning. In this paradigm, the experimental subjects have one of their hands covered so that they cannot see it. What they can see instead is a rubber hand which is visible exactly above their covered hand. The experimenter then touches the subjects’ covered-hands and the rubber hands in exactly the same way; as a consequence, the subjects include the rubber hand into their body schemes and mistake the rubber hand for their own hand. For instance, if the experimenter threatens to injure the rubber hand with a hammer, the subject will retract his real hand. This demonstrates again that the body scheme is something sub-personal, and that it is at least a good candidate for constituting basic self-familiarity.

But there is a second possible candidate for an explanation of pre-reflexive self-consciousness namely perspective-taking, which takes a completely different route. It is cognitive, to some extent personal, and involves taking the perspective of a person with whom one is cooperating or communicating. There are certain requirements that have to be met in order to have this capacity. If perspective-taking is a good candidate for the development of self-consciousness, small children should be able to develop such abilities from a very early age on. In the following, I will give an account of the constitutive elements of perspective-taking, and show that children master these before they develop self-consciousness.

The first requirement in order to take someone else’s perspective is to
identify those beings whose perspective you can take. If you attempt to take the perspective of, say, a table or a camera, you will fail. However, the ability to distinguish between living beings and non-living beings is already present in infants at 2 to 3 months. Amanda Woodward (1998) shows that at 5 months, infants expect different behaviors from living and nonliving agents. In this study, a small child watched both a human and a mechanical arm reaching for an object. When the object was displaced the 5 months old child expected only the human arm but not the mechanical arm to account for this displacement and to reach for the object at the new location. Apparently, the children expected that the person had the intention to get the object, whereas they expected that the mechanical arm would just repeat its movement without any intention. This finding has been replicated several times and it shows that even at this early age, children can selectively attribute intentions to human agents. This capacity does not include perspective taking but it looks like a good starting point for developing this ability.

Between 7 and 9 months, infants are able to distinguish between human and non-human animals. This is important, as it seems to show that they can make such basic distinctions even before they are able to learn language (Pauen & Zauner, 1999). Of course, merely distinguishing human agents from non-human animals does not suffice for real perspective-taking. Children also have to be aware of, or be able to identify the behavior of their conspecifics. Studies conducted by Andrew Meltzoff (Meltzoff & Moore 1983; Meltzoff 1988a, 1988b) have shown that children have the ability to imitate soon after birth. If you stick your tongue out in front of a newborn, he will likely imitate your behavior. Interestingly, he will refrain to do so when a similar movement is performed by a mechanical object.

Even very young infants thus seem to be aware of particular kinds of behavior that their conspecifics display. There is a theory about why they are able to translate the visual information into behavioral output, that does not assume complicated inferential processes happening in the infant’s mind when he is imitating – such as identifying his tongue, finding out where it is, and how to move it. This theory suggests that there is a direct coupling between the motor system and the cognitive system, which draws on motor abilities to improve our understanding of observed behavior (Gallese & Sinigaglia 2014). It seems that something like this happens in the case of imitative behavior in newborns.

Finally, there is a last requirement for perspective-taking, which appears to be the most important one. So far, I tried to show that children are able to identify candidates for perspective-taking and that they can interpret their behavior and their movements from very early on. However, in order
to be able to take someone else’s perspective they should also be able to
distinguish between their own mental states and others’ mental states,
more particularly between their own feelings and others’ feelings as well
as between their own beliefs and others’ beliefs. This means that they have
to make a perspectival distinction and recognize mental states in others
that they do not experience themselves. There is evidence that children also
develop at least rudiments of this capacity at an early age.
For instance, 6 weeks old infants are able to distinguish emotions in the
facial expressions of their caregivers; and 4-months olds are able to use
emotion expressions to assign a voice to a character in a movie: if they see a
movie character and hear a voice that does not fit the emotional expression
shown, they get irritated (Oerter & Montada 1995, p. 230). 9-month-olds
can use facial expressions of their caregivers in order to assess special
situations, a phenomenon termed “social referencing” (Feinman, 1982).
When 9-month-olds are exposed to a potentially dangerous situation, for
example an unfamiliar toy, they look at their caretaker: if the caretaker is
relaxed, they take the toy; if the caretaker does not look relaxed, they do not
take it. The assumption here is that the children themselves do not have this
emotion; rather, they are able to interpret the emotion of their caretaker
and therefore take his (or her) perspective in order to assess the situation.
Generally, children develop basic forms of perspective taking as early as
with 9 months when they acquire secondary intersubjectivity: “At 9 months of
age infants begin to understand that other people perceive the world and
have intentions and feelings toward it” (Tomasello 1993, p. 175).

Now let us look at data indicating when the first signs of self-awareness
occur: when are children able to recognize themselves as themselves? An
empirical test to shed light on this question is the so-called rouge test or
mirror test, which children are typically able to pass between 15 and 21
months of age (Neisser 1993, p. 16). Of course, as with any empirical method,
it can be debated what this test really shows (Loveland 1986). However,
although far from perfect, it is still one of the best tests for the existence of
self-awareness. In this paradigm, unbeknownst to the children, a red dot is
placed on their nose. Subsequently, the children are put in front of a mirror.
If the children do not recognize themselves, they will try to play with the
children in the mirror; however, about the age of 15 to 21 months, they
start to touch their own nose in order to clean it. This is taken to show that
the children recognize themselves in the mirror and know that they are
looking at their own nose. Thus, a form of self-awareness is displayed here:
children passing the rouge test are aware of themselves as themselves; they
understand that it is themselves what they see. Between 18 and 24 months, children start using the expression “I” in the right way, they understand how the first-person pronoun works. At 24 months, children can do self-ascriptions of persistent features, such as their gender (Fogel 1997, p. 380). Arguably this is the time when they make the first steps to develop something like a self-concept. It is important to note that in the course of a child’s development, the conceptual development comes after the development of the underlying non-verbal abilities – even if conceptual abilities are certainly essential for full-fledged self-awareness. In any case, there seems to be quite some evidence that self-awareness, understood as the ability to recognize yourself as yourself, is based on perspective-taking. Small children who try to hide themselves by closing their eyes do not have self-consciousness regarding their own perceptions. They confuse their own perspective with what is going on in the world: when they cannot see anything, then no one can. In order to recognize their perceptions as their perceptions, they have to recognize the difference between their own and another person’s perceptions. Note that all this can be done without referring to any criteria that identify yourself. So unlike the old reflection theory, understanding self-consciousness as perspective taking does not lead to a vicious circle. Even more important, it demystifies self-consciousness and makes it a subject of empirical research.

But what about the pre-reflexive self, this direct awareness of ourselves as it was assumed by the Heidelberg School? According to the present approach, two aspects are important: First, immediate access to our own emotions, perceptions, and beliefs comes for free, as soon as we are conscious. Having a feeling of pain means having direct first person access to this experience. What self-consciousness adds is, second, the ability to recognize this experience as your experience. And this is what is explained by perspective taking, that is, which helps us to recognize what is specific regarding our own emotions, perceptions and beliefs.

So being able to recognize yourself as yourself means being able to recognize the difference between yourself and someone else, and this ability develops in degrees and it can be investigated with empirical methods. As we have seen, it starts with recognizing a difference in emotions, which seems to work already at 9 months, and is followed by recognizing differences in perception. The next step is recognizing a difference in beliefs, which requires a theory of mind in the stronger sense – a theory about the other’s mental states and beliefs. The question here is, how can a child understand her beliefs as her beliefs, and know that someone else might have different beliefs?
There are many studies investigating this capacity, of which I will only refer to one: the so-called Sally-Anne-test, or false belief test, which was introduced in the 1980s (Wimmer & Perner 1983; Baron-Cohen et al. 1985). Here children are shown a scenario where two characters, Sally and Anne, are in a room together. Sally places an object, a ball, into a box and then leaves the room; while she is gone, Anne moves the ball into a different box. The experimenter then asks the child observing this scene: “When Sally comes back, where will she look for the ball – in the first or the second box?”. The right answer would obviously be the former: Sally still has the belief that the ball is in the first box, where she left it. However, children up to around 4 years say that she will look in the second box, because they are not yet able to distinguish between what is really the case and another person’s, Sally’s, false beliefs. This shows that the ability to recognize your beliefs as your beliefs develops at about 4 years. Of course, again, this ability develops in stages, and its precursors are instantiated much earlier. But a very stable finding of this experimental paradigm is that it takes around 46 to 49 months to develop.

The conclusion to take away from this is that recognizing yourself as yourself is not an ability that you either have or do not have; rather it comes in different degrees: You may be more or less proficient in this ability which develops at different stages of the ontogenetic development with regard to emotions, perceptions and beliefs.

I have tried to argue, against skepticism by Dennett, Metzinger and others, that self-awareness is not a fiction. It does not require reference to a self-object or some kind of property that happens to be specific for yourself. What it does require is the ability to recognize yourself as yourself, which can be explained as the ability to distinguish between your own experiences, beliefs, and other features and the related mental states and features of others. These abilities, particularly perspective taking, are not mysterious at all. They can be investigated in empirical studies which improve our understanding of this ability. So there is no reason to doubt that the self exists and naturalism can explain how it emerges.

---

2 I would like to thank Laura Schlingloff for revising the manuscript.
REFERENCES
Frank, M. (1991), Selbstbewußtsein und Selbsterkenntnis. Essays zur analytischen Philosophie der Subjektivität, Reclam, Stuttgart;
Kant, I. (1821), Vorlesungen über die Metaphysik. Zum Drucke befördert von dem Herausgeber der Kantischen Vorlesungen über die philosophische Religionslehre, Keysersche Buchhandlung, Erfurt (Metaphysik Pöltitz);