GÖTEBORG UNIVERSITY
DEPARTMENT OF PHILOSOPHY

PERSONS OR THINGS

On the ontology and conceptualization of morally relevant entities

Ida Hallgren Carlson

Bachelor of Practical Philosophy thesis
2007
Advisor: Gunnar Björnsson

Contents

1. Introduction

This essay deals with the use of the concept of personhood. In their (2007) article Martha J. Farah and Andrea S. Heberlein conclude that in the outer world there are simply no such objects as the ones we try to refer to when talking about persons: the idea of such entities is altogether illusionary. It is this conclusion, and its suggested implications for moral theory and practice, that will be presented and discussed below. Finally, it will be suggested that the important empirical facts presented by Farah and Heberlein point in other directions than the ones proposed in their article.

My main reasons for wanting to investigate this subject are my intuitions that there *are* morally relevant entities that we try to talk about when talking about persons, and that these entities are different from, or at least more stable than, momentarily existing subjects of experience, but also that there *are* moral problems connected to the use of the concept person.

Are there morally relevant entities or do we have to base our morals on illusions? The illusion of a soul, like the idea of a Cartesian ego interacting with our worldly bodies, is likely to have evolutionary benefits for the being acting upon this illusion. It is good for us, (for the passing on of our genes), to believe that we are something special - something stable and united, and this may be why we keep having these persistent ideas about being persons that exist over time. Is there also some truth to these beliefs we have about being entities that can exist over time? If so; who are "we"? What kind of biological beings are morally relevant entities?

In his influential book Reasons and Persons, Derek Parfit thoroughly investigated the concept of personhood and pointed out common false ideas about a need to use further facts such as Cartesian egos when talking about persons. The ideas of Farah and Heberlein, in ways similar to those of Parfit's, lead them to conclude that even though our everyday beliefs in persons are beneficial, we should centre our theoretical moral discussions around the quest for awareness and certain psychological criteria that influence the moments of awareness in morally relevant ways. Their main conclusion, different from that of Parfit's, is that the concept of personhood is based on an illusion and therefore irrelevant to moral theory. I believe that we should not settle with this conclusion. Despite our common misunderstandings of the true nature of persons there might very well be reasons to use a metaphysical concept of personhood. Contrary to Farah and Heberlein I will claim that there are reasons to talk about moral agents, which could be referred to as persons. Besides reasons to talk about persons as moral agents, there are also reasons to talk about entities that are moral objects. Moral objects

are not the same things as moral agents, and hence there are reasons for not referring to all types of moral objects as persons. The fact that moral objects are not necessarily the same things as moral agents becomes illuminated if we ask questions about how we should treat non-human animals. Non-human sentient beings can be moral objects, i.e. morally relevant entities that we should show moral concern, even if they are not themselves moral agents that can show moral concern to other beings.

The problem that often seems to arise when advocates of animal rights enter the debate on where to draw the line between persons and things is precisely that we do tend to seek a line to draw between persons and things. Those who strive for moving this line so that some non-human animals fall within the category of persons might succeed in moving this problem but they will not remove it. Further, as will hopefully become clear in the summary of Farah and Heberlein's article, there are reasons for believing that this strife for moving our intuitive line between persons and things will never become fruitful. The advocates of an expanded category of persons will fight a war on innate mechanisms in the human brains. I find it unlikely that Peter Singer, an animal rights advocate, claiming that it is not directly wrong to kill a non-person but that there are non-human animals that are persons (1979), will ever succeed in making it widely accepted to use the person concept for non-humans. What Farah and Heberlein's mapping of empirical evidence from the field of cognitive neuroscience shows, is that the tendency to draw a line between human-like persons and (other) things is actually a deeply ingrained mechanism of the human psyche. Personhood is generally ascribed through the use of a simplistic feature-based categorization-system that gives weight to morally irrelevant outer features such as bodily shapes, movements and faces.

My conclusions will be that instead of following Farah and Heberlein in abandoning the person concept, we can keep the metaphysical concept of personhood as well as investigate a new concept that is less bound to innate prejudices. The personhood concept can be kept for discussions that involve human agents while other concepts are needed for discussions about 'moral objecthood'. To be useful and meaningful, a large part of morals has to deal with talk about persons, agents, responsibility and other metaphysical concepts. Besides this, morals need to deal with the facts of the external world. Moral reasoning and answers to questions about how to act morally have to be linked to investigations of the true nature of moral objects. I will hypothesize that moral objects, as morally relevant entities existing over time, can be conceptualized in an advantageous way by references to the having of self, which I will propose is something that can not only exist over time, but can also exist in degrees.

Focusing on the inner qualities of selfhood might allow for more truthful and relevant descriptions of moral entities than what is made possible by our brains prejudiced way of categorizing the world in persons and things. The dividing of the world in persons and things has often been followed by a dividing of those who are worthy of moral concern and has a right to their own lives, and those who are not and has not. Such a simplistic kind of thinking and acting is morally problematic.

2. Summary of Farah and Heberlein's article

2.1. The definitional problem of personhood

In their article Personhood and Neuroscience: Naturalizing or Nihilating? Farah and Heberlein start out by describing the puzzle of personhood. Over the centuries the debate on the nature of persons has produced many attempts to specify criteria for personhood. Such attempts have ended up with criteria such as *"rationality, self-awareness and the awareness of being an individual that exists over time"*, i.e. Locke's definition of personhood, *"self-consciousness, rationality and ability to be concerned with worthiness of blame or praise"*, i.e. Englehardt's definition, and *"intelligence, self-awareness, the capacity to view others as having intentional mental states, to use language, and to be conscious in some special way not shared by other animals"*, i.e. Dennett's list of criteria, (Farah and Heberlein 2007, p 37.). Farah and Heberlein give an account of these and other suggestions, some of which deal with passing criteria for personhood such as at least an IQ 20, and conclude that all defining criteria have shown to be elusive and arbitrary.

2.2. Naturalizing personhood

As a response to the elusiveness of the concept of personhood, Farah and Heberlein suggest an alternative take on the problem. By looking for empirical data, perhaps we could find a "natural kind" that corresponds to our intuitions about personhood. Having found such a category of objects that corresponds to the concept of person, these biological criteria could then solve cases where intuitions have tended to collide. After first having discussed the concept of personhood in philosophy, e.g. Kant's opinion that all moral action is dependent on the ability to distinguish between persons and things, Farah and Heberlein now turn to science asking for possibilities to develop a scientific taxonomy that could replace a folk taxonomy and philosophical quarrels about how this line between persons and things is to be drawn.

Maybe biology could reveal the essential difference between persons and non-persons in the same way biology once stated that the essential difference between plants and animals is not that plants are green, (some are not), or tend not to move, (some do), but that only plants can photosynthesize.

However, using neuroscience to find a place for personhood in nature does not take Farah and Heberlein very far. Soon they conclude that measuring EEG from fetal brains and other attempts to empirically measure correlates of suggested psychological criteria for personhood, only lead to quantitative data that do not reveal any qualitative transition point that could be correlated with the transition from non-person to person. Likewise, a definite point of transition from person to non-person at the end of the lifespan is equally hard to find. While the loss of a human life has come to mean the same thing as the death of the brain, there has not been a consensus reached on what that corresponds to the death of the person. Generally, the transition from being a person to being a non-person has been associated with the loss of higher cortical brain functions. This correlation between personhood and higher cortical functions raises a number of questions for Farah and Heberlein. *"Which cortical systems in which combinations are critical and how much functionality is required of each of these systems?"* (2007, p.40), they ask, troubled by the lack of hope of finding qualitative transition points between persons and non-persons.

What Farah and Heberlein find to be an even harder problem is again the arbitrariness they claim to have found when investigating the hunt for defining criteria of personhood. Since, they conclude, we still lack knowledge about which psychological capabilities that are crucial for personhood and therefore do not even know which systems that do matter, we will not be able to naturalize personhood via neuroscientific explorations. Therefore:

> *The real contribution of neuroscience to understanding personhood may be in revealing not what persons are, but rather why we have the intuition that there are persons. Perhaps this intuition does not come from our experiences with persons and non-persons in the world, and thus does not reflect the nature of the external world; perhaps it is innate and structures of our experience of the world from the outset. Thus, instead of naturalizing the concept of personhood by identifying its essential characteristics in the natural world, neuroscience may show us that personhood is illusory, constructed by our brains and projected onto the world.* (Ibid).

2.3. The person network

Farah and Heberlein write that it is a not "widely appreciated" fact that our perceiving and understanding of the world through our brains has important consequences for metaphysics and epistemology. They state that; *"We can only understand categories of reality and their regularities and interrelationships if our brains are capable of representing these categories".* (2007, p. 40.)

Clearly we can understand the category of persons, but how the perceived objects belonging to the mental category of persons are related to the outer world takes further investigation. Farah and Heberlein claim that we have reasons to believe that our concept of personhood is not related to any real objects in the same way that concepts referring to physical objects generally seem to be. A suggested reason for believing this is the way we gain knowledge about persons. While our understanding of the physical world is based on a fairly nice fit between our perceptions and the outer world that these are mapped onto, it could be somewhat different with our understanding of our social world. Farah and Heberlein point out three ways in which the way we gain knowledge about the social concept of personhood differ from the way we gain knowledge about the physical world. These three ways of gaining knowledge about social concepts will be described more thoroughly below, but I will mention them briefly already at this stage. First, we are born mentally equipped with a brain system that reacts to person-related triggers. Farah and Heberlein hold that we do not seem to create the concept of personhood by gaining data from the outer world, as when we gain our understanding of physical objects through a process of learning. Secondly, our sense that the world consists of persons and non-persons seem to be activated automatically: we do not gain this idea by careful reasoning but by automatic activation of certain parts of the brain. Thirdly, this process of judging percepts as persons or non-persons takes place irrepressibly. For evolutionary reasons we get a strong sense of perceiving persons when anything triggers what Farah and Heberlein labels our person network. This happens even if we have conscious knowledge about the trigger not being related to any living being at all.

The empirical data that Farah and Heberlein present to back up the idea of the existence of a person network can be summarized as follows:

In cases of prosopagnosia, patients suffering from damage to the fusiform gyrus on the ventral surface of the brain, fail to recognize human faces. Object recognition and even recognition of animal faces may be spared. Further, patients with preserved face recognition might show poor object recognition. This functional double dissociation observed gives

strong evidence for the existence of a brain area exclusively specialised on recognizing faces, or possibly on reacting to a set of characteristics associated with faces. Further, emotional facial expressions activate additional brain areas including the amygdala.

Nearby the face area and on the lateral surface of the brain near the temporoparietal juncture are two areas activated when we see human bodies, silhouettes or even stick figures. Shapes that are not bodies, but equally complex, do not activate these areas.

Bodily movements and actions activate another part of the temporoparietal juncture. It is characteristic human motion, (think of the sight of reflectors attached to points of a human walking in the dark), that is triggering this area. Specific parts of the temporoparietal juncture are activated by goal directed behavior and yet other parts by our thinking about people's mental states, an activity that also activate the medial prefrontal cortex. The medial prefrontal cortex has been shown to be involved in a variety of person-related processes such as the thinking of mental traits, the understanding of other people's cognitions and the describing of people as opposed to using the same adjectives to describe other objects. Anytime the research subject believes himself or herself to be interacting with or for other reasons come to process people, this area gets activated.

Farah and Heberlein establish that:

> *The weight of the evidence, from a sizable literature only sampled here, clearly supports the conclusion that the human brain represents the appearance, actions, and thoughts of people in a distinct set of regions, different from those used to represent the appearance, movements and properties of other entities. These regions together form a network that is sometimes referred to as "the social brain" (e.g., Brothers 1990; Adolphs 2003; Skuse et al. 2003) but could equally well be termed a network for person representation.* (2007, p. 42.)

2.4. Characteristics of the person network

Farah and Heberlein thoroughly describe the *automaticity* and the *innateness* of the person network, i.e. *"the tendency of the person network to be triggered by certain stimulus features even when we are aware that the stimulus is not a person",* and *"the genetically preprogrammed nature of the system, without a need to learn that persons exist in the world"*, (2007, p. 42). The first evidence of the automaticity of the network is drawn from an additional case of prosopagnosia where the patient, unable to turn off his dysfunctional face recognition system, was shown to be more able to process faces that were presented upside-down. More evidence put forward by Farah and Heberlein are the cases of trigger features such as smiley faces or stick figures that can activate the person system. People playing a

computer-run economic game came to choose more generous strategies when a pair of cartoon eyes was presented on the computer screen.

Any seemingly non-mechanical interactions between entities, even between very simple figures, will activate the whole person network. In one famous study these entities are animated geometric shapes moving in an interrelated manner and hence evoking psychological descriptions implying these entities having intentions and desires. This kind of evidence is said to indicate the person network being independent of rational, conscious beliefs.

The innateness of a system dividing the world in persons and non-persons is depicted by research on newborn infants. Babies tested within 30 minutes of birth show more interest in face-like patterns than in other patterns. Unlike other parts of the brain that when damaged may have great capabilities of recovery or of having its typical tasks being taken over by other parts of the brain, the face-system shows none of this plasticity. Damage to the face area occurring as early as the first day after birth is irreparable, which *"...implies that the category of human face, as well as its representation by specific brain tissue, is determined essentially at birth."* (2007, p. 43.) This conclusion of Farah and Heberlein is backed up by references to more studies on older infants.

There have been several studies showing specifically how infants have different expectations on the movements or 'behavior' of mechanical versus intentional objects. One study referred to by Farah and Heberlein is the work of Kulhmeier et al. where infants' different reactions to physical versus non-physical objects are *"interpret[ed] as evidence for a distinction in the infant's mind between persons and things."* (2007, p. 43.)

A final source of evidence for the innateness of the person system presented in Farah and Heberlein's article is the case of autism. Autistic children do not show the expected brain activity when presented to human faces, the kind of shapes that normally activate person-related cognition, or when reading stories involving the mental states of other people. Autistic individuals show abnormal interpersonal behavior and, according to Farah and Heberlein, generally prefer to interact with inanimate objects.

To summarize the empirical evidence presented above; the person network described gets activated autonomously, whenever triggered by the right kind of stimulus and independently of conscious beliefs. This brain system is genetically preprogrammed to divide the world in persons and non-persons; in objects that do and do not possess the trigger features.

2.5. Farah and Heberlein's conclusions

While having identified objective criteria for the category of plants, science has not yet been able to identify criteria for personhood. Farah and Heberlein suggest that this lack of clear criteria remains because the concept of personhood in fact does not respond to any real category of objects in the world. As made clear in their reply to comments on their article, Farah and Heberlein's main point is not that the concept of personhood is vague. *"The personhood concept"*, they write, *"suffers from more than fuzziness; its definitional problems run deeper than the lack of a sharp boundary."* (2007, p. W1.) To make this point clear they refer to a study by Rosch where some subjects indeed disagree on whether a lamp is furniture, but certainly agree on lamp belonging between a doorknob and a sofa in its degree of "furnitureness". Though, what that appears when placing a fetus, an ape and a patient in a persistent vegetative state on a ranking scale is no such matters of degree, but a deep lack of consensus on where to place such examples on a scale. Different persons could place different examples like this on different sides of the divide. The problems of defining and naturalizing personhood gives Farah and Heberlein reason to ask for other reasons of us having strong intuitions about the existence of persons, i.e. intuitions about the existence of a natural kind that could be described with further facts than being a merely human, or a thinking or a sentient being. The suggestion presented is that:

> *Our sense that the world contains two fundamentally different categories of things, persons and non-persons, may be a result of the periodic activation of this person network by certain stimuli rather than any fundamental distinction between the stimuli that do and do not tend to trigger it.* (2007, p. 44.)

Farah and Heberlein conclude that the belief that the concept of persons respond to a natural kind in the outer world is an illusion. To show how mental representations can be activated in a systematic way they use 'phlogiston' as an analogy. Though, while the 17th and 18th century scientific belief that the substance 'phlogiston' could explain the process of burning, Farah and Heberlein writes that the belief in persons is probably not as thoroughly wrong as the belief in 'phlogiston'. They lay down that there are other things in the world that have minds, and that mindedness can differ in degree and perhaps also in kind. *The illusion described seems to consist of a false belief in persons and non-persons being fundamentally different.*

Two reasons for this belief arising are presented; the separateness of the person network from other systems and the tendency of the network to get activated automatically by triggering features. Further, the person network is presented through a framework of evolutionary psychology and a clear adaptive value of having such a network, at least in a

previous world with less ambiguous cases, is elucidated and makes its existence even more plausible.

A lack of hope of finding facts that can determine whether something is a person or not gives Farah and Heberlein reasons to abandon philosophical and bioethical trials to find objective criteria for personhood. Their conclusions about how this is to effect morals are different for moral theory and everyday moral practice.

In ethics the opinion presented is that the question of personhood should be exchanged with questions about interests related to the existence of capacity for enjoying different psychological traits such as intelligence and self-awareness. The only alternative to a person-based ethics that is offered in the article is utilitarianism.

Though, it is stated, in everyday behavior our preprogrammed thinking in terms of persons can not be avoided, and, says Farah and Heberlein, neither should it be. Acting as if there are persons is nothing that we should try to avoid. This, they say, is because even if being bad metaphysics and in an evolutionary perspective a concept somewhat out of date, the concept of personhood does guide our everyday behavior in an overall beneficial way.

3. Questioning Farah and Heberlein's conclusions

Three conclusions reached by Farah and Heberlein can be summarized as follows; 1) Persons are illusions. 2) Moral theory should give up the focus on persons in favor of a utilitarian approach. 3) In everyday moral practice the illusory concept of persons is useful and overall unproblematic.

I will now take a closer look at these three conclusions and the arguments, or the lack thereof, that precede them.

3.1. The illusion of personhood

Farah and Heberlein compare their suggested illusion of personhood to famous visual illusions. Against our knowledge about what there is to perceive, visual illusions can make us see illusory spots or movements that are not there in the real figure or pattern. Our ideas about personhood can be as persistent and stubborn as such visual illusions.

In cognitive science it is a well-known fact that in normal object perception what we see can and will change if we come to realize that what we see is not for example a sheep running

loose in the park but actually just a hairy grey dog. Even if first having had a firm belief in seeing a sheep, this type of normal perception developed through learning about the world is easily adjusted and updated. Top-down processes; our conceptual knowledge and expectations, affect our understanding of our bottom-up, or data-driven, perceptions. (Reisberg 2001). When having realized that what we see is actually a dog, we will not keep having a part of our brain telling us that it is a sheep running around over there. Illusions can not be adjusted by top-down understandings in a normal way. I do agree on a likeness between tendencies in our visual system that make us keep seeing movements in a pattern that is clearly not moving, and our tendencies to keep acting and responding to facial patterns etc. A computerized voice on the telephone failing to "understand" what we say can make us frustrated in a way that a not so human-like input could never succeed in, no matter how aware we are of the fact that we are talking to a machine.

Our person percepts can be compared to visual illusions, but does this imply that there is nothing out there? That persons are illusions and that the moon, seeming to change in size due to its position on the sky, is not really there? Farah and Heberlein do state that we probably are not as thoroughly wrong when believing in persons as when believing in 'phlogiston'. Is this point not given enough weight when they decide to label personhood illusory? The answer to that question probably depends on how wrong we have to be to be said to have illusions.

Farah and Heberlein do state that there is clearly something out there triggering our person network. There are minds in the world. Still Farah and Heberlein conclude that our perceptions of persons are illusory. This conclusion is based on the premise that the brain system described is actually activating ideas about *persons,* some innate ideas that is. When we see a photograph of a face, they would mean, we instantly perceive *a person*. The premise that the social network actually does activate the category of personhood, and not just specifically faces or minds etc, is supported by the presented facts showing some trigger features to trigger the whole person network. If added up the data indicate that it is not just our pragmatic or culture-specific concepts that get triggered by the photo of the face, but our stereotypical, foundational and faulty thinking about the world as persisting of either persons or things.

As far as I know no cross-cultural studies have been testing the idea of the person category being a universally human one. Therefore one could ask if the focus on persons vs. non-persons is something that has developed alongside a cultural evolution in the western

world where a bias from Christianity and individualism-oriented thinking have created a conceptual framework that have come to effect the categorized kind of thinking depicted by Farah and Heberlein. However, besides the evidence showing at least some parts of the person network functioning before it could possibly have been influenced by any learned concepts, there is further evidence that ties in well with Farah and Heberlein's description of a preprogrammed social brain. Philip Robbins and Anthony I. Jack (2006) thoroughly discuss data depicting subsystems guiding our understanding of other people, as opposed to systems guiding our understanding of the mechanical world. Through what they describe as the intentional stance, the phenomenal stance and the physical stance, we automatically interpret the world in intentional, mental or physical terms. What Farah and Heberlein's presented data seem to suggest is that whenever we perceive something as mental or intentional, our entire social system, or what they label the person network, gets activated. This is clearly different from our common object recognition where learning processes have developed a fairly nice fit between the object and what we see, (e.g. normally a few lines symbolizing a cube does not *automatically and irrepressibly* activate the idea of a house, even though of course we could *choose* to think about the lines as a house).

To me it is clear to see that while it must have been important for our survival in this world to have a truthful understanding of its physical properties our beliefs about our social world could develop in all sorts of fancy ways and still be useful in their activation of adaptive responses to specifically social situations. Social norms are not true in the same sense as physical laws. The reason why a person with Antisocial Personality Disorder does not see a difference between moral imperatives and social norms is that there is no such difference in the outer world, and thus a lack of a certain 'normal' adaptive type of understanding of social situations can result in 'unnormal' responses to social situations. Normally, the perception of other persons instantly activates certain beliefs and also certain activation patterns. *"It is thanks to the person network's hair trigger that we slam on then brakes at the first glimpse of a human form in the road, rather than wait until our conscious mind has arrived at the belief that there is someone there."* (Farah and Heberlein 2007, p. 46.)

As an ophidiophobic person can get the whole flight or fight system running by perceiving something just resembling a snake, (a process that bypasses our slow and conscious thought-processes), something just triggering a part of the person network can get the whole system working. It is this type of uncalled for reaction that Farah and Heberlein are referring to as the illusion activated. Triggers activate a distinct category of objects and our

beliefs tied to this category are illusory. *"...the concept does not correspond to any real category of objects in the world."* (2007, p. 37.)

One could easily believe that the concept of personhood like any other concept is learnt, that the concept is a label that we through a learning process have come to associate with certain features whereof some just happen to be features that are triggered as early as 30 minutes after birth. However, what Farah and Heberlein seem to claim is that the fault made is due to our predisposition to categorize the world in certain ways.

To use an analogy; there is plenty of evidence for humans having the most stereotypical ideas about gender and gender roles at a very early age, (Golombok and Fivush, 1994). One could hypothesize this being a result of not only the acquiring of gender stereotype concepts but also being due to an organising of these concepts with the guidance of innate tendencies of categorizing the world in feminine and masculine. Such a tendency would not encourage a mapping of people in those degrees of both masculine and feminine characteristics that empirically driven theories of androgynity have shown all humans to possess, but would force an either-or division on the outer world[1]. Either-or divisions might sometimes be practical, but there are also often reasons to take a closer look at the reality behind such stereotypical concepts.

Even if agreeing on the non-existence of such categorical divide between persons and things one could still ask the following question; when we do not see a photograph of a face or the like, but an actual human being in front of us, could it not still be true that we do see a person? This question calls for a clarification of what that is meant by person. Of course we now see something that is really there, but is there a further fact that justifies separating out persons from the rest of the things in the world? Or is perceiving persons as over and beyond the other parts of the world like the hallucination in a toxicated mind where the sight of a tree branch triggers the perception of a monster? Farah and Heberlein would say that believing that what that is perceived as a person could not be described in the terms of being merely human or sentient etc, that it would take some additional facts maybe such as a reference to an immaterial soul or to some godlike type of beings distinct from other beings or objects, would

[1] When considering the gender analogy we do have reasons for doubting that the tendency to categorize is innate in the same way as our tendency to categorize persons vs. non-persons since gender triggers are not automatized in the same way and can be very culture specific, e.g. a child can believe that only women can have long hair and that cutting the hair changes the gender, (Golombok and Fivush, 1994). Recognizing faces on the other hand, is as we have seen not something that has to be learnt.

be to have illusory ideas about the nature of what we perceive as persons. To believe in such further facts would be something like having a religious belief triggered by arbitrary symbols perceived as meaningful. In this sense I would agree on the illusionary nature of what that can be referred to as persons, but this point adds nothing to what that have already come clear in Parfit's book Reasons and Persons (1984). There are no further facts like Cartesian egos or other such special entities that make up the person and that stay the same throughout our lives.

No one can argue against the fact that the debate on defining criteria for personhood has been long, though other persons than Farah and Heberlein might still have considered it fruitful. Farah and Heberlein do not explicitly label the debate a waste of time but impatiently looks for other takes on the problem.

According to Farah and Heberlein the elusiveness of the concept and the lack of a natural kind that correspond to the concept beg for a new approach where the concept of personhood is abandoned in favor of a focus on awareness and on relations between capacities for enjoyment and therefore morally relevant psychological traits. Farah and Heberlein do point out an alternative route for dealing with the questions about personhood and they do show that persons could be illusions. What that is meant by this claim of persons being illusions is as we have seen that the perceived fundamental dividing line between persons and things is a construction of the mind and that even if there are obviously human biological beings in the world, we do not need further facts to describe these individuals; it does not add anything to the description of the natural world to label human individuals 'persons'. Hence the very concept of person is by Farah and Heberlein seen as unnecessary.

The article is a suggestion and an offer of a new approach in the light of new empirical facts. Though, what that is not accomplished in the article is the closing of other roads: Farah and Heberlein never prove other projects to be logically or practically impossible. The argument presented could call for at least an agnostic approach to the existence of 'persons' as a real category of objects in the world, but this does not cancel out a need for a pragmatic use of the concept of personhood.

Even if not necessarily taking them where they want to end up; in the abandonment of the theoretical concept of personhood, I believe that Farah and Heberlein's argument does call for a deconstruction of the concept. Perceiving persons is by Farah and Heberlein said to be similar to perceiving visual illusions in that both are affected by specific brain mechanisms that help determining the outcome of the perceptual process and in that both are unaffected by

our conscious knowledge about their true nature. There are often evolutionary explanations behind visual illusions. Inborn tendencies that have been overall beneficial for our survival can distort the mapping between our phenomenal world and the outer world so that the moon will always look bigger by the horizon and smiley faces will always have us interpret emotional meanings.

For these reasons it seems like a good idea to be cautious about what that is really out there whenever there are preprogrammed systems involved in our perceptions. And being cautious might not be enough. Since we are in the debate on personhood dealing with a concept that guide important parts of our behavior it could be a very good idea to deconstruct the person concept and see which parts of it or which parts of what the concept refers to that should be acted upon. But while I do believe in perceptions of persons resembling visual illusions, while I deny the existence of a qualitative difference between persons and things and do not believe in it taking any further facts to describe what persons are, I still would be hesitant to label persons as illusions. I do not believe that *what we try to refer to* when talking about persons are illusions that necessarily do not belong in moral theory. This leads us over to the discussion on the relationship between our moral ideas and the outer world.

3.2. Moral theory

To talk about moral concepts as illusions implies that a certain relationship between metaethics and reality is necessary for a concept to be morally relevant. Farah and Heberlein clarify this premise in their reply to comments on their article.

> *What we are doing is addressing the relation between the moral concept of a person and the natural world, and in that sense we are assuming that the natural world is relevant to moral theory. Although moral principles themselves may not require empirical validation, they do refer to entities in the real world, and for bioethics in particular the way in which we anchor such principles in empirical reality is crucial.* (2007, p. 63.)

I do agree with Farah and Heberlein's underlying opinion that it is the true nature of morally relevant entities that should guide our actions towards these entities. There has to be some 1st or 3rd person type of investigation that can help provide the answer to whether or not someone is a person. I believe that to be or not to be a person must have something to do with the person itself, (rather than just rely on commonly shared beliefs about what persons are and look like). How we are to act towards moral objects must have to do with what moral objects are like.

It is a clear mistake though, to draw the conclusion that utilitarianism is the only option if personhood is illusory, a mistake that is actually pointed out, but overlooked, in the quote above. Farah and Heberlein correctly state that moral principles may not require empirical validation. Since utilitarianism is a moral theory on how to act and not a theory about the nature of morally relevant entities, the conclusion drawn has no support. There are also famous utilitarians, e.g. Peter Singer, that are in favor of the person concept. Though, Farah and Heberlein view utilitarianism as an *alternative* to person based ethics. There is a great gap between the subject dealt with in Farah and Heberlein's article, the nature of moral entities, and the subject of moral principles.

Having concluded that what moral principles to act upon, is beyond the scope of this essay, I will now focus on the nature of moral entities. No data presented so far has been pointing in the direction of there being a fundamental and qualitative line to be drawn between persons and non-persons. No data suggests that the categorical difference promoted by our automatized person network relates to any real conditions. However, as in other cases where agnosticism might seem to be the only alternative, pragmatism can be an option. We can investigate our vocabulary in moral theory and see if there is a place for a concept like personhood between moral principles and values. What words do we need to be able to talk about those entities towards which we are to act in certain ways, to secure values relevant to those entities? It might be that we find the pragmatic need for a concept like personhood and it can turn out that what we try to talk about when using the concept person is something existing in degrees. Or rather, we can find a need for making distinctions between different kinds of moral objects because of a difference in certain characteristics, and these characteristica can be of kinds that can vary in degree.

Are we now back to where we started? In a never-ending debate on what criteria to choose for describing persons, left with a deep lack of consensus? No, at least learning about the somewhat primitive person network has taught us that Kant was wrong when he let his innate tendencies to categorize the world dictate his opinions about the world being simply divided in persons and things.

Having agreed on the theoretical possibility that being a person could be a matter of degree, we do not have to listen to Farah and Heberlein complaining about the lack of qualitative transition points between persons and non-persons. We can again try the short-cut that the quest for neuroscientific correlates of criteria is. Would we not easily find such correlates but are able to establish the psychological traits without its biological correlates,

then this will once again be a another possible path to take. Unlike the preprogrammed category of personhood, a new concept, formed for pragmatic reasons, like most of our concepts, could be interwoven with our empirically based understanding of those entities as they exist in the world; interwoven with those entities that the concept has been created to relate to. This new moral concept could be related to the natural world in the way that Farah and Heberlein for good reasons demand our moral concepts to be. We would now no longer, like Farah and Heberlein, be asking the question of whether or not personhood is in the world. Neither would we, like an agnostic but for pragmatic reasons religious person, simply invent concepts that we find useful but cannot validate. What we would do is asking how the morally relevant parts of the outer world should be conceptualized.

Still there is one important question to answer before creating such a concept; do we really need it? The pragmatic approach to the matter of personhood takes that we start by investigating the role of and the need for the concept. Farah and Heberlein do not see a role for the concept of personhood. Rather, spending time on debating personhood is described as a distraction that takes focus from important issues like determining to what degree a being is aware. Other thinkers have other opinions and intuitions on this matter.

In 'I' The Philosophy and Psychology of Personal Identity, Jonathan Glover concludes that our morals and all our close relationships depend on us assuming that there are such things as persons. Glover believes that we should believe what we tend to believe about persons. In this he also address the question of personal identity, arguing against Parfit (1984) who discussed and dismissed our ideas about persons being stable entities that persist over time and pretty much stay the same over time[2]. Parfit suggested that we would feel liberated and be more prone to act morally if giving up such illusions about persons. Glover on the other hand emphasizes the importance of treating the persons that we are closely interrelated with as persons that do stay the same over time. This is according to Glover plainly something that we have to do to be able to maintain such relationships.

> *Seeing the boundaries between people as less important would have problematic effect on relationships. Part of loving someone is an intense interest in what she is like, together with an awareness of her response to you. It matters that you are with her. If she cannot come, it is not almost as good if*

[2] In this essay I am focusing on what persons might or might not be and the question of personal identity is not to be addressed per se, since it is clear that if nothing exists at all it is irrelevant to ask if this nothing can exist over time.

> *she sends her sister or her aunt instead. The boundaries between people you love and other people cannot seem relatively unimportant.* (1991, p. 168.)
>
> *If, instead of saying, 'She will be dead,' I say, 'There will be no future experiences psychologically connected to her present experiences, I do not find this cheers me up at all. It will not be nearly as good if similar experiences will still occur, though located somewhere else.* (1991, p.169.)

To Glover, the answer to the question on whether we have to base our morals on illusions is that yes, we have to. General ideas such as the boundaries between people *cannot* seem relatively unimportant. This is a moral type of have to, an imperative, an instruction on what we should do. No matter what the true nature of persons is, we have to act upon our intuitions about them. To Glover it is clear that we do need the concept of personhood.

3.3. Moral practice

As mentioned earlier, Farah and Heberlein come to different conclusions about the need for a concept like personhood depending on whether we talk about ethics as a discipline or about everyday moral behavior. When it comes to the latter, this is a context where Farah and Heberlein agree fully with Glover's opinion that we should act upon our beliefs about persons. They write: *"We cannot reprogram ourselves to stop thinking in terms of persons, nor would we want to."* (2007, p.46.)

Though, in stating that in our daily behaviour it is a good thing that we can not stop acting upon our pre-programmed ways of thinking about persons, Farah and Heberlein let their conclusion about moral behaviour rest on two questionable premises. To start with, their view of daily moral practice may be a little naive. They simply assume that there are no moral dilemmas related to our ideas about persons in our everyday lives and that the person system serves us well. In addition they have the idea carved in stone that we can not learn to act against our ideas about persons.

Surely, thanks to our highly automatized brain functions that guide our moral behaviour, our everyday living does not *seem* morally problematic. But might it not be that our trust in our social brain makes us overlook the fact that in a global world our actions do not always have the results that they seem to have? Might it not be that our trust in our person network makes us fail to see when this part of the brain does *not* tell us to act morally and hence does not guide us in the right direction? TV-images of human silhouettes falling from the twin towers on 9/11 can keep an urge to act determining behavior for many years, while just hearing about the fact that 16, 000 children starve to death every day can keep being an

abstract number that fail to activate our social brain and that will most likely drift away from our conscious minds shortly after having been processed, without causing any behavior at all.

Surely, to act as we do when we believe that there are persons around us is often the well-needed cause of moral behavior. The problem is that we believe this to be enough. We allow ourselves to be guided by what we see, by outer features, and we so deeply trust our person network to activate adequate moral behavior that there never *seem* to be reasons to look beyond the imperatives from the categorical person network.

In case our ideas about persons are somewhat illusory and do not even serve us very well in guiding moral behaviour, do we then still have to keep acting upon these ideas? After having given most attention to the nature of persons, I now come to dig into the first part of a question that I wanted to raise in this essay: Do we have to base our morals on illusions?

Farah and Heberlein write that we cannot reprogram ourselves, but since they see no need to do so, they choose not to discuss whether or not we could act against what our person network direct us to do. They do not let us know what their opinions are about to what degree our beliefs about persons determine our behavior. I do, of course, agree on there being many, many cases where we should act like we normally do when we believe that there are persons around us. Like with Parfit's philosophical figure "Claire", I often believe it to be a good idea to leave our predispositions alone. In the case of Claire, these predispositions direct her to treat her children as if they have a certain moral status that let her favor them over other children[3]. In a similar way, it is clear that we should often act as we are predisposed to act on our beliefs about persons, but is this always the case?

The question of what we have to or ought to *do* can be discussed separately from the question of whether or not we can *believe* the right things about persons. Whether or not we can act against our beliefs about others and against the imperatives from our person network is a question that I will here leave for further empirical investigation, but still I find that there is more to be said about what we can and should try to *believe* about others.

When Farah and Heberlein discuss the role of the personhood concept in the discipline of ethics, they suggest that we can believe that persons are illusions and then rearrange our ethical systems accordingly. Parfit also put forth such a trust in our intellectual capacities

[3] In my opinion, this example of Parfit's is in a way rather confusing in that he uses this thought experiment to suggest that one could at any moment t_1 choose what disposition or character to take on. E.g.; should one in a given moment have a character that makes one act in favor of distant unrelated children instead of having the character of putting family first? This is confusing since the choice of what dispositions or character one would want to try to develop, in an Aristotelian way, must in fact have been made at a time t_{-1}, long before the situation that Parfit describes as a moral dilemma.

being able to guide us to true beliefs. (When here referring to his thoughts about personal identity, let us remember that Parfit does believe that there are persons[4].) On the matter of personal identity, Parfit has concluded that there is no such persistent entity as a Cartesian ego or a person that remains the same person over one's life, but after concluding this, he turns around and asks; Is the true view believable? This question is analogical to the one we can ask knowing about Farah and Heberlein's claims about our distorted ideas about the things we perceive as persons. Parfit's answer is:

> *Nagel once claimed that it is psychologically impossible to believe the Reductionist View. Buddha claimed that, though this is very hard, it is possible. I find Buddha's claim to be true. After reviewing my arguments, I find that, at the reflective or intellectual level, though it is very hard to believe the Reductionist View, this is possible. My remaining doubts or fears seem to me irrational. Since I can believe this view, I assume that others can do so too. We can believe the true view about ourselves.* (1984, p.280.)

I believe that Parfit is right about our intellectual capacity for having the true view about *ourselves*, but partly due to what Farah and Heberlein have had to say about our person network, I am still skeptical about our capacities for believing the right things about *others*, and of course this is a relevant question when dealing with moral questions.

Farah and Heberlein imply that in our everyday living we can not have beliefs that correspond to the true nature of persons, and Parfit seem to be of the same opinion when it comes to his everyday beliefs about personal identity:

> *I can believe [the reductionist view] at the intellectual level. […] At the reflective or intellectual level, I would remain convinced that the Reductionist View is true. But at some other level I would still be inclined to believe that there must always be a real difference between some future person's being me, and he being someone else. Something similar is true when I look through a window at the top of a sky-scraper. I know that I am in no danger. But, looking down from this dizzying height, I am afraid.* (1984, p.279.)

Despite it being difficult, and maybe in everyday practice impossible, Parfit believes that we *should try* to find the true view about ourselves. In this I strongly agree. I do believe that the question of how to treat ourselves and thus the future beings that we will become is a moral question and that true beliefs about ourselves are beneficial for us and are thus something that we should strive for.

---

[4] "*On the reductionist view, that I defend, persons exist. And a person is distinct from his brain and body, and his experiences. But persons are not separately existing entities. The existence of a person, during any period, just consists in the existence of his brain and body, and the thinking of his thoughts, and the doing of his deeds, and the occurrence of many other physical and mental events."* (Parfit 1984, p. 275.)

> *The truth is very different from what we are inclined to believe. Even if we are not aware of this, most of us are Non-Reductionists. […] we would be strongly inclined to believe that our continued existence is a deep further fact, distinct from physical and psychological continuity, and a fact that must be all-or-nothing. This belief is not true.*
> *Is the truth depressing? Some might find it so. But I find it liberating, and consoling. When I believed that my existence was such a further fact, I seemed imprisoned in myself. My life seemed like a glass tunnel. Through which I was moving faster every year, and at the end of which there was darkness. When I changed my view, the walls of my glass tunnel disappeared. I now live in the open air. There is still a difference between my life and the lives of other people. But the difference is less. Other people are closer. I am less concerned about the rest of my own life, and more concerned about the lives of others.* (1984, p.281.)

When it comes to the striving for true beliefs about ourselves I do agree with Parfit on this being something that we can and should do, but for Parfit striving to have the true view about ourselves conveniently coincides with and promotes moral behavior and I am not so sure that it is as easy as that. To me it seems things get a bit more complicated when it comes to the question about what we should believe about others. Maybe there are illusory beliefs that we should have to be able to be moral human beings?

As I have come to understand Buddhist psychology and Buddhist moral philosophy, and as it comes forth tied in with Parfit's reasoning above, it is basically very hedonistic and utilitarian and promotes that liberating ourselves from false views and from attachments to our selves and other things will lead to happiness. Empirical evidence suggests that there is much truth to this, research on mindfulness meditation have shown good and fast results and is already used as a tool for improving health at over 240 hospitals in the US, (Bear 2003). Despite these promising clinical results, when it comes to deeper discussions on value theory, happiness and attachments I do see much room for uncertainties that for me have not been filled by the Buddhist approach. To what I understand, attachments - including attachments to other 'persons', is in the Buddhist tradition considered to be something deeply problematic that will always beg for disappointments, stress and suffering. *"In general, liberation was defined as the perception of the emptiness or nonsubstantiality of persons. Monks and nuns would analyse their minds and see them as composities of numerous transitory and interdependent phenomena."* (Blackstone & Josipovic, 1986 p. 29.)

Since I do not have a Buddhist's belief in afterlife, I do not see our detachment from this world or from the people in it as the final goal. Quite the opposite I am wondering if not attachments to certain beliefs about other persons are something that should <u>not</u> be questioned

and dissolved in the quest for liberation. I believe that the question on what beliefs we should have about others still needs to be answered. I am not fully satisfied with Parfit's Buddhistic ideas. (In some Buddhist traditions the helping of others is considered prior to one's own liberation. This might not be the traditions that inspired Parfit's view of liberation as preceding moral behavior.)

Neither am I fully satisfied with the answer given by Glover, Farah and Heberlein. I find it likely that it often is not problematic if we have the wrong beliefs about the nature of other people but I believe that Glover, and Farah and Heberlein are mistaken if they believe that this is always the case. Of course I do along with Glover, and Farah and Heberlein find it likely that we should often act the way we do when we follow our everyday intuitions and beliefs about persons, but as in the cases with non-human animals that I will discuss more in depth further down, there may be moral problems arising along with and due to our everyday beliefs.

When now finishing off my discussion of Farah and Heberlein's conclusions, I am not yet convinced about the superiority of any general principles on how we should relate to and act upon false ideas about persons, and I am not completely convinced about the superiority of any underlying value theory. What I am convinced about is that there is not enough evidence or arguments presented in the article by Farah and Heberlein for promoting a final settlement on a non-person based utilitarian ethics followed by some kind of act-utilitarianism in our daily lives. The reason behind Farah and Heberlein stating that: *"For ethics, the only alternative we can see is the shift to a more utilitarian approach.",* (2007, p.46), is that they do not see the whole picture with all its moral dilemmas and questions.

4. Discussion; alternative conclusions

Inspired by Parfit and assuming that all the glass walls around our beliefs can disappear, I will now all together leave the question about what persons and our beliefs about them *could* be like. I will for the moment believe that any kind of beliefs are possible and I will now turn to discussing what that we have learnt about the person network might tell us about how we should handle these new facts. For reasons following below, I believe that we should: 1) In moral practice be aware of how flaws in our person networks might affect our everyday moral behavior. 2) In moral theory pick the words and concepts for the moral language game with great care. 3) Investigate a new ontological foundation for moral entities.

4.1. Everyday morals is more complicated than it seems

I have already stated that I do believe that our person networks often help us to act morally. What I now wonder is how often this is. What do the new empirical facts tell us about this?

I would suggest that the facts presented by Farah and Heberlein point out potentially overlooked moral problems in our everyday moral practice. Philosophy and science should try to answer question such as: Who are the ones that we should be morally concerned about and are we able to see them? I fear that the person network described often fail to give us the right answers to these questions. The outer world is not always what it seems to be, especially not that world that lays a little bit beyond the environment that includes that most intimate social group that we have been evolutionary designed to focus on and to base our behavior around.

The person network is not based on carefully selected features representing those objects that we have chosen to try to label in moral philosophy. It is based on a number of pre-selected trigger-features. Human-looking faces and human-looking bodies and motions are examples of what we base our everyday moral behaviour on. That these features are doubtful foundations not only for ethics as a discipline but also for daily moral decisions, this is overlooked by Farah and Heberlein. Moral situations involving a human in a persistent vegetative state or a featus with human-like features are examples of situations that go beyond everyday situations, but there are also examples of everyday situations where the having or not-having of human features tends to determine decisions and behavior in a perhaps morally problematic way. One example of daily situations where it is no longer clear how we are to act are situations where we are confronted with other animals lacking the typical human body shape or motions as well as the typical human facial features. Whether or not to put meat on our plates would be an example of everyday behavior that should not be directed by innate predispositions and frameworks *if* these are misdirecting in a morally relevant way.

Are we generally rightly guided by our neural systems? Are the person networks usually serving our values? It is of great importance to ask these questions. Farah and Heberlein state that: *"Although the concept of personhood may be bad metaphysics and better suited to an earlier world, even today it serves us well."* (2007, p. 46.) Is this true?

As so often in morals this is fundamentally a question about who the ”us” is in the expression ”serves us well”. Consider this quote by the Jewish philosopher Adorno: *"Auschwitz begins wherever someone looks at a slaughterhouse and thinks: They're only animals."* (Quoted in Patterson, p. 53.) I propose that this quote divides people into two groups as sharply divided as the person vs. things categories in our heads. I believe the two different kinds of reactions to this quote by Adorno support Farah and Heberlein’s description

of an irrepressibly categorical type of thinking. Reactions to this quote tend to show a complete lack of nuances. Few people would reply with degrees of 'yes there is probably some truth to that'. Either people see an almost (or explicit) divine line between persons and animals, and hence find the quote infantile and still extremely provocative; or the quote will make people sense an important similarity between the persons around them and the non-human entities at the slaughter houses and hence find the quote horrifyingly mind-boggling. Depending on which of these reactions the quote evokes it can single out two different attitudes that tend to affect our everyday behavior in fundamentally different ways. It is of the greatest importance to determine which one of these two categorically different outlooks that most resembles the truth about the natural world.

Robbins and Jack (2006) nicely depict the subsystems that we use to understand the social and physical world. What that happens if the phenomenal stance is excessively employed when we interpret our surroundings is what Robbins (2007) labelled "benthamism". I believe that this could be exemplified with a Jainist or strict vegan showing great moral concern for insects. An analogy would be the child making the same kind of mistake of over-extension when using the phrase "Fido" to talk about all 4-legged animals. The even extremer form of mentalising without discrimination would be animism. Here the social brain is running wild and can no longer distinguish between any differences or degrees of mindedness. To make such a mistake is to loose all ability to give the special kind of moral concern to those special kind of mentally equipped entities that should be treated in other ways than we can treat stones and cars. In our modern western society animism is unlikely to be what determines our everyday behaviour. What is more likely in our society is that the tendencies of dualistic thinking stemming from the Christian division between godlike humans and automata-like animals will reinforce the categorical labels forced upon the world by our person network. To overlook what kind of catastrophic results a daily action pattern based on such instinctual beliefs could have, *if faulty*, is naive. The risk would be that we did not see the ones that are right there around us or on our plates.

Our person networks might not guarantee moral behavior. Another question is, will functional abnormalities in our person network guarantee immoral behaviour? I believe that this is something that is commonly but falsely believed. Farah and Heberlein mention how autistic children have been described as "treating people like objects". Autistic children tend not to be very interested in other people and according to Farah and Heberlein they generally prefer to interact with inanimate objects. I wonder if *inanimate objects* should here be

exchanged for *non-person-like objects*. Surely autistics have difficulties with so called normal social interactions, but I am not sure how characteristic a choice of interacting with inanimate objects is for autistic individuals. What I do know from reading the book Animals In Translation, Using The Mysteries Of Autism To Decode Animal Behavior, written by the autistic woman Temple Grandin, is that autistic individuals surely can have a great interest in and profound understanding of non-human animals, i.e. in non-human but animate objects. While suffering from difficulties to understand implicit social, (but as far as I know not necessarily moral), rules for interactions between human individuals, Grandin is working as a researcher, gives talks on animal behaviour and works to improve the conditions at American slaughter houses. When it comes to how her abnormal neural wiring guide her actions towards non-human individuals, Grandin might possess an advantage improving her likelihood to act morally compared to those whose actions are guided by a person network! For sure a dysfunctional person network give rise to social difficulties and being autistic makes it hard to interact with other humans as the others would like the autistic to act. My point here is that if there are situations, e.g. where non-human animals are considered, where someone with a dysfunctional person network would more easily reach morally preferred conclusions, then this gives reason to question our automatized reliance on and everyday complete trust in this network.

How are we then to know how we should act towards entities possessing different degrees of mindedness if we cannot trust our preprogrammed systems for understanding the world? The answer to this might be found in the article Animal minds, animal rights and human mindreading, where Mameli and Bortolotti suggest that:

> *Current knowledge about the evolution and cognitive structure of mindreading indicates that human ascriptions of mental states to non-human animals are very inaccurate. The accuracy of human mindreading can be improved with the help of scientific studies of animal minds.* (2006, p. 84.)

Where our person network is at risk at failing in guiding us in the right directions, our everyday behavior should be directed by a moral that is based on scientific investigations of the natural world and the different beings around us. But our possibilities to understand the true nature of those who do not look like us are not only determined by innate predispositions causing prejudices that can only be fought by hard-core science. Besides a biological make-up and beliefs gained through scientific work, there is also social learning. Cultural determinants and behavioural/cognitive plasticity have often been overlooked by natural scientists. Mameli and Bortolotti (2006) do not at all discuss the importance of social learning. Their conclusions

about humans' inaccurate judgements of the mental capabilities of other animals might be too much influenced by experiences that have arisen in a society where humans do no longer share their environment with individuals from other species. That our moral beliefs about those who do not look like us are not determined when we are newborns is proved by that fact that humans in different social contexts develop different attitudes. Humans living with pets can come to look upon these particular animals as persons and family members while their attitudes about factory animals might have developed into cultural specific ideas about these types of animals being more like inanimate objects[5]. In those days and areas where people lived/live among animals the understanding for (i.e. the correct judgements of the mental capacities of) these particular animals might have been greater. Considering cultures where the godlike-person vs. inanimate-objects distinction has not been emphasized by religious beliefs suggests that this is the case. Other beliefs than the human-centered one are possible. It is well known that Buddhists do not separate human suffering from animal suffering the way we do in the Christian-cartesian tradition and an Inuit who would have to kill an animal for survival would strive for a respectful treatment asking the animal for forgiveness so that the animal's soul will not come back for revenge, (Fine 2000).

Which beliefs are the true ones? Which ones should guide everyday moral behaviour? These are complex questions with implications beyond our everyday worlds where we live embedded in habits typical for some specific human culture. Humans do make many inaccurate judgements about other animate objects, i.e. about animals, which is pointed out by Mamel and Bortolotti (2006). But these inaccuracies are not limited to the cross-species problems. Evolution does not, as Mameli and Bortolotti imply, operate on a species level but on a gene level. We are evolved to pass on genes, that is, we are evolved to first and foremost care for those in our own group and those who are the most genetically related to ourselves. The moral disasters that are prone to happen due to such predispositions are known to us all. Luckily, innate predispositions are not the only determinants of behavior.

Our values and our intellectual understandings are, as Parfit and others have shown, open for adjustments and are examples of determinants that can stretch our alternatives away from evolutionary and biologically predisposed options. Our innate neural subsystems guiding our actions are and should be subordinate to our ideas about what that is valuable, but this

[5] If some humans do look upon pets as persons and distinctly different from farm animals, such a cognitive divide and lack of thinking in terms of similarities in degrees could give further support for the categorical thinking described by Farah and Heberlein. It would be interesting to see if the pets' faces would activate the face area of their owners' brains while an unknown pig's face would fail to do so.

may take some thinking that goes beyond those innate tendencies that want to guide our everyday living in certain ways.

4.2. We need moral concepts that are free from prejudice

Albert Einstein once wrote:

> *A human being is a part of the whole called by us universe, a part limited in time and space. He experiences himself, his thoughts and feeling as something separated from the rest, a kind of optical delusion of his consciousness. This delusion is a kind of prison for us, restricting us to our personal desires and to affection for a few persons nearest to us. Our task must be to free ourselves from this prison by widening our circle of compassion to embrace all living creatures and the whole of nature in its beauty.*[6]

Can all our moral dilemmas be dissolved if we just expand our circle of compassion and develop inner peace, love and cross-border understanding? Well the problem is, as for example Glover stressed, that we seem to need and according to Glover should, separate out distinct entities that are to us special and intrinsically different from the rest of the universe. Einstein's ideal and Parfit's extinguishing of the glass walls around us might help ourselves to greater mental health and harmony, but do not give us any guidance in if and how to recognize entities around us that beg for special kinds of moral concern.

Persons are important in our everyday lives, but the moral concept of personhood is, as we have seen, built on prejudices. We are wired to believe that there are persons that are categorically different from non-persons and we are prone to assume that the relevance of personhood always comes together with, but never comes without, the appearance of human-like movements, shapes and facial features.

There are different theories on how our brains go about creating concepts. Many concepts are developed through an adding up of the knowledge learnt about the objects that the concept refers to. Then our brains end up with prototypes that are an average of those objects that we have experienced. Categorization can also be based on typical exemplars. Other theories describe the association of certain features with members of the category to be essential when we create concepts, (Reisberg 2001). I is likely, and I believe that animal studies suggest, (Wynne 2001), that there is truth to all different theories of categorization and that different concepts are construed through different kinds of cognitive processes.

[6] http://www.wisdomquotes.com/000762.html

Maybe the concept of personhood is too much associated with those trigger features that have never been shown to be relevant to personhood. When feature-based concepts are determining categorization there is no allowance for degrees of membership since those who have the features are placed within the category and the ones that do not are seen as a non-members. To me it seems like feature-based concepts are more simplistic than for example abstract concepts, and that 'person' is an example of such a concept. I believe that our ideas about 'persons' rests on a concept that have arisen early on in an evolutionary perspective and is driven by a somewhat primitive form of categorization. It is often the case that concepts that are derived from innate forms of knowledge are harder to adjust and less open for influence from conscious processes then abstract concepts. Therefore I take the evidence for the existence of a fairly rigid and simplistic person network as pointing out a need for a new conceptual framework in moral theory, a conceptual framework that is deliberated from prejudices and that contains new concepts that allows for usage in complex types of reasoning.

Farah and Heberlein do come to a similar final conclusion:

> *The concept of "person" comes with heavy baggage from our intuitive ways of responding to humans and other entities, and it is unclear whether we can free ourselves of the grip of, say, faces when we deliberate about the objective status of a candidate person. For this reason, we advocate focusing on the psychological traits themselves rather than a concept of person defined by those traits (and as yet unsatisfactorily defined).* (2007, p.62).

My conclusion differs from the one of Farah and Heberlein's in its second part. For reasons similar to the ones described by Glover I am not yet ready to finally give up on the concept of personhood or rather, on the need for some similar but maybe more well-functioning, less stereotypical concept for referring to special moral objects. Glover and I would here share the same view, that in order to recognize differences between morally relevant entities, we do need to be able to label and describe such entities. We need to relate to and label these moral objects, but maybe there is some inner entity to refer to instead of a metaphysical concept triggered by outer features.

A sentientist and utilitarian might not have to be concerned about any lasting moral entities nor about anything else than the facts that there are moments when someone senses something pleasurable. He or she could be concerned with acting to maximize the number of those moments, (or the number of entities that allow for those moments), to decrease the number of unpleasant moments, and to increase the intensity of the former as well as to decrease the

intensity of the latter. If the someone that senses something has changed into someone else in the next moment, and if nothing makes this someone belong to one and the same entity over time, then this does not matter for the sentientist utilitarian[7]. In the new now the new someone is the one for us to be concerned about, not some metaphysical substance that can be described by further facts and that might be supposed to be persisting over time.

At first it might seem like Farah and Heberlein and other sentientist utilitarians provide a good enough base for morals; we can be morally concerned about sentient beings and act upon these concerns in every moment. But still it does seem like the sentientist utilitarian runs into problems when considering a person in a persistent vegetative state, or in a coma, or just plainly being in a deep state of dreamless sleep. Glover's concern for those whom we have special relationships with and Parfit's fictional figure the mother Claire who wonder if she is right in giving special concern to her own children would be other examples where our intuitions might tell us that the utilitarian can not provide a satisfying answer to the question of how to act. In these cases it seems like we need to assume that we are confronted with unique and lasting morally relevant entities. To be able to claim that it does matter if we let the sleeping person wake up in the morning, instead of replacing him or her with someone else as capable of sensing the same things in the upcoming moments, we have to assume the existence of some kind of morally relevant object to relate to. Another example of the same type of moral question, that might seem more real and easier to grasp, would be whether or not some utilitarians are right in claiming that we are doing no wrong in killing farm animals, (or pets), for our own pleasure of eating meat, as long as these animals are replaced with new animals feeling at least the same amount of physical and psychological wellbeing as the ones we killed. Someone claiming that farm animals are also 'persons', would of course not agree on this.

Facts presented have shown a need for a deconstruction of the concept of personhood. Still being in a need for moral entities to refer to, we will now, guided by our empirically based knowledge, have to reconstruct those concepts that we need.

Could we not just treat personhood as a purely metaphysical concept? Would not the best approach be to altogether try to avoid confusing moral concepts with empirically derived knowledge? No, the most important message in Farah and Heberlein's article could be that it has to be clear that moral philosophers are also affected by their person networks. Actually,

[7] Though as already have been pointed out, there are utilitarians in favour of the person-concept.

when perceiving outer objects there are always inner cognitive structures, innate or more lately developed, determining how we understand what we see. Empirical evidence supporting this fact was nicely summarized in an article in Scientific American where the author concluded that; *"It is no longer possible to divide the process of seeing from that of understanding."* (Zeki, 1992 p. 43.) Of course this now empirically backed up knowledge is not new to philosophers. 'Concepts without percepts are empty; percepts without concepts are blind', as Kant put it 1781. (Quoted in Hollis 1994).

What the new empirical facts about the percepts that we label persons point out, is how great a role is played by our innate beliefs about what persons must be like. These facts beg for a more cautious approach when aiming at making progress in the refinement of the philosophical concept of personhood. It would be easy to fall prey to prejudices created by the human perception system, and this is, I believe, what that has already so often happened in the hunt for defining criteria of personhood. The search for defining criteria has been effected and affected by preferences for certain shapes and triggers. Philosophers discussing what a person is might have started off with thinking about persons and therefore automatically getting images of what persons generally look like and then he or she have tried to imagine what psychological characteristics these types of beings generally possess. Due to such philosophically doubtful procedures we end up with such suggestions of defining criteria as Dennett's demand on capability to use language or to be *"conscious in some special way not shared by other animals"*. Although no moral philosopher could seriously claim facial features or bodily shapes or movement patterns to be morally relevant, these characteristics or rather the human-looking archetype associated with such features, might be the only things that the different beings depicted by a vast number of defining trials would have in common. Concepts can be refined. A feature-based category can develop into a more complex perhaps prototype-based perhaps semantic concept. The problem is that despite such plasticity of concepts the archetype of for example personhood seems to be irresistibly connected to and based upon those trigger features that the first pre-verbal concept of person was based upon already when we were newborn.

Not until our inborn prejudices have been made explicit will we be able to steer free from neurological pitfalls. When this has been done the construction or reconstruction of the concepts that we need can start and a philosophical discussion liberated from innate frameworks can proceed. Arbitrariness can now be ruled out thanks to defining criteria now being linked to those psychological criteria that are morally relevant.

Still it is not obvious which words and concepts to use. Even if the status of personhood should not be arbitrarily based upon the membership of one certain biological species or on certain physical traits, our perceptual networks might never allow us to completely drop the common use of the person-word as referring to exclusively human individuals having certain facial or bodily features etc. Because of this baggage the very word person might turn out not to be the best one to use in moral discussions. If so, what we need will be new words with certain specific morally relevant meanings and implications, words that can be used parallel with and independent of the folk psychological concept of persons. We should start with stating what that is morally relevant and then create the appropriate frame-work for talking about these morally relevant characteristics or entities that we would like to talk about.

If we want morals to be liberated from neurologically determined prejudices about what a certain morally relevant entity must look like, then maybe we should use concepts that do not trigger too simplistic and potentially automatically misleading ideas or images. A discussion about which words and concepts that we need could be driven by a search for a truthful ontological description of those entities that we need to talk about and relate to.

4.3. Look for ontological foundations for moral objects

Talk about persons is probably unavoidable, and most of the time that certainly is a good thing; thinking and talking about persons often makes us recognize and act in good ways towards those human beings that we see around us. The problem with the word person is that it, at least in our culture, is a word too loaded with sometimes misleading associations. The association to a perhaps divine dividing line between 'persons' and on the other hand 'things', have definitely had a problematic impact on the forming of our morals. Further, in the history of categorical thinking about persons, personhood has been used synonymously to 'moral objecthood' so that a biological being has either been looked upon as a person and hence seen as worthy of moral concern, or has been 'just an animal'. Those not falling within the person category have for example often been denied the right to their own lives. Contrary to these categorical intuitions, moral objecthood seem to vary in degree; we should show more concern about complex human beings than about some less conscious animal. Hence we need to talk about moral objects without having to rely on the person concept that is for neurological and historical reasons too categorical and simplistic. What kind of entities is it that we *try to* refer to when talking about persons? What kind of entities are moral objects?

I will now discuss what kind of entity it might be that we should try to refer to when using the person concept in morals. I will do so by first showing one way that I disagree with

Parfit, and secondly by pointing out what I believe to be another common problem with many other philosophical ideas about persons; the use of one and the same concept (person) for referring to moral objects as well as for referring to moral agents.

Parfit seem to suggest that subjective experience is enough to qualify for personhood. When discussing the cases of people having had their brain hemispheres separated in clinical attempts to limit the effects of severe epilepsy, Parfit states the following concerning the fact that such patients show evidence for having two streams of consciousness:

> *We might claim that, in such a case, there are two different people in the same body. /.../ Our alternative is to claim, about these actual cases, that there is a single person with two streams of consciousness. /.../ If we believe that the unity of consciousness must be explained by ascribing different experiences to a particular subject, we must claim that in these cases, though there is only a single person, there are two subjects of experiences. We must therefore claim that there are, in a person's life, subjects of experiences that are* ***not*** *persons. It is hard to believe that there are really such things. (1984, p.276.)*

Contrary to Parfit I have no problem believing that there are subjects of experiences that are not persons. Any conscious being is a subject of experiences. What it means to be conscious is to have a subjective perspective where it is like something to be that being. Consciousness as such is not to be mixed up with reflexive consciousness, which is consciousness about conscious contents and which allows for self-reflexive consciousness and self-awareness. Conscious, but not necessarily self-conscious, beings can take notice of what that is happening to them. They can be affected in good or bad ways and hence they are moral objects. The split-brain patient in Parfit's example might be hosting two such moral objects.

Mere consciousness will give at least the sentientist an ontological base[8] for moral objects. This would be a good start for the building of an ontology of moral objects. Though, I believe

[8] Is this true? Is consciousness real? That is real which has real causes, said Spinoza, and if consciousness is not a natural category what is it then? Just the steam of the machinery says the emergentist. Though, at a first glance evolutionary biology seem to suggest that any biological phenomena, and consciousness has been suggested to be a biological phenomenon, has beneficial effects for the carriers of the genes that brings it about, i.e. that consciousness must have some beneficial causal effects on its carrier. However, this does not have to be the case. Evolution has no goal and whatever phenomena that does not significantly decrease the chances of its carrier's survival, can keep on being inherited to new generations. Therefore, even if we do put an evolutionary perspective on the discussion, it could be true that consciousness is just a phenomenon without causes. It could be that we are conscious just because we didn't die of it. Though, if nothing else, it seems rather unlikely that consciousness itself is just another blind gut, an illusionary one. Many, or most, psychological phenomena as well as bodily appearances seem to have functions and consciousness has been suggested to have many specified functions (Baars 1997). Consciousness does something and therefore consciousness is something.

that, like Parfit, many thinkers have been focusing too much on the thought of a more or less currently present stream of consciousness when debating criteria for personhood. When asking whether or not some being in a deep coma is a special kind of moral object that we would do wrong in killing, we are not primarily interested in whether or not the being is conscious. When talking about persons we refer to something more than beings that are momentarily conscious. Since consciousness is repeatedly interrupted, it is no good base for the entity we refer to when talking about persons. When considering patients in persistent vegetative state, or sleeping, we see that it seems to take more than a stream of consciousness to make up a person. Maybe after all the question is not; Can they suffer?[9] but; Will they be able to remember their suffering? This would not be a question only suitable for a cynical anaesthesiologist, remember that to determine if we are dealing with a moral object in the first place, it is essential to know if some being is conscious. The; Can they suffer? question will always be morally relevant. Though, to ask what suffering that will be remembered could be essential in figuring out *what type* of moral object that we are dealing with. I will soon try to argue for self as a beneficial concept to use when discussing special types of moral objects. A self contains memories, plans and much more. A being with a self will definitely not be a morally irrelevant 'thing', neither will it be a mere conscious being that can suffer but not remember its suffering or have ideas about how to avoid future suffering, nor will it necessarily be a 'person'. A being with a somewhat complex self will be caused some harm if being killed and should be shown greater moral concern than a mere conscious being.

Many definitions of 'person' are tying onto Locke's old definition in that they try to single out some kind of entity that is aware of its own existence over time, it is aware of itself as a distinct entity. Singer (1979) for example, use 'person' in the sense of a rational and self-conscious being, and since Singer is a preference utilitarian, for him it is the capability of having desires about the future that makes these two criteria important. The problem here is that self-consciousness is an even more instable entity, or state, than consciousness itself. Being self-conscious and having certain preferences could be morally relevant for the moral object having them in the moment the being is self-conscious, but is this moment of self-consciousness relevant for some specific lasting entity? To be relevant for a being now *and* in any later moment there must exist such an entity that will not only *sense* itself lasting over

---

[9] *"The question is not, Can they reason? nor Can they talk? but Can they suffer?",* Famous quote by Jeremy Bentham (1748-1832), e.g. in Singer 1979 p. 50.

time, but will also in reality be an entity that does lasts over time. At least if we want our description of the moral entity that we try to refer to, to actually depict something in the real world. Is there something, and if so what is it, which makes a morally relevant entity to be the same morally relevant entity at some later moment in time?

Here we are back in the good old debate of personal identity. I believe, and will below try to argue for, that it is the *having of self* that is the morally relevant characteristic that we should be focusing on. The clinical psychologist Cullberg (1999) has described the self as the subjectively experienced side of one's own person. This type of talk brings attention from outer features and the person's body to what that must be relevant for determining degrees of moral objecthood – the inner experiences. A self, I will claim, can exist in degrees and is something that is stable over time, and when considering for example the case of someone in a coma or someone sleeping we need to ask what that is the same when the person wakes up and what that is interrupted and killed if the unconscious someone is killed. To me, the fact that those outer properties that make someone look like a human, properties that will activate ideas about a person, is not enough of a morally relevant foundation. I am interested in the morally relevant inner characteristics of this someone who is at the moment unconscious. This someone might have had an accident and lost all capabilities of using human language and might not be a moral agent or worthy of praise or blame etc, but all these characteristics that have been used as defining criteria for personhood is not what I am interesting in when asking what that makes this someone into a special kind of moral object. This special entity with its special characteristics must be able to last over time. Self-consciousness, language or agency can come and go during the days.

Self might actually be what thinkers like Locke and Singer are intuitively referring to when talking about self-consciousness. A being that is able to have moments or periods of self-consciousness is able to develop a more complex type of self than mere conscious beings with more rudimentary selves. It is the complex self of a being that *can* be self-conscious that Singer should be focusing on rather than the capability of having self-conscious moments per se. Shifting focus from the person concept to the self concept will also be a way to push for more complex reasoning about non-human moral objects without violating our by the person network shaped intuitions about persons. Humans are and should be extra important to us, but we need to be able to talk about moral objects who are not human-like but still are stable and lasting entities who exist in their own right. What is important about these moral objects are not that they are self-conscious beings who are aware of themselves as distinct entities, with a

past and a future (Singer 1979), but that they, maybe thanks to some self-conscious processing[10], *are* distinct entities with a past and a future.

'Self' is an often poorly defined concept but despite this fact it is the necessary, (although slippery), stepping-stone for discussions about self-awareness and self-reflexive consciousness. A lot of attention has been drawn towards interesting studies trying to answer questions about what sort of creatures that are self-aware. Important to remember here though is that discussing the concept of self is not the same as discussing self-awareness or self-consciousness. The self itself, the something that awareness gets directed towards, must reasonably be present before awareness about it can arise. Self is unlike self-consciousness and consciousness an entity that continuous to exist over time. Like players on a soccer field that do not cease to exist when the game is over, the self does not go away when the self-consciousness is turned off over night. Now we know what entity to look for and one recent attempt to approach the concept of self empirically consists of Northoff's et al. (2005) searching for neural correlates of the self. When taking a closer look upon activation patterns in the brain, we are faced with close relationships between emotions and for the research subject self-related items.

What I will soon argue for is that it is this self itself, an entity that we can approach empirically, that is the morally relevant entity that we should try to single out as a special type of moral object that begs for special concern. Before getting to this I will just point out the earlier mentioned other common source of confusion in the personhood debate.

Besides the somewhat hastily treatment of consciousness in the discussion of the ontology of moral objects, I believe that also the role of moral agency has confused the personhood-debate. A few thinkers, for example those who have been striving for ascribing personhood to beings of non-human species, have not seen moral agency to be relevant for personhood. Contrary to such uses of the person concept, many other lists of defining criteria have included precisely the demand for moral agency. Rationality and capability to be concerned with worthiness of blame or praise are examples of such suggested criteria that relates to such demands on agency. I believe that such beliefs are deeply intertwined with the everyday use of the word or concept person and to use 'person' synonymously to moral agent might not cause the same problems as linking 'person' to moral objecthood. The everyday meaning of

[10] The capability of being self-conscious or to be able to at least momentarily reflect over ones own behavior, might very well be the mechanism that is crucial for developing a complex self. Hence an entity having this mechanism could be the type of entity that we would cause harm if killing, but this is a suggestion that I will not have room to investigate further in this essay.

the word person does often seem to involve agency and therefore I will now leave the metaphysical concept of personhood for those who are more interested in discussing metaphysical questions of agency, blame and free will etc.

There is one important difference between those who act in morals; the moral agents or moral subjects, and the ones who those agents should be concerned with; the moral objects. I wanted to investigate entities that are moral objects of a special kind, and I do not want to confuse this quest with the search for criteria for moral subjects. I am interested in the concerns of Glover's and I now want to return to his intuitions:

> *The undercutting of relationships would be a disaster. But, if people did fade fast, the disaster would be unavoidable once we saw things clearly. But there are two reasons for thinking people do not fade so fast. One is that many of the features that contribute to a person's distinctiveness (such as likes and dislikes, or style) seem usually fairly stable. The other is the importance of the inner story which spans our lives. Through mutual recognition, we can share the creation of our inner story with others. So we can change in partly interlocking ways, and avoid the disaster of our commitments and relationships fading away.* (1988 p. 169.)

My intuition is that what that Glover is referring to as a somewhat stable entity is *the self.* What that we are morally concerned with here is not the human body nor its' facial features or movements. We are concerned with a special part of the, in lack of a better word, person. This special part can be described as self. For sure Glover's paragraph above could be associated with a few other concepts; personality, identity, narrative and interrelatedness, but these concepts can all be treated as intimately related to or parts of the broader concept of self.

The inner story that Glover writes about has by many others been referred to as the narrative self and the mutual recognition and interrelatedness can be referred to as the social self.

A person's distinctiveness or its' likes or dislikes and style could be talked about in terms of personality or identity. Personality has to do with traits that show some kind of stability over a long period of time. Certain traits, it has been debated how many, show consistency in cross-cultural research. That is, some personality types, or rather, traits, are found all over the world. Personality research could even draw on cross-species claims where studies of other animals bring understanding about some basic behavioural tendencies. A lot of personality research is based on evolutionary psychology and can show for example why a bold fish have certain beneficial tendencies in a certain environment whereas a curious fish gain other benefits in another less dangerous environment (Sneddon 2003). Personality has the same kind of influence on the individual self over a life-time. Identity, could be looked

upon as a sub-category of self-reflexive consciousness. Not that our identities are at all conscious for most of the time, but they seem to have been originally construed by conscious processing. This would mean that it takes that there have at moments been self-conscious processes going on for an identity to be developed. The term identity is important within the social sciences and there it basically refers to the individual's comprehension of himself or herself. While our identities are construed our personalities often seem to have stronger ties to innate tendencies. Of course there are also more philosophical concepts that can easily be associated with what Glover writes about distinctiveness, style and likes or dislikes. I am here thinking about character, the tendencies to act morally as discussed by Aristotele, or the dispositions to act in certain ways towards for example ones own children, as discussed by Parfit. My reason for bringing up these different concepts from somewhat different disciplines is that I believe that all these concepts are to a great extent parts of the same greater whole, of what we are talking about as 'self'.

The social self, the narrative self, identity, character and to a great extent personality are parts of the self. What Glover seems to grasp for is this self, a self that is organized on different levels and that is to some degree consistent and lasting.

The self has been described in many different ways, often without being clearly defined. In contrast to those vague descriptions I find Bernard Baars' approach to the matter to be a useful one, based on and consistent with a variety of research from the areas of neuroscience and psychology. Baars depicts a self organized on different levels. His idea about self is 'self as context', i.e. self as a type of underlying information that is shaping our conscious moment to moment experiences. *"The "self" of everyday life can be seen as a context that maintains long-term stability in our experiences and actions."* (Baars 1997, p.142.)

Self is not any kind of context. *"Unconscious and involuntary processes do not mandate a connection with self. "We" do not acknowledge unconscious knowledge as our own, and "we" disavow responsibility for slips and unintentional errors. Yet self is not something we experience directly, as we might experience a musical phrase in a song."* (Ibid.)

Evidence for 'self as deep context' is that loosing a relative or a close friend can feel like loosing a part of oneself. Here the interlocked social self gets shaken and changed. Self is a kind of deep context through which we experience our lives. Another source of evidence for this being true, is the fact that when our life goals are interrupted, we are affected deeply and fundamentally. We might sometimes have to reorganize ourselves so that the parts and levels

come together again. An integrated self can help our inner worlds to stick together and to be understood as united, even when the most diverse kinds of experiences and circumstances may seem to 'tear us apart'. Such integration can be made easier thanks to Glover's 'inner story' or 'the narrative self'. *"At the highest levels of organization we encounter a kind of self that neuroscientist Michael Gazzaniga has dubbed "the interpreter". Unlike the sensorimotor self, the interpreter engages in a narrative, and therefore involves the speaking center of the left hemisphere."*(Baars 1997, p. 147.)

One could hypothesize that for being functional and efficient, any consciousness would have to rely on unconscious context-like structures guiding behaviour. Where there is consciousness there is likely to be unconscious context shaping this conscious element. Rudimentary consciousness would thus be likely to have evolved in an interrelationship with a rudimentary self. Self and consciousness are closely interconnected. It could be that anything that can act to increase its own likelihood to persist over time is likely to have a self. Though, Baars does define his self as deep context rather narrowly, claiming that only some context is self.

> *One way to think of "self" is as a framework that remains largely stable across many different life situations. In theater terms,*[11] *the implicit self seems to involve the deepest layers of context-the most basic expectations and intentions that guide our lives. Like any context, self seems to be largely unconscious, but it profoundly shapes our conscious thoughts and experiences. It seems to work behind the scenes of the theater, pulling invisible strings to control the spotlight, shaping the actions planned and carried out with the aid of the theater, and to some extent perhaps, the actors themselves.* (Baars 1997, p. 145.)

Baars' description of self could be a description of a special kind of entity that is somewhat stable over time and that is morally relevant. It could be a description of a certain kind of moral object that separates those beings that have self from those beings that are merely conscious. It clearly depicts an entity that can vary in degree and complexity among different beings. Even though some attention mechanisms[12] and possibly rudimentary consciousness have been observed in such relatively simple animals as fruit flies, we would not believe them to have more than at most rudimentary selves. Fruit flies are not likely to have the type of deep, complex and layered context described by Baars and thus you would not act morally wrong if killing fruit flies, (at least not if replacing it with new as sentient fruit flies). The self

[11] Read more in Baars' *In the theater of consciousness: The workspace of the mind*.
[12] Fox, 2004.

is a special type of context that may be accessed by the conscious I. This view of the self might become clearer if we, like Baars, consider a quote by William James. *"The total self (is) partly known and partly knower, partly object and partly subject. ... we may call one the Me and the other the I. ...I shall therefore treat the self as known as the me, and the self as knower, as the I. ..."* (1997, p. 142.)

To further relate this to what has been discussed earlier, I would like to suggest that Glover has been stressing the importance of the, to use James's terminology, 'Me', whilst Parfit has been talking about James's 'I'. This distinction between the conscious I and the Me - that which can possibly be accessed by consciousness, singles out two referents; the mere consciousness and the persisting psychological self. Both of these are morally relevant and at least one must be present for us to be able to talk about a moral object. In case of someone in a persistent vegetative state, there must be a sleeping self and there must be a possibility for this self to at some later moment again be consciously experienced. Friends to someone 'suffering' from a complete lack of personal memories do not have a moral duty to keep acting as if the now inaccessible (part?) of the self is still there.

I have tried to single out what is, and what is not, the morally relevant self that is the concept and entity that I am looking for as an alternative to the problematic concept of personhood when wanting to talk about moral objects. Baars has stated that: *"The "self " of everyday life can be seen as a context that maintains long-term stability in our experiences and actions."* And he continued; "*Over many different situations we still manage to maintain a sense of predictability about who and what we are. A review of "disorders of self" such as multiple personality disorders shows that any fundamental changes in one's expectations and intentions are experienced as self-alien."*

As a last take on the subject of self, I will now test my intuitions of self as a morally relevant entity on the cases of Multiple Personality Disorder, or what that is nowadays describes as Dissociative Identity Disorder. DID is a disturbance of the usual development of the self, in many cases caused by traumatic childhood abuse. This disorder, or unusual order, is understood to arise when the child enters a dissociative state in order to cope with traumatic events, and this state over time develops into a complete and differentiated self, (Baars, 1997). Far more than one alternative subpersonality, or as more often labelled today, alternative self, or simply 'alter', can develop.

The experience of unity, of having one self is a common experience and, I would say, is likely to be a real phenomena with a neural correlate. The same thing should be the case with

the uncommon experience of having more than one self, but the DID diagnosis has been a hot topic and by some not considered to be a real phenomena. Even clinicians who must have had much confrontation with strange or un-normal psychological phenomena have labelled the diagnosis absurd. For sure the idea of many selves in one body is contra-intuitive but let us consider the idea that those who describe their experiences of having multiple selves or alters to be speaking the truth. What would happen with the 'self as morally relevant entity' hypothesis if tested against these unusual experiences? Would there be many moral objects in the same individual body?

Consider the following quotes by DID survivors:

"We hate the word integrated. We work as a team and will stay as a team."-R.C., (Cohen et al., 1991 p. 170).

"Intgration is something I look forward to, although my parts would take issues with that! But just as the patterns of dissociative behaviours lasted for many, many years, it is understandable to me that it will take a while to unravel the reasons my mind originally "divided." It will take time to discover the role each of us had to play in our survival. Each part was a victim of traumatic betrayal. As those traumas are given time to heal, I think we will eventually know each other better and finally "come together".
-Vickie G., (ibib, p. 170-171).

"Who am we really and how do I talk about myselves? English has no cupped hands to carry my meaning."-Gregory B., (ibid, p. 174).

"What I wish my spouse had known:
[...]
That each alter really thinks he/she is an independent being." -Charlie Anderson, (ibid, p.192).

Experiences like these described by DID survivors could suggest that a self really is a morally relevant entity that should be given a special kind of concern that goes beyond the concern we show mere conscious beings. DID survivors express a wish for each part, each self or alter, to be treated morally. But still, will not the notion of self as a morally relevant entity only raise new dilemmas? Should someone who deliberately kills a someone with multiple selves have the sentence multiplied by the number of alters that used to live in the brain of the victim(s)? What about the liberated self? Does Parfit abandon himself when letting go of self? What is it that is liberated from what? To me the answers to these questions actually seem to go along with my other moral intuitions rather than create new dilemmas. I believe that each alter or each self appearing through consciousness should be treated with respect and I believe that in the case of a liberated self the moral question on how to treat this self might have dissolved.

The typical human clinging to self can give raise to a lot of unnecessary suffering for the conscious 'I' that tries to protect its ego, (not the Cartesian ego but the Self as has been discussed here). So then, is liberation from self something that one should strive for? Well, liberation from the self might resolve this question in quite the same way as when someone who undergoes an irreversible complete loss of personal memories will not have obligations towards the lost (part of?) self, because a self must have the ability to once again get access to consciousness for being morally relevant. The self of a person in a coma is a lasting and morally relevant entity *if* there is hope for this self to once again reach some conscious state. It takes a conscious subjective experience to make a self meaningful. When liberated or irreversibly separated from the capability of having access to a conscious 'I', there is no longer anyone to whom the question of what to strive for can have meaning; there will only be the mere stream of consciousness. To the self-less Buddhist monk, liberated or detached, only the present moment would have value and the clinging to the faith of some lasting entities would be perceived as morally irrelevant and even selfish. Still, it is reasonable to believe that this liberation from self would consist in the non-attachment and non-ascribing of value to the self but that even the perceptions and insights of self-less monks or nuns would come about through and thanks to the somewhat stable contexts of a self. Hence even the one liberated from self could have or be a self worthy of moral concern and then should be treated accordingly even if they themselves, like Farah and Heberlein, would feel that the question of morally relevant entities has dissolved.

What about integration of different selves in DID? Is integrating through a therapeutic process the same thing as killing the separate selves? Will this not be the absurd conclusion that will at last rule out the possibility of using self as a morally relevant entity? No. For some person, or rather, for some conscious 'I' who is altering between two or more separate selves with separate sets of attitudes, opinions and identities, the therapeutic goal of integrating the different identities in DID can indeed appear like nothing but a murdering of the alters. *"As hard as multiplicity can be to live with, choosing to work towards integration is a painful and frightening process. I would like friends and family members to know that the joining of alters can be like the death of a beloved family member. Although joining is a sign of healing, I still grieve for the loss of a friend who was always there for me."* – Susan B., (Cohen et al., 1991 p. 177).

Probably it is only the individual subjects of experiences or the individual selves that can know when processes like integration or liberation are to strive for. What is important for moral theory is that knowledge derived from unusual experiences can help build answers to questions about the nature of persons and selves, as well as to questions of how we should act towards those who have experienced DID, and perhaps also on how we are to act towards any being having at least one self. These experiences, if shared, may help us all to overcome prejudices about what certain moral objects must look like and show that what appears to be a person can be described in more complex, morally relevant and accurate ways, and that behind or beyond a face there can be more to find.

But like so often happens, where we find answers we will also find many more questions:

> *I am so different from myself sometimes. Like matter and anti-matter. Am I waiting to bump into the other on some unexpected dark corner? What will be left? Will there be a grand flash of light, pure energy in a moment of space and time, like the forming of some distant sun? Or will I cancel myself out? Will I become a vacuum? Blank time? Void? Darkness? Can I then only be expressed in the past tense or the future tense? And when one is gone, this Siamese twin I have connected at the mind, will I mourn the loss? Will I visit this man's grave and cry for myself. Or is gone the wrong concept? Is absorption, or metamorphoses better? Is this my time for wrapping up tightly into a cocoon and loosing matter to whirl and fly changing myselves from us to me? Will I find that inside myself I will then be like two taking up the space of only one?* By Gregory B., (Cohen et al., 1991 p. 174).

5. Summary

Empirical studies summarized by Farah and Heberlein depict a neural person network that is quick to judge. Moral theory and practice concerned with personhood stems from a categorical type of thinking, put forth by this network. The network is categorical and its reliance on certain trigger features, such as faces and bodily movements, has a great impact on our thinking about persons. Farah and Heberlein write that if we had a plant network that functioned similarly to our person network, then we might feel the urge to sniff the flowers on a friend's Hawaiian shirt or to water carpets that are green. Our beliefs can never be neatly separated from our desires and intentions to act, and knowing this should make us hesitant when basing moral theory and practice on innate and rather plump categorization-systems.

Parfit claimed that being a person or not is not an all-or-nothing type of question, but this is precisely what the person network described by Farah and Heberlein most often makes us think. One question is how easy it is to make our brains think differently about the subject. We are not slaves to our innate predispositions. Though one question still, is what happens when we consider moral dilemmas and try to make changes in our beliefs about which individuals that are persons; do we just move our categorical dividing line between persons and things, or do we come closer to a true understanding of personhood as being a matter of degree? Farah and Heberlein's article suggest that we are always influenced by a faulty kind of categorical thinking. In case we do not come closer to a true understanding, it may be important to look for other entities to refer to when discussing moral obligations towards those others that are more to us than mere momentarily conscious beings. I have suggested and argued for that this entity that we should talk about is 'the self'.

As Glover puts emphasis on, we really should talk about and act as if there are morally important entities that exist over time. Using the word and concept self may allow a thinking about moral entities that is free from hindrances formed by our evolutionary formed tendencies as well as from our personhood-related socially construed associations with agency and legacy etc. Discussing self as a morally relevant entity can give input to and gain validation from the discussion of experiences of Dissociative Identity Disorder as well as discussions on other self and consciousness related matters. What may be most important is that where personhood turns out to be a concept hinting at problematic division lines where there are none, selfhood is not tied to irrelevant trigger features and selfhood can easily be understood in terms of degrees.

6. Bibliography

Baars, B., (1997). *In the theatre of consciousness: The workspace of the mind*. New York: Oxford University Press

Bear, R. A., (2003). Mindfulness training as a clinical intervention: a conceptual and empirical review. *Clinical Psychology: Science and Practice, vol. 10, 125-140.*

Blackstone, J. & Josipovic, Z., (1986). *Zen for beginners.* New York: Unwin Paperbacks

Cohen, B.M. et al., (1991). *Multiple personality disorder from the inside out.* Baltimore: The Sidran Press

Cullberg, J., (1999). *Dynamisk psykiatri: i teori och praktik.* Stockholm: Natur och kultur

Farah, M. J. & Heberlein, A. S., (2007). Personhood and neuroscience: naturalizing or nihilating? *The American Journal of Bioethics, 7(1), 37-48, W1-W4.*

Fine, A. H., (2000). *Handbook on animal-assisted therapy, theoretical foundations and guidelines for practice.* London: Academic press

Fox, D., (2004). Do fruit flies dream of electric bananas? *New Scientist, February 32-35.*

Glover, J., (1991). *I: The philosophy and psychology of personal identity.* Harmondsworth: Penguin Books Ltd.

Golombok, S. & Fivush, R., (1994). *Gender development.* Cambridge: Cambridge University Press

Hollis, M., (1994). *The philosophy os social science, an introduction.* New York: Cambridge University Press

Mameli, M. & Bortolotti, L., (2006). Animal minds, animal rights and human mindreading. *Journal of Medical Ethics, 32, 84-89.*

Northoff, G. et. al., (2005). Self-referential processing in our brains – a meta-analysis of imaging studies on the self. *NeuroImage 31, 440–457.*

Parfit, D., (1984). *Reasons and persons.* New York, Oxford University Press

Patterson, C., (2002). *Eternal Treblinka, our treatment of animals and the holocaust.* New York: Lantern Books

Reisberg, D., (2001). *Cognition, exploring the science of the mind, (second edition).* New York: W.W. Norton & Company, Inc.

Robbins, P. & Jack, A. I., (2006). The phenomenal stance. *Philosophical Studies, 127, 59-85.*

Robbins, 2007-06-25, talk at the 11th Annual Meeting of the Association for the Scientific Studies of Consciousness

Singer, P., (1979). *Practical Ethics.* New York: Cambridge University Press

Smith, M., (1994). *The moral problem.* Malden: Blackwell publishers

Sneddon, L.U., (2003). The bold and the shy: individual differences in rainbow trout. *Journal of Fish Biology, 62 (4), 971–975.*

Wynne, C., (2001). *Animal cognition; the mental lives of animals.* New York: Palgrave Macmillian

Zeki, S., (1992). The visual image in mind and brain. *Scientific American. September, 43-50.*

Einstein, A. http://www.wisdomquotes.com/000762.html