Philosophy of Mind and the Neuroscientific Image of Man

Working Towards a Stereoscopic View of Self-Consciousness

By

Jochem Bevers

HPS

Master's thesis

Thesis supervisor: Dr. Niels van Miltenburg

Utrecht University

Student number: 4016939

For Elsbeth

Contents

Introduction

When you look into the depths of your own soul, what is it that you find? Do you find yourself there? Do you find a construct of yourself based upon past experiences? Do you find yourself to be as you want yourself to be? Can you influence how you perceive yourself? Is your self an entity? Can your self be spatio-temporally located within you? Can your self be equated with physical processes? Are you your brains? Is the self something that you can grasp? Is it something which is transcendental? Can your self be equated with what you call "I"? What does it mean to be conscious of your self? These are examples of questions that one could pose about self-consciousness. We are not certain whether the right answers have been provided yet regarding many of the questions that have supposedly been answered. In this essay, I will analyse some of those answers.

Self-consciousness can be described as the awareness that a person has of herself as being a thinking being in the world. A self-consciousness person knows that she is an individual thinking being, in the sense that she knows that other persons are indeed *other*, and that they do not share in her *selfhood*. The description of self-consciousness that I have just given is still a very superficial description, in the sense that not much more is said beside that self-consciousness is a form of individual awareness, or some sort of personal mental abstraction of individuality. How can we gain more knowledge of something that is so personal, but at the same time shared by all people? A philosopher might try doing this by putting what she thinks she knows about it to paper. As she tries to establish what self-consciousness truly is by reflecting on her own self-consciousness, she might develop far-reaching, unheard-of ideas about it. A neuroscientist might try to investigate a very particular location within the human body, which she presumes to be responsible for self-consciousness based upon grounded scientific knowledge. She tries to establish how and why certain biochemical tissues interact with each other and create this abstraction of individuality. The philosopher thinks about thinking from within, by being a first-hand thinker. The neuroscientist not only thinks about thinking from within, by being a first-hand thinker, but also from without, through empirical investigation as a 'third-hand' experiencer.

In this thesis I will analyse how philosophy of mind and neuroscience can contribute to our understanding of self-consciousness, and whether their insights can be combined. I have chosen the

concept of self-consciousness because it has proven to be a thorny issue in both philosophy and the neurosciences, and because there seems to be much that we can still learn about it. Philosophers and neuroscientists use a multitude of approaches and methods to study self-consciousness. For example, whereas the philosopher might make use of introspection or conceptual analysis to establish what the self is, and to determine what it means to be conscious of this self, the neuroscientist might make use of magnetic resonance image (MRI) scans or biochemical theories to establish what the physical preconditions for having self-consciousness are. These are two different ways of attempting to come to a better understanding of self-consciousness, not only with regard to the results, but also methodologically speaking. This is problematic, because, as the methods and results are so different, it is hard to see how they can be combined in one comprehensive understanding of self-consciousness. In this essay, I shall analyse and compare these two ways of investigating self-consciousness, to see how philosophical conceptions of self-consciousness and neuroscientific conceptions differ from each other, and to see whether they can complement each other.

In order to approach this multidisciplinary debate, I will make use of the theoretical framework as expounded by Wilfrid Sellars in his book *Science, Perception, and Reality*.[1] In the first chapter of this essay I will write an account of Sellars' two images of man-in-the-world, which are two frameworks in their own right. Sellars argues for a distinction to be made between *the manifest image of man* and *the scientific image of man*. The former represents the framework of 'sophisticated common sense' in which man first became aware of himself as man-in-the-world. It is the image of man that we encounter in 'perennial philosophy'. The latter represents the framework in which we place scientific knowledge of man-in-the-world. This image could also be called the 'postulational' or 'theoretical' image. Ultimately, these two pictures are to be synthesised into a stereoscopic view, in which neither image should overwhelm the other. This theoretical framework will enable me to place the theories, findings, and ideas by philosophers and neuroscientists within either one of these two frameworks, or in both.

After expounding the theoretical framework in the first chapter, I will discuss several neuroscientific conceptions of self-consciousness in the second chapter by presenting six case studies.

[1] Wilfrid Sellars, *Science, Perception, and Reality* (London: Routledge and Kegan Paul Ltd, 1963).

These case studies will enable to present an argument against the philosophical position called 'eliminative materialism'. The proponents of this position uphold that "our common-sense understanding of psychological states and processes is deeply mistaken and that some or all of our ordinary notions of mental states will have no home, at any level of analysis, in a sophisticated and accurate account of the mind."[2] I will argue that there are certain manifest, common-sense aspects of self-consciousness that do have a home in a sophisticated and accurate account of the mind. In the third chapter I will present two influential philosophical conceptions of self-consciousness by Immanuel Kant and Sebastian Rödl, and I will show that these conceptions have the potential of having a home in a stereoscopic view of self-consciousness. Most philosophical conceptions of self-consciousness can be placed within the manifest image of man-in-the-world, the framework of 'sophisticated common sense', and they are regarded as inferior to scientific conceptions by the eliminative materialist. I shall argue why I believe that a Sellarsian stereoscopic view should be strived for when it comes to the concept of self-consciousness in the fourth chapter, a view in which neither image holds primacy, to counter the eliminative materialist's claim.

With this study I hope that I can add to the debate on self-consciousness in three ways. Firstly, I wish to ascertain whether philosophical and neuroscientific conceptions of self-consciousness can be placed within the manifest and/or the scientific images, and whether primacy is granted to either of these images by their respective researchers, in order to better place the concept of self-consciousness within the human web of knowledge. Secondly, by arguing against eliminative materialism, I will try to show that philosophers with manifest, common sense ideas still have a fundamental role to play when it comes to investigating self-consciousness. And thirdly, by reflecting on the material presented in the first three chapters, I wish to offer some suggestions which might lead towards a stronger unification of philosophical and neuroscientific knowledge on self-consciousness, and towards a conceptualisation of the self and its qualities, both rational and physical, in a stereoscopic view. This essay is primarily concerned with the philosophy of science, and with the relation between science and philosophy. Its

[2] William Ramsey, "Eliminative Materialism," *Stanford Encyclopedia of Philosophy*, accessed May 31, 2017, https://plato.stanford.edu/entries/materialism-eliminative/. The most famous example of such a notion is a 'sensation', which is an important concept in philosophy of mind. It is a notion that we will return to several times in this essay.

thesis, presented below in *italics*, is twofold: the philosophical position of eliminative materialism is rejected, and suggestions are proposed for a view of self-consciousness in which the philosophical foundation for neuroscientific research on self-consciousness is inspired by said research and vice versa.

Certain manifest aspects of self-consciousness are necessary conditions for understanding the possibility of human cognition in general, and these ideas can function as philosophical principles upon which a neuroscientific and stereoscopic image of self-consciousness can be built.

Chapter One
Wilfrid Sellars and the Two Images of Man

Before we can establish how philosophers and neuroscientists investigate and conceptualise self-consciousness, and before we can establish how both disciplines might complement each other, a theoretical framework is required, in which the different conceptions, and several possibilities for synthesis can be placed. In order to discuss the neuroscientific and philosophical images of self-consciousness, I will make use of my own adaptation of Wilfrid Sellars' theory on the relation between philosophy and science as presented in the chapter "Philosophy and the Scientific Image of Man" in his book *Science, Perception, and Reality*.[3] This text will prove useful because of several reasons. Firstly, Sellars discusses the relation between philosophy and science in a similar manner as I hope to be able to do in this essay; he not only discusses how they differ, but also how they could be brought together. Secondly, Sellars explicitly discusses neurophysiology and its place in the intellectual landscape. And thirdly, Sellars discusses *self-awareness*, which can be regarded as an important aspect of self-consciousness, in the contexts of introspection and neurophysiology. By assessing the different neuroscientific and philosophical ideas on self-consciousness within this particular framework, a different light can be shed on these conceptions as they can be compared more easily, a light which has the potential of strengthening the bonds between neuroscience and philosophy, bridging the disciplinary gap on the topic of self-consciousness. The chapter of Sellars' book, "Philosophy and the Scientific Image of Man", which I shall now discuss, originally consisted of two lectures given at the University of Pittsburgh in December 1960. This chapter will follow the structure that Sellars uses in his own chapter. I will make use of his subdivision of sections and of the titles, to which I have added a few words behind each hyphen '–'.

[3] Sellars, *Science, Perception, and Reality*, 1-40.

I. THE PHILOSOPHICAL QUEST – TO KNOW ALL

"THE aim of philosophy, abstractly formulated, is to understand how things in the broadest possible sense of the term hang together in the broadest possible sense of the term."[4] Sellars describes this aim as "the reflective knowing one's way around in the scheme of things."[5] This holistic view on the aim of philosophy is worth pursuing, but it is also something which is hard to achieve. Understanding how our minds work, from Sellars perspective, is only one 'thing' of all the things hanging together, but it is fundamental for understanding how things hang together. For, it is us, human beings, that do the understanding, and understanding how we are able to understand even a fraction of the world around us, and the world within us, is a fundamental basis for the grounding of all knowledge. Understanding how things hang together metaphysically, goes hand in hand with understanding one's own place as a perceiving and acting subject within a web of knowledge. Surely almost every person has a conception of him- or herself as man/woman-in-the-world, a manifest image.[6]

Sellars argues that what is characteristic of philosophy is the aim of knowing one's way around with respect to the subject-matters of all the special disciplines; having 'the eye on the whole' of the specialisations within the sciences. Whereas the philosopher is supposed to know a little of all the disciplines, the specialist has in-depth knowledge about her specific discipline, but she must also have some understanding of how her discipline fits within the intellectual landscape.[7] According to Sellars, there is little unity to be found within the picture that one gets when having one's 'eye on the whole'. In principle, there are two pictures the philosopher is confronted by, but in fact, Sellars argues, there are many. Sellars does not go on to explain why there are in fact many pictures, but instead, he focusses his attention on two pictures which purport to be a complete picture of man-in-the-world. Perhaps because he wishes to make a point which would become too confused if more than two general idealisations are used. These two pictures, conceptions, or perspectives, are essentially of the same order of complexity, are equally non-arbitrary, equally public, and they must be fused into one vision, or stereoscopic view, by the philosopher, which is something I hope to set steps towards in this essay regarding the concept

[4] Sellars, *Science, Perception, and Reality*, 1.
[5] Ibid., 2.
[6] In the rest of this essay I will make use of the term 'man-in-the-world' in a similar manner as Sellars.
[7] Sellars, *Science, Perception, and Reality*, 2-3.

of self-consciousness. These pictures are the *manifest* and the *scientific* images of man-in-the-world. Sellars makes use of the ambiguous term 'image' because it suggests its dependent status of an object as a sort of projection or conception of that image, which makes it 'usefully ambiguous' as a contrast between an existing object and a projection of this object which also exists, be it dependently as objects of philosophical reflection and evaluation, and because it is something which is imagined that *can* exist.[8] Both images are idealisations with their own histories.

> They are designed to illuminate the inner dynamics of the development of philosophical ideas, as scientific idealizations illuminate the development of physical systems. […] [W]hile the main outlines of what I shall call the manifest image took shape in the mists of pre-history, the scientific image, promissory notes apart, has taken shape before our very eyes.[9]

Although both images are equal in the ways described above, the manifest image is many millennia old, whereas the coming to be of the scientific image seems to be mostly a twentieth and nineteenth century development. It is interesting that earlier in the chapter Sellars writes that the special subject-matters that philosophers did have, have been turned over to non-philosophers over the past 2500 years, starting with mathematics, and more recently psychology and sociology.[10] This indicates that although more and more sciences became separate disciplines over the centuries, they only started providing a scientific image of man-in-the-world quite recently, which illustrates the philosophical and historical origins of the modern sciences. In the following three sections both images will be described and explained in greater detail.

II. THE MANIFEST IMAGE – LEAPING AWAY FROM THE ORIGINAL IMAGE

Sellars characterises the manifest image in two different, but supplementary ways. The first is 'quasi-historical', in the sense that it is the framework in which man first became aware of himself as man-in-the-world, or as Sellars also describes it, it is man's first encounter with himself; he became aware of himself as an individual being which exists in the world. In a sense, this is the 'birth' of self-

[8] Sellars, *Science, Perception, and Reality*, 4-5.
[9] Ibid., 5.
[10] Ibid., 2.

consciousness. Sellars upholds that a jump has been made here; this was a transition from pre-conceptual patterns of behaviour, to conceptual thinking, a transition which was the coming into being of man. It is likely that we will never know how this transition occurred, since it took place in pre-historical times. Sellars believes it to have been so sudden, because a conceptual framework functions as a whole that is *prior* to its parts. When only one, or a few parts are conceptual in character, whilst the framework is not, then these parts will not function in the framework as having conceptual characteristics, for they are not grasped within the framework, which proves the impossibility of having merely conceptual parts. The paradox is that man could not be man before he encountered himself and developed conceptual thinking.[11] It does seem strange, however, that this transition should have occurred in an entire species in the form of a 'jump'. Would this have occurred in one individual who enlightened the rest? Is conceptual thinking something which we have acquired rationally or something which was activated through experience? These are hard questions to answer. The non-historical connection between (conceptual) thinking and man being aware of himself as man-in-the-world, or as self-in-the-world, will be scrutinized extensively in the third chapter, as Kant's and Rödl's conceptions of self-consciousness both place a great emphasis on this connection.

The second way in which Sellars characterises the manifest image is by describing it as an empirical and categorical refinement or sophistication of the 'original', pre-conceptual image. The manifest image is categorical, because of the alteration of the classification of the framework. It is empirical, because the manifest image is itself a scientific image, in a limited sense, that makes use of the scientific method of correlational induction. Sellars stresses that the manifest image is not a historical and bygone stage. He demarcates the manifest from the scientific image by arguing that there is "one type of scientific reasoning which it, by stipulation, does *not* include, namely that which involves the postulation of imperceptible entities, and principles pertaining to them, to explain the behaviour of perceptible things."[12] That which is imperceptible is of course not manifest, hence the name 'manifest image' for the image that does not include the postulation of these imperceptible entities. Therefore, the scientific image of man-in-the-world could also be called the 'postulational' or 'theoretical' image.

[11] Sellars, *Science, Perception, and Reality*, 6.
[12] Ibid., 7.

Sellars argues that correlational and postulational methods, being dialectically related, have gone hand in hand in the evolution of science, whilst the scientific image proper has only taken shape before our very eyes relatively recently.[13]

The manifest image defines that to which philosophical reflection has been drawn over the centuries, according to Sellars, ranging from the speculative systems of ancient and medieval philosophy, to systems and 'quasi-systems' in recent and contemporary thought, both Continental and Anglo-Saxon. He groups these philosophies under the term 'perennial philosophy of man-in-the-world', and they include the Platonic tradition in its broadest sense, and philosophies of 'common sense' and ordinary usage. In it, the manifest image is promoted as real, as an adequate account for understanding the reality of man and the world, and science is merely used as a *tool* to better understand specific aspects of this reality; the manifest image is dominant, which Sellars believes to be a distorted view of things. Sellars seems to uphold that the perennial philosophers refine and endorse a misconceived image, which does not accommodate the scientific image properly.[14] In the fourth chapter I will compare the neuroscientific and philosophical images of man to establish whether this is valid with respect to the concept of self-consciousness. Sellars argues that both images are to be blended in a stereoscopic view. In this synthesis, neither image should overwhelm the other.[15]

Before the scientific image and the possibilities of a synthesis can be characterised, we must first delve a little deeper into the manifest image as a conceptual framework as described by Sellars in this section and the next. With regard to the general structure of the manifest image, Sellars argues that determining the basic objects of a (given) framework is to provide a more or less abstract classification. For the perennial philosopher the classification provides a general view of the contents of the framework, which does not directly refer to them as objects or entities. Sellars refers to the classification as a list of objects echoing the lower stages of the 'great chain of being' of the Platonic tradition, such as persons,

[13] Sellars, *Science, Perception, and Reality*, 5 and 7.
[14] Ibid., 8, and 18-19. I use the word 'seems' here, because in the text Sellars is quite ambiguous: "For I have also implied that […] I seem, therefore, to be saying […] If this were what I wished to say".
[15] Ibid., 9.

animals, plants, and material things like minerals.[16] The image below is a sixteenth century Neoplatonist pictorial representation of the chain, which springs from God and descends to the seven archangels, angels, people, animals, plants, and minerals respectively. Sellars argues that there is an important sense in which the primary objects of the manifest image are persons. To understand how persons are the primary objects, is to understand fundamental themes in the history of philosophy, for the first layer beneath the transcendent, that of which we truly know that it exists, is the human being.[17]

Pictorial representation of the great chain of being in Didacus Valades' *Rhetorica Christiana* (1579).[18]

[16] Sellars, *Science, Perception, and Reality*, 9. Interestingly, in the same paragraph Sellars writes that to ask for objects, is to *not* ask for a list, but for a classification. It must be said, however, that the list is solely intended for approaching an answer to the question.
[17] Ibid., 9-10.
[18] Image found in: Anthony Fletcher, *Gender, Sex, & Subordination in England 1500-1800* (New Haven: Yale U.P., 1999), 138-139, plate 15. The original image can be found in the collection of the British Library.

Whereas in the original image of man-in-the-world all objects were persons, "*originally* to be a tree was *a way of being a person*", after the shift to the manifest image, trees were no longer thought of as persons.[19] This change, argues Sellars, is not a gradual change in belief, but a radical change in category. Again we face the problem of how this transition could have occurred in an entire species in the form of a sudden 'jump', especially in regard to the prehistorical aspect, but we will let this rest as it is not a fundamental problem with regard to the thesis. Sellars remarks that in general a *person* is not a team of a mind and a body, but simply a single person. Even though the dualism in philosophical theories where mind and body are two different substances is prominent, the *essential dualism* in the manifest image, which Sellars returns to later, is between two kind of ways in which an individual human is related to the world.[20] We will return to the dualism of body and mind, or to its seeming absence in contemporary thought in the fourth chapter.

III. CLASSICAL PHILOSOPHY AND THE MANIFEST IMAGE – THE PERENNIAL TRADITION SCRUTINISED

According to Sellars, the manifest image has an objective existence in philosophical thinking 'in its own way', and in human thought generally. It is one of the poles to which philosophical thinking is drawn, and it transcends 'in some way' the individual thought of individual thinkers. There is truth and error in respect to the image, because it has 'a being' which transcends the individual thinker. Academic philosophers try to precisely describe the manifest image as an existent image which transcends their thinking. Sellars suggests that the manifest image might ultimately have to be rejected as false 'in the last analysis'. The analytic tradition, a continuity of the perennial tradition inspired by the later Wittgenstein, has tried to come to a pure form of the manifest image, and by doing so, it has shown that it is foolish to try to *replace* it step by step with the scientific image.[21] The logical positivists have presumably contributed to this the most. It is important to delineate the image according to Sellars, "man is what he is because he thinks of himself in terms of this image, and the latter must be understood before it is proper to ask, 'to what extent does manifest man survive in the synoptic view which does equal

[19] Sellars, *Science, Perception, and Reality*, 10. Emphasis in original (as in all further citations in this thesis).
[20] Ibid., 10-11.
[21] Ibid., 14-15.

justice to the scientific image which now confronts us?'"[22] This is one of the most important questions that I wish to answer in this thesis, only then slightly altered: 'To what extent does manifest man survive in the synoptic view of the neuroscientific image of man?'

"Now one of the most interesting features of the perennial philosophy is its attempt to understand the status in the individual thinker of the framework of ideas in terms of which he grasps himself as a person in the world."[23] Does this result in a regress? When one tries to understand the status of this framework in the individual, which is the framework in which the individual understands himself as man-in-the-world, does one then not think of himself as man-in-the-world, resulting in a never-ending quest for defining the framework that allows oneself to understand oneself in the world? Sellars states that there are two things to be noticed when we try to understand how individuals come to be able to think in terms of the framework, and how they come to have the manifest image: (1) The manifest image presents conceptual thinking as a complex of items which, considered in themselves and apart from these relations, are conceptual in character. (2) The process of conceptual thinking itself as it occurs in the individual mind must be in accordance with, or at least echo, the intelligible structure of the world.[24] Now, the first point seems clear; the items of conceptual thinking must surely be conceptual when considered philosophically. As we shall see, the neuroscientist might argue otherwise. The second point raises a lot of interesting questions which cannot be properly answered due to the scope of this essay: What is the intelligible structure of the world? What is the process of conceptual thinking? Can it be defined? Doesn't our imagination enable us to think conceptually in a way that has nothing to do with the intelligible structure of the world or with any structure whatsoever? Must conceptual thinking always abide to a set of rules? Although they remain unanswered, these questions will be touched upon in the next two chapters.

Sellars makes two further generalisations of the perennial tradition: "*(a)* that association of

[22] Sellars, *Science, Perception, and Reality*, 15.
[23] Ibid., 15.
[24] Ibid., 15. I have rephrased these two sentences in order to remove some ambiguity from the original statement; the first sentence contained a double negative, and in the second sentence Sellars uses the words "must echo, more or less adequately", when referring to the process. The original sentences are: "(1) The manifest image does not present conceptual thinking as a complex of items which, considered in themselves and apart from these relations, are not conceptual in character." "(2) Whatever the ultimate constituents of conceptual thinking, the process itself as it occurs in the individual mind must echo, more or less adequately, the intelligible structure of the world."

thoughts is not association of images, and, as presupposing a framework of conceptual thinking, cannot account for it; *(b)* that the direct action of perceptible nature, *as perceptible*, on the *individual* can account for associative connection, *but not the rational connections of conceptual thinking*."[25] Point (a) seems logical, for thoughts contain more than images, and so it follows that there are more possibilities of association. This is not to say that images are not important. Point (b) requires more extensive consideration. Let us suppose that the photons which strike the retinas in a person's eyes are a direct action of perceptible nature on an individual. She does not perceive the photons individually, but she does perceive light, which is a key part of visually perceptible nature. This can indeed account for associative connection, for it can function as an input of experience, which might lead to thoughts, and then in turn to associative connections of conceptual thinking. But why might they not account for the rational connections? It seems strange to assume that the entire perennial tradition agrees with this idea, for there have most certainly been philosophers who believed that man is born without any rational connections or ideas, which are gained solely through experience. Sellars does stress the words '*as perceptible*', so it might be argued that the transition from perception to rational connection is something different than implied in point (b).

Starting with Plato, and existing to the present day, is the philosophical idea of a direct causal influence of the world as intelligible on the individual mind, be it through illumination from the One, the Good, God, or simply as a result of causal relations present in the world itself, as in strict determinism. It was not until the nineteenth century, argues Sellars, that the fundamental role of the group as a mediating factor in the causation of the world on the individual mind was recognised, most prominently in Hegel's writings. "Yet the essentially social character of conceptual thinking comes clearly to mind when we recognize that there is no thinking apart from common standards of correctness and relevance, which relate what *I do* think to what *anyone ought to* think. The contrast between "*I*" and *anyone* is essential to rational thought."[26] Each individual thinks of himself as an "I" in regard to the other members of the group, or simply put the 'other'. The group in turn exists in the way the individuals represent themselves; the individual and the group are thus strongly connected, most importantly in the

[25] Sellars, *Science, Perception, and Reality*, 16.
[26] Ibid., 16-17.

sense that the group mediates between the individual and the intelligible order of the world. Although the manifest image is to be construed as containing a conception of itself as a group phenomenon, any attempt to explain the mediation within its framework is unachievable. Sellars argues that it can only provide the foundation on which a scientific theory can be built which explains the meditation, and that it cannot be defined within the framework itself.[27] I think that Sellars upholds that whereas in the scientific image of man we can explain the rise of the meditation of the group in evolutionary terms, as advantageous for the perseverance of human species, from the manifest image we cannot know *why* it has come to be that the group mediates. Sellars uses this argument to point out that within the manifest image there is an irreducibility of all forms of conceptual thinking to more elementary processes. This is an important point, which shall be discussed in more detail later in this chapter. The irreducibility is related to the primary and essential dualism of the perennial philosophy, i.e. the two types of causal impact of the world on the individual I have just discussed; of direct causal influence of the world as intelligible on the individual mind, and of indirect causal influence of the world as intelligible on the individual mind mediated by the group, as a group phenomenon.[28]

Sellars finishes this section with his primary concern: "'in what sense, and to what extent, does the manifest image of man-in-the-world survive the attempt to unite this image in one field of intellectual vision with man as conceived in terms of the postulated objects of scientific theory?'" Because man conceives himself as a being in terms of the manifest image, not much might remain of man as he knows himself once scientists are done with it.[29] We shall see how much remains of the self in the theories of neuroscientists.

IV. THE SCIENTIFIC IMAGE – THE UNIFICATION OF ENTITIES AND PRINCIPLES

> I suggested that the most fruitful way of approaching the problem of integrating theoretical science with the framework of sophisticated common sense into one comprehensive synoptic vision is to view it not as a piecemeal task—e.g. first a fitting together of the common sense conception of physical objects with that of theoretical physics, and then, as a separate venture, a fitting together of

[27] Sellars, *Science, Perception, and Reality*, 17.
[28] Ibid., 17-18.
[29] Ibid., 18.

> the common sense conception of man with that of theoretical psychology—but rather as a matter of articulating two whole ways of seeing the sum of things, two images of man-in-the-world and attempting to bring them together in a 'stereoscopic' view.[30]

Before we can look at the problems and possibilities of this promising 'stereoscopic' view, we must first delve deeper into Sellars' vision of the scientific image, the image which is supposed to be given primacy in the combined image of man-in-the-world. According to Sellars, the scientific image is still under development, which makes it a greater idealisation than the manifest image. Whereas the manifest image is the ideal of a correlational and categorical refinement of the original image, the scientific image is the ideal in which the image is procured from the results of postulational theory construction. The images of man constructed with these theories are as diverse as the many sciences which have something to say about man, like the chemist, the physicist, and the psychologist. It is therefore that *the* scientific image of man is an idealisation "in the sense that it is a conception of an integration of a manifold of images, each of which is the application to man of a framework of concepts which have a certain autonomy."[31] Thus, the scientific image is a construct of several images provided by diverse specialisations. These specialisations are autonomous sciences in which certain concepts are applied to man. For example, chemical concepts, biological concepts, or physical concepts. What is important, is that Sellars believes that each of these separate images, created by the sciences, is supported by the manifest world. So, the world that is apparent to the scientist, supports the scientific image, which is created with postulational theory. A question that arises here is how the manifest world could support anything that is postulated, or imagined. "[E]ach theoretical image is a construction on a foundation provided by the manifest image".[32] Thus, the scientific image is supported by the manifest world, whereas the separate theoretical images, which form a manifold resulting in the scientific image, are supported by the manifest image. Another key aspect of the scientific image is that although different sciences use different methods and instruments, the theoretical entities they produce have an intrinsic identity which is unifiable. Sellars gives the example of biochemical compounds being 'identical' with patterns of subatomic particles, in

[30] Sellars, *Science, Perception, and Reality*, 19.
[31] Ibid., 20.
[32] Ibid., 20.

the sense that they are two theoretical structures, replaceable by one theoretical framework, *a* scientific image. This framework should be regarded as being connected at 'two levels of complexity' with the perceptible, manifest world trough different instruments and procedures.[33]

Although, the scientific image presupposes the manifest image in a methodological sense, as a foundation, it is not prior to it in any other way, for the scientific image purports to be a *complete* image. When the manifold of images from a number of sciences is combined, from it emerges the scientific image as a framework which has the potential, argues Sellars, to be the *whole truth* about that which belongs to the image, e.g. the working of the brain as constructed with chemical, physical, and neurological theory. The scientific image is a *rival* image, even though it is a development within the manifest image methodologically. "From its point of view the manifest image on which it rests is an 'inadequate' but pragmatically useful likeness of a reality which first finds its adequate (in principle) likeness in the scientific image."[34] It is hard to understand why the scientific image should have such superiority in regard to the manifest image, which is ought to find the scientific image to be superior to itself due its 'adequate likeness' of reality, especially since the scientific image is still under development and since it is so reliant on the manifest image. Although Sellars tries to view both images separately throughout the chapter, be it as idealisations, it does seem that both images are more strongly intertwined than he is suggesting: if the manifest image supports the other as a methodological foundation, and if the scientific image presents a more truthful account of reality to the manifest image, then it seems as if both images must be in stereoscopic use all the time without much primacy credited to either image. Furthermore, Sellars states himself that "in point of historical fact many of the latter correlations were suggested by theories introduced to explain previously established correlations, so that there has been a dialectical interplay between correlational and postulational procedures."[35] The merger of the manifest image and the scientific image has been under development for quite some time.

In the remainder of this section, Sellars primarily discusses the place of behaviouristics in the scientific image, which is troublesome due to the subjective nature of behaviour. Most of this is not

[33] Sellars, *Science, Perception, and Reality*, 21.
[34] Ibid., 20.
[35] Ibid., 19.

relevant for this essay, and therefore I will not discuss it in detail here. What I will discuss, however, are Sellars' remarks on behaviouristics link with neurophysiology, which is relevant as it is a branch of the neurosciences. Neurophysiology is first mentioned in relation to the unification of theoretical principles of sciences. Whereas the laws of biochemical substances are 'special cases' of the laws of particle physics, i.e. the former forming more complex patterns than single particles, the theoretical principles of the neurophysiological image of man are 'special cases' of the principles pertaining to the biochemical image of man, i.e. nerves and brain cells forming more complex patterns than biochemical substances. In both these comparisons the entities can be equated, whereas the principles should be considered as special cases of the more 'basic' or 'finer grained' science.[36] Sellars argues that in a scientific account of behaviouristics, the correlations that are found pertaining to human behaviour must find their counterparts in the postulational image, which are neurophysiological and, consequently, biochemical connections. "I shall, therefore, provisionally assume that although behaviouristics and neurophysiology remain distinctive sciences, the correlational content of behaviouristics points to a structure of postulated processes and principles which telescope together with those of neurophysiological theory, with all the consequences which this entails. On this assumption, if we trace out these consequences, the scientific image of man turns out to be that of a complex physical system."[37] I believe that Sellars here applies a unification of theoretical principles of sciences when he speaks of 'tracing out the consequences'. The postulated processes and principles telescoping with those of neurophysiological theory are special cases of processes and principles in biochemistry, which in turn are special cases of principles in particle physics, and it is therefore that the scientific image of man turns out to be that of a complex neurophysiological, biochemical, and ultimately physical system. A final issue I wish to point out is that we again stumble upon an intertwinement between aspects of the manifest and the scientific images, for it is the *correlational* content of behaviouristics pointing to postulational content of other sciences. When the former content 'points to', it seems that, conversely, the latter content makes use of the former, in the sense that it is supported by the former epistemologically. In the next chapter, the intertwinement between key aspects of the manifest and

[36] Sellars, *Science, Perception, and Reality*, 21-22.
[37] Ibid., 25.

scientific images will be laid to bare when it comes to neuroscientific conceptions of self-consciousness, and in the fourth chapter I will suggest how the manifest and scientific images of self-in-the-world can do more than merely 'point to' each other, but strengthen each other, despite the seeming lack of intertwinement emanating in the third chapter in which two philosophical conceptions of self-consciousness are discussed.

V. THE CLASH OF THE IMAGES – THE FUSION OF WORLDS

In this section Sellars investigates how we are to evaluate the conflicting claims of both images of being 'the true and complete account of man-in-the-world'. The clash is studied from a historical perspective, starting with Descartes, as he is an example of a perennial philosopher who lived during an early stage of postulational science, and who was clearly influenced by this 'new' way of looking at the world, which resulted in an attempt at a synthesis. It seems that Sellars has quite some esteem for at least one branch of the neurosciences, for he states that it is obvious that at the time of Descartes, theoretical science had not yet reached the neurophysiological level. It was *inanimate* nature which was challenged by the upcoming scientific image. Physical things were to be construed as systems of imperceptible particles, and thus not part of perceptible, manifest nature. Sellars argues that this was done in a manner foreshadowed by Greek atomism.[38] Perhaps this is not just an adumbration of a scientific theory, but, in a sense, a postulational theory in its own right, as imperceptible entities are key elements in this worldview.

Sellars presents three lines of thought which seemed to be open in regard to this 'new' postulational vision: "(1) Manifest objects are identical with systems of imperceptible particles in that simple sense in which a forest is identical with a number of trees. (2) Manifest objects are what really exist; systems of imperceptible particles being 'abstract' or 'symbolic' ways of representing them. (3) Manifest objects are 'appearances' to human minds of a reality which is constituted by systems of imperceptible particles."[39] Sellars is primarily interested in thought (3), followed by thought (1). Although thought (2) has its merits, it is not considered by Sellars. The trees in thought (1) are

[38] Sellars, *Science, Perception, and Reality*, 25-26.
[39] Ibid., 26.

imperceptible entities which are part of a perceptible object, the forest. A bit of a strange metaphor considering that trees are highly perceptible entities, but the message is clear. In this way of thinking systems can have properties which their parts do not have; a heart is a pump, but a heart cell on its own is not considered to be a pump. In this line of reasoning, the same holds for the property of perceptibility, but *not* when it comes to a quality like colour. A red rock presents itself to us as *ultimately homogenous*, as an object which is a red continuum, and which remains red to us no matter how many times you would divide it into smaller pieces. Reflecting on this, Sellars formulates the following principle: "If an object is *in a strict sense* a system of objects, then every property of the object must consist in the fact that its constituents have such and such qualities and stand in such and such relations or, roughly, every property of a system of objects consists of properties of, and relations between, its constituents."[40] So the parts form the whole, and it's due to the parts, that objects have certain qualities. Now, states Sellars, it was argued that if these parts are imperceptible particles, it is not possible that the whole has the perceptible qualities which are characteristic of physical objects in the manifest image; (3) the imperceptible particles *appear* to us as perceptible objects, but it is the appearance of a system of particles, and we do not perceive the system as it truly is.[41] Physical objects have perceptible qualities, but they cannot be beheld as they truly are, as *Dinge an sich*.

Sellars takes this thought further. In response to G.E. Moore's objection that objects, like tables, cannot be appearances, because we *know* that they are real objects, Sellars replies that it is wrong to uphold that the claim that physical objects do not really have perceptible qualities, is analogous to the claim that something is generally believed to be true about a certain kind of thing is actually false, for the perception of an appearance is not false, but it is our capacities to perceive the world that limit us. "It is not the denial *of* a belief *within a framework*, but a challenge to the framework. It is the claim that although the framework of perceptible objects, the manifest framework of everyday life, is adequate for the everyday purposes of life, it is ultimately inadequate and should not be accepted as an account of what there is *all things considered.*"[42] This philosophical position, which emphasises the primacy of the

[40] Sellars, *Science, Perception, and Reality*, 27.
[41] Ibid., 27-28.
[42] Ibid., 27.

scientific image, poses a challenge to the 'common sense' framework. An account of what there is *all things considered*, will always be an incomplete account, be it manifest or scientific, although the latter might indeed come closer to the 'truth'. Sellars certainly seems to uphold that the success of living, thinking, and acting in terms of the manifest framework can be accounted for more properly by the scientific image.[43] I am still not truly convinced, however, that these two idealisations can be regarded separately, let alone as opposing each other. Beside the images, the frameworks also seem to be intertwined. A scientific framework without human beings is void, and scientific knowledge encompasses a lot more than mere postulational techniques. Furthermore, scientific knowledge seeps through in peoples' manifest worldview, even if they are not aware of it. And does the scientist in his everyday life make us of the manifest framework, whilst when he is at work he will only make use of the scientific framework?

Sellars convincingly rebukes defences of the reality of the manifest image, by showing how they do not prove what they set out to prove because they operate only *within* the manifest framework; they lack an external point of view with which to defend it. The scientific image on the other hand does provide an external point of view in regard to the manifest framework.[44] Especially when we have thought (3) in mind, in which manifest objects are 'appearances' to the human mind in the framework of space and time, combined with the 'objective' and postulational nature of science, we can better see how our understanding of the world arounds us is clouded by our perceptive and rational limitations. As will become clear later in this essay, this argument gives good grounds for the establishment of a stereoscopic view of self-in-the-world, as the manifest philosophical conceptions are found to be wanting with respect to this argument.[45] A question that arises is: can the scientific image provide an external point of view with regard to the scientific framework? This certainly seems possible, but not without always taking into account the manifest image, for it must still be humans who judge a framework, which seems impossible to do in a solely scientific and purely objective manner. I believe that this is why Sellars is striving for a stereoscopic view of the world, in which the scientific image is

[43] Sellars, *Science, Perception, and Reality*, 28.
[44] Ibid., 27-28.
[45] We will return to this point in the fourth chapter in the sections "II. Bridging the Sellarsian Gap" and "IV. Transcendental Idealism and the Neuroscientific Object: Locating and Delimiting Self-Consciousness".

the more intelligible account of what there is, whilst it necessarily coexists with the manifest image that is not overwhelmed by it. This point alone provides us with sufficient reason to doubt the eliminative materialist position, which gives primacy to the scientific image, overwhelming the manifest image. In order to better place the eliminative materialist's philosophical position, I will now present four materialist views.

In their essay 'What is eliminative materialism?', William Lycan and George Pappas provide the following four views: (1) The materialist view simpliciter is: "If there are sensations, then they are identical with certain physical items (or with physical states of physical items, etc.; let us call all such things 'brain-processes', for short)."[46] (2) The reductive materialist view is: "There are sensations, but they are not distinct from brain-processes."[47] (3) The weak eliminative materialist view is: "'What we called "sensations" turn out to be nothing but brain-processes'."[48] (4) The strong eliminative materialist view is: "It is just (non-trivially) false that there are any sensations. The term 'sensation', for all its apparent referring function, fails to denote anything, just as 'unicorn' does."[49] Lycan and Pappas convincingly show that the weak eliminative materialist view (3) is unsatisfactory as it ultimately collapses in either the view of the reductive materialist (2) or of the strong eliminative materialist (4). In this essay, I will argue against eliminative materialism, because the term 'sensation' does denote something, it denotes something which is essential for understanding human cognition. (a) If the eliminative materialist is correct then there would be no sensations, and (b) sensation is essential for self-consciousness. In the next section I will present Sellars' discussion of the concept of sensation, and in the conclusion I will show why sensation is essential for having self-consciousness, and why the eliminative materialist is therefore incorrect, based on arguments provided in the third chapter of this essay. But first, let us turn to Sellars' analysis of 'thought' in the two images of man.

Descartes, and other philosophers in the era of the new physics, ascribed those things which do not partake in mechanical explanation to the mind of the perceiver, argues Sellars. Traits of objects, like

[46] William G. Lycan and George S. Pappas, "What is eliminative materialism?," *Australasian Journal of Philosophy* 50:2 (1972): 149.
[47] Ibid., 149.
[48] Ibid., 150. This view is a citation excerpted by Lycan and Pappas from Richard Rorty's essay "Mind-Body Identity, Privacy, and Categories," *Review of Metaphysics*, Vol. XIX (1956-1966): 28.
[49] Lycan and Pappas, "What is eliminative materialism?," 150.

colour, were regarded as conceptual constructions, which held no ontological status in physical reality, only in thought. It was considerations like this which led these philosophers to a dualistic theory of man, in which the body is a system of particles, which cannot be the subject of thinking and feeling, making the mind an entirely different substance, which interacts with the body.[50] Sellars believes that man is a complex physical system, whose thoughts can be construed as complex interactions of physical particles in a postulational image, which, as we shall see, is in line with the thoughts of many neuroscientists. If Descartes also had believed that there are neurophysiological processes which are analogues to conceptual thinking, which many philosophers and neuroscientists especially nowadays believe, then he must have had to change his position in regard to the reality of physical objects, or he must have assumed that conceptual thinking really is a neurophysiological process.[51] There would be no proper foundation for conceptual thinking to rest on, except for the *cogito* argument, which only convinces the individual of the certainty that she is a *thinking being*. This tells us nothing about *how* our thinking works, only *that* it works. The *cogito* is a clear example of an image which regards man-in-the-world from within, which is only manifested in the grounding of the being of the self through conceptual thinking, versus an image which regards man-in-the-world from without, which is a scientific approach for understanding how conceptual thinking works in a physical way. This is an important distinction that we will return to several times in this essay.

"Just as the claim that 'physical objects are complexes of imperceptible particles' left us with the problem of accounting for the status of the perceptible qualities of manifest objects, so the claim that 'thoughts, etc., are complex neurophysiological processes' leaves us with the problems of accounting for the status of the *introspectable qualities* of thoughts."[52] The ontological status of introspectable qualities in the scientific image is somewhat problematic; these qualities exist in introspective awareness *of* the thoughts which seem to have them, argues Sellars, but not in the thoughts themselves, resulting in a regress. Sellars demarcation of awareness and thoughts is troublesome, for it seems hard to uphold that being aware of something is not also a form of thought. Thinking, or being aware of a thought,

[50] According to Descartes, this interaction takes place in the pineal gland.
[51] Sellars, *Science, Perception, and Reality*, 29-31.
[52] Ibid., 31.

which seems to have an introspectable quality, seems to be such a complex neurophysiological process, that I wonder if a purely scientific image can ever be developed to such an extent that it will be truly understood as a physical process. These problems make it understandable, according to Sellars, that people are enticed to say that the manifest image is real, and that the postulated entities are 'symbolic tools' and that the processes in the brain run parallel to conceptual thinking, but that they are not the same. But, believes Sellars, the manifest image should not be the measure of what there really is.[53] We will discuss the awareness of one's own thoughts extensively in the third chapter, especially in light of Sebastian Rödl's conception of self-consciousness.

VI. THE PRIMACY OF THE SCIENTIFIC IMAGE: A PROLEGOMENON – AN INCOMPLETE IMAGE

In order to discuss the supposed primacy of the scientific image, Sellars tries to tackle the problem of introspectable qualities which seems to stand in the way of identifying thoughts with neurophysiological processes. Here he phrases the problem differently; supposing that in self-awareness conceptual thinking presents itself to us in a qualitative guise, is a mistake. Thus, the introspectable qualities of thoughts seem to be the same thing as self-awareness of conceptual thinking in Sellars' view. Sensations and images present themselves to us in a qualitative character, but they cannot be equated with conceptual thinking, or with the complexes consisting of them.[54] It is unclear how introspectable qualities and qualities of sensations and images are related to each other, as they seem to be two quite different things. It might be the case that Sellars tries to show that conceptual thought is not qualitative in nature, and that self-awareness of conceptual thinking is something which should not be regarded as qualitative in nature. This does seem so when he compares the direct knowledge of our thoughts to introspection to the perceptual knowledge from observation. They are only analogous in the fact that they are basic forms of non-inferential knowledge, but they differ in the fact that the latter is qualitative, whereas the former pertains to something which is going on *in me*.[55] In chapter three I will discuss whether the object or subject that is the self, is something which can or cannot be considered as having qualities like other

[53] Sellars, *Science, Perception, and Reality*, 31-32
[54] Ibid., 32.
[55] Ibid., 32-33.

objects in space and time.

If we consider the concept of thought abstractly, i.e. disregarding the essential character of thoughts and regarding only their functionality, "*as items which can occur in patterns of relationships which are analogous to the way in which sentences are related to one another and* to the contexts in which they are used", then, "then there is no barrier *in principle* to the identification of conceptual thinking with neurophysiological process," and there would be nothing qualitative about the concept remaining.[56] The identification would be easier to accomplish, than accomplishing the identification of physical things in the manifest image with complex systems of physical particles, believes Sellars. "And in this key, if not decisive, respect, the respect in which both images are concerned with conceptual thinking, which is the distinctive trait of man, *the manifest and scientific images could merge without clash in the synoptic view*."[57] Now, there are several problems with these findings. Firstly, what remains of a thought if only its functionality remains? Kant's words "thoughts without content are empty, intuitions without concepts are blind" are quite applicable here.[58] And secondly, if the principles used by neurophysiology are 'special cases' of particle physics, then isn't the identification still harder to achieve because the interaction of particles in the brain lies at the basis of thought and neurophysiological processes? Still, the point Sellars is making here is clear: one could try to understand conceptual thought as a neurological process, and by doing so conjoin the manifest image with the scientific image into a stereoscopic view. Whether they can merge *without clash* remains to be seen, at least when it comes to the concept of self-consciousness.

In the manifest image, conceptual thinking is constructed by analogy with observable speech, and sensation is constructed by analogy with its external cause, the sensation being brought about by an object. It could therefore be supposed that through this abstract and functional way of regarding sensation, we could once again identify them with complex systems of physical particles, believes Sellars. This is difficult to assume, however, because it is easier to believe that thoughts are completely different than overt speech when their *role* is concerned, than it is to believe that sensations are radically

[56] Sellars, *Science, Perception, and Reality*, 34.
[57] Ibid., 34.
[58] Immanuel Kant, *Critique of Pure Reason*, trans. Paul Guyer and Allen W. Wood (Cambridge: Cambridge U.P., 2000), 193-194 [A51/B75].

different by analogy from their external causes, because the analogy concerns the *quality* itself.[59] The role of sensation is qualitative categorically. This finding leads Sellars to a crucial issue: "can we define, in the framework of neurophysiology, states which are sufficiently analogous in their *intrinsic* character to sensations to make identification plausible?"[60] He believes that this cannot be the case, because of the issue of 'ultimate homogeneity' which was discussed earlier. The surface of an object is a surface with a certain colour, and all smaller areas of this surface also have a colour, ad infinitum, whereas the state of a cluster of neurons has ultimately regions which are single neurons, which also holds for the finer grained level of biochemical processes.[61] The main question that arises here is: why should the analogy be made illicit by the limitations of the levels of reduction? If a colour expanse is reduced in size in sensation, should then also the size of the group of neurons reduce in an analogous manner? We are limited in our visual perception, when it comes to the size of objects. Even though a molecule may have a surface with a colour, we cannot perceive it solely with our eyes. We can make it visible with a microscope, but we do not see it in it's true size. Would we need less than a single neuron to have a sensation of a very small object? There are probably many neurons involved in every sensation. It must be said that Sellars makes explicit that it does not make sense to say of the particles of physical theory that they are coloured. He is interested here in the reoccurring problem of Eddington, the two tables of being: "*how to reconcile the ultimate homogeneity of the manifest image with the ultimate non-homogeneity of the system of scientific objects*."[62] Thus, it is this homogeneity of perceptible qualities which prevents identifying it with complex system of physical particles, which also stands in the way of identifying sensations with complex neural processes, according to Sellars. Sensations are essential to construct appearances in the manifest world, but the reality of the scientific image should be given primacy, whilst it is not identifiable because of the absence of ultimate homogeneity. Sellars believes that we are therefore confronted with an antinomy: (a) Either the neurophysiological image is incomplete, or, (b) the ultimate homogeneity of the sense qualities and the qualities themselves, are mere appearances and do not exist in space and time at all. In regard to (a), Sellars suggests that a solution

[59] Sellars, *Science, Perception, and Reality*, 35.
[60] Ibid., 35.
[61] Ibid., 35.
[62] Ibid., 36.

could be the supplementation of new objects, representative of activity of the visual cortex as a system of particles, which he calls 'sense fields', that do have ultimate homogeneity.[63] Personally, I think (b) is the most credulous alternative. Sense qualities are *qualities* which are very subjective in that they differ in how they are perceived and imagined by each individual. This is not to say that I believe in hard idealism; there is a lot of common ground between all human beings. We all live in the same objective reality, and we all have comparable ways of processing experience, but the colour of an object is something that we make of it. As a *Ding an sich* an object has no colour. What it does have is a certain composition of molecules and atoms, which reflects light in a particular way, but when this reflected light is perceived by us, it us only us who see it as having colour. Admittedly, it could be argued that I am rejecting a stereoscopic view in favour of a scientific image with this explanation. This is not the case, however, for this line of thinking, considering sense qualities as mere appearances, is very much part of a manifest image when the explanation of the reflection of light is not included, which only strengthens it.

Sellars wonder whether the situation is irremediable, for it seems that the manifest and the scientific conception of sensations and conceptual thinking beheld together, can only fit into a synoptic view as parallel processes, for they can only be truly regarded as mere analogues, and are not perfectly identifiable as of yet, or at least in the time that Sellars wrote. He could thus be placed in the reductive materialist's camp: "There are sensations, but they are not distinct from brain-processes."[64] He is not a materialist because they cannot be regarded as being identical, yet. And he is certainly not an eliminative materialist because for him the term 'sensation' does denote something. The dualism could be avoided if the scientific image as a whole is to be regarded as a symbolic device, but this is not what Sellars wants. The scientific image is simply not yet complete. At the end of the section, Sellars comes up with a beautiful and almost prophesying suggestion, which might appeal to the physicist specialised in quantum mechanics. He argues that if it should turn out that the smallest physical particles could one day be cut up into interacting particles, so that the interactions would be the most primitive entities, then consciousness does not have to be identified with systems of particles, but with non-particulate

[63] Sellars, *Science, Perception, and Reality*, 36.
[64] Lycan and Pappas, "What is eliminative materialism?," 149.

interactions, whatever they may be. In this non-particulate image, the qualities of sense are a dimension of natural processes occurring in connection with cut up complex physical processes, which become the complex system that is the central nervous system.[65] Thus, the problem of ultimate homogeneity would be overcome, as the non-particulate interactions would be considered ultimately homogenous and undividable.

VII. PUTTING MAN INTO THE SCIENTIFIC IMAGE – THE SEARCH FOR A COMPLETE IMAGE

The scientific image is, as it stands, still a very idealistic image. Even if the scientific image could remove all the problems posed above, it's primacy would be barely of the ground, believes Sellars. "There would remain the task of showing that categories pertaining to man as a *person* who finds himself confronted by standards (ethical, logical, etc.) which often conflict with his desires and impulses, and to which he may or may not conform, can be reconciled with the idea that man is what science says he is."[66] I believe that the primacy of the scientific image will always remain an idealisation, for how can you take that which is human, that which is us manifestly, away from humanity? Even though we may achieve high levels of rationality in the future, we will try to maintain our humanity. Perhaps when the limitations of our perception and thinking are removed by advanced technologies, we will be able to regard the world from a purely scientific point of view, but then again, would we still be humans? Sellars argues that besides the 'free will' objection that the proposed scientific reconstruction of the image of man faces, in which human beings are considered to be responsible agents making genuine choices between genuine alternatives, that could perhaps one day be met with extraordinarily complex defined concepts in the scientific image, it fails decisively on another count: such a reconstruction is logically impossible in principle. Unfortunately, Sellars does not argue the point explicitly, but provides some remarks which contain 'the essential clues'.[67]

If the reconstruction is impossible then there seem to be three alternatives: "*(a)* a dualism in which men as scientific objects are contrasted with the 'minds' which are the source and principle of

[65] Sellars, *Science, Perception, and Reality*, 37.
[66] Ibid., 38.
[67] Ibid., 38.

their existence as persons; *(b)* abandoning the reality of persons as well as manifest physical objects in favour of the exclusive reality of scientific objects; *(c)* returning once and for all to the thesis of the merely 'calculational' or 'auxiliary' status of theoretical frameworks and to the affirmation of the primacy of the manifest image."[68] Sellars finds these three alternatives unsatisfactory, and he presents a fourth, which can be represented as: *(d)* the manifest framework of persons does not need to be reconciled with the scientific image, but rather joined to it. Although Sellars presents the impossibility as a logic one, the core of it seems to lie in *intentionality* and *community*. A person construes its behaviour in regard to a group, or community. The fundamental principles of a community can be equated with its most general intentions. The logical aspect of the impossibility is that to think these thoughts is not to classify or explain but to 'rehearse an intention'.[69] Now, it is because of the intentionality of the individual and the group, which is strongly reminiscent of the 'free will' objection, that the scientific image cannot provide the absolute truth about what man thinks he is as man-in-the-world, and thus the conceptual framework of persons needs to be joined to it.

> Thus, to complete the scientific image we need to enrich it *not* with more ways of saying what is the case, but with the language of community and individual intentions, so that by construing the actions we intend to do and the circumstances in which we intend to do them in scientific terms, we *directly* relate the world as conceived by scientific theory to our purposes, and make it *our* world and no longer an alien appendage to the world in which we do our living. We can, of course, as matters now stand, realize this direct incorporation of the scientific image into our way of life only in imagination. But to do so is, if only in imagination, to transcend the dualism of the manifest and scientific images of man-of-the-world.[70]

Thus, it seems as if we must reduce the primacy of the scientific image in order to come to a viable stereoscopic view of man in the world, which seems reasonable. For if we would not, little that is truly human would remain. Whereas Sellars believes in stereoscopic view in which neither image should overwhelm the other, the eliminative materialist strives for the elimination of common sense

[68] Sellars, *Science, Perception, and Reality*, 38-39.
[69] Ibid., 39-40.
[70] Ibid., 40.

manifest ideas when it comes to theory of mind in favour of a scientific account. In the next chapter, we will discuss neuroscientific conceptions of self-consciousness, and see whether primacy is given, or ought to be given, to the scientific image, or if we should strive for a stereoscopic view. With Sellars' theoretical framework kept in mind, let us now discuss the concept of self-consciousness proper, and let us search for a complete image of self-in-the-world.

Chapter Two

Neuroscientific Conceptions of Self-Consciousness

In this chapter, I will present six case studies taken from two books and four articles written by neuroscientists whose research revolves around the topic of self-consciousness, in order to establish to which extent one image, or both images appear in their research. I have also selected these case studies as some of the ideas that we will come across offer possibilities for comparative analysis with respect to philosophy of mind, as for example regarding the absence of the dualism of body and mind. I will first discuss the concept of self-consciousness as proposed by the neurologist Antonio Damasio in his book *The Feeling of What Happens*.[71] This book can be regarded as a theoretical cornerstone on self-consciousness in neuroscience, as it is very often referred to in neuroscientific literature. The main topic of the book is 'our stepping into the light of consciousness' as Damasio poetically describes it. "I write about the sense of self and about the transition from innocence and ignorance to knowingness and selfness."[72] Damasio's goal is to consider the biological circumstances that permit this transition from an interdisciplinary perspective. "The traditional worlds of philosophy and psychology have gradually joined forces with the world of biology and created an odd but productive alliance."[73] Damasio makes use of scientific theories based upon the findings of experiments with MRI, fMRI and PET scans, and he makes of findings regarding molecular events by neuroanatomists, neurophysiologists, neuropharmacologists, and neurobiologists, to study the organism's private mind, the organism's public behaviour, and the organism's brain, and how they are related to each other.[74] The joining of forces and the emphasis on scientific methodology and theory formation already point in the direction of a stereoscopic view, the likes of which Sellars had in mind. In this chapter, I set out to prove that neuroscientific ideas on self-consciousness are aligned with, and reliant on the manifest image of man-in-the-world, which tells us that eliminative materialism is a flawed philosophical position when it

[71] Antonio Damasio, *The Feeling of What Happens: Body, Emotion and the Making of Consciousness* (London: Vintage, 2000).
[72] Ibid., 4.
[73] Ibid., 13. It is interesting that Damasio groups philosophy and psychology together; a perennial tradition and a relatively recent area of expertise. Perhaps he uses the word 'traditional' here because psychologists and philosophers are not as receptive to changes in theory as biologists are.
[74] Ibid., 14-15.

comes to investigating the self and self-consciousness.

Damasio regards *consciousness* as a turning point in the history of life. Consciousness' standard dictionary definition is presented "as an organism's awareness of its own self and its surroundings."[75] To overcome 'the obstacle of self', i.e. understanding how it is that we know that certain feelings occur within our own organism, we must come to understand the neurological underpinnings of self-consciousness, which might also help elucidate the neural underpinnings of consciousness in general, believes Damasio. He argues that, at its most basic level, consciousness provides us with the need to survive and develop a concern for the self. At its most complex level, consciousness provides us with a concern for other selves.[76] Thus, in this way of describing consciousness, it is the awareness of "I" being a person, and of "I" being a member of a community aware of the other selves within the group, which is here presented in the tradition of perennial philosophy. Damasio also presents an interesting anecdote which is explained in a more scientific way. It is a story about a man who stopped interacting with Damasio as a person, as he did not respond to his name, and looked and moved about in a room randomly, which Damasio describes as an impairment of consciousness. "Neurologically speaking, he had an absence of seizure followed by an absence automatism, two among the many manifestations of epilepsy, a condition caused by brain dysfunction."[77] He interpreted the meaning of this impairment of consciousness as a transition to a mind deprived of the sense of self, leading him to believe that the self is an indispensable part of the conscious mind. Already, we come across a stereoscopic view of the impairment of consciousness, explained in the scientific sense as an episode of epilepsy, and explained manifestly as the absence of self.

Damasio's own definition of consciousness is as follows: "Consciousness is an entirely private, first-person phenomenon which occurs as part of the private, first-person process we call mind."[78] This position is hard to place with regard to Sellars' essential dualism of the perennial philosophy, i.e. the two types of causal impact of the world on the individual; of direct causal influence of the world as intelligible on the individual mind, and of indirect causal influence of the world as intelligible on the

[75] Damasio, *The Feeling of What Happens*, 4.
[76] Ibid., 3-5 and 8.
[77] Ibid., 6.
[78] Ibid., 12.

individual mind mediated by the group as a group phenomenon, as consciousness is described as something that is entirely private. In their book, *Philosophical Foundations of Neuroscience*, Bennett and Hacker rightfully criticise this 'private mind argument', which is common among neuroscientists, as (self-)consciousness is something which can be explained and shared by the individual through language.[79] Damasio studies consciousness by looking at the organism, the human and the object, and the relationships between the latter two. He believes that consciousness constructs knowledge about two 'facts': "that the organism is involved in relating to some object, and that the object in the relation is causing a change in the organism."[80] These two 'facts' are important to keep in mind when we look into Damasio's theories in the following two sections, as they lie at the core of Damasio's theories on both consciousness and self-consciousness. Furthermore, they touch upon certain fundamental philosophical ideas that will be discussed in the next chapter.

I. THE PROBLEM OF CONSCIOUSNESS

Damasio divides the problem of consciousness, from the perspective of neurobiology, into two related problems. The first problem is understanding how the brain engenders the mental patterns Damasio calls 'the images of an object.' Objects are described by Damasio as entities like persons, places, sounds, aches, and even emotions. Images stand for mental patterns in any form of sensory modality, be it visual, auditory, olfactory, gustatory, and somatosensory. It is the problem of how we get a 'movie-in-the-brain'. If it is to be solved from the neurobiological perspective, neuroscientists must find out "how the brain makes neural patterns in its nerve-cell circuit and manages to turn those neural patterns into the explicit mental patterns which constitute the highest level of biological phenomenon, which I like to call images."[81] He describes the flow of said images as *thought*. There are two issues which arise here. The first is Damasio's use of the word 'image', as it is rather confusing to use the word 'image' for such an enormous range of items, especially if we take into account the two possible ambiguous uses of this word as described in the first chapter. Couldn't such images more appropriately be called thoughts and feelings? For how is the feeling of pain in any way an image? Or how is the hearing of a melody an

[79] M.R. Bennet and P.M.S. Hacker, *Philosophical Foundations of Neuroscience* (Hoboken: Wiley, 2003).
[80] Damasio, *The Feeling of What Happens*, 133.
[81] Ibid., 9.

image? The second issue is the problem of translating neural patterns into the mental patterns which Damasio calls images, as they are analogues, but not necessarily identical. Let us therefore see what neuroscientists can tell us about the similarities between those patterns, in particular with regard to the concept of self-consciousness.[82]

The second problem of consciousness is understanding how the brain engenders a sense of self in the act of knowing. It is parallel with the first problem of engendering mental patterns for an object. Damasio describes this sense of self as a presence which is always there when awake, it is a presence of yourself as the feeling of what happens when your being is modified by the acts of apprehending something, of knowing. "The presence must be there or there is no you."[83] It is a feeling that signifies you as observer and potential actor. There is this feeling of a relation between you and any object you think about. If you understand the self in this philosophical sense, you can attempt to solve the problem by discovering the biological underpinnings for the ability we have of constructing mental patterns which convey, automatically and naturally, the sense of a self in the act of knowing.[84]

It seems that the first problem, the engendering of mental patters for an object by the brain, is strongly related to the second. To understand how the brain makes neural patterns in its nerve-cell circuit and manages to turn those neural patterns into the explicit mental patterns which constitute the highest level of biological phenomena, images, would also entail that one could understand the mental patterns of the sense of a self. The second problem, the engendering of a sense of self in the act of knowing, is in fact a part of the first problem, for the sense of self must accompany all the 'images' that Damasio is writing about. "Consciousness, as we commonly think of it, from its basic levels to its most complex, is the unified mental pattern that brings together the object and the self."[85] The second problem pertains to the observer of the movie-in-the-brain. The brain 'generates' the sense that there is an owner and observer for this movie, making the self an appearance.[86] This could be interpreted as a form of idealism; if the self is merely an appearance which is created by the brain, like the 'images' are, then what remains

[82] Damasio, *The Feeling of What Happens*, 8-9 and 317-318.
[83] Ibid., 10.
[84] Ibid., 10-11.
[85] Ibid., 11.
[86] Ibid., 11.

of that which is you? If you would truly believe this, then there is no truth to anything. Consider the *cogito* argument: When the "I" in "I think and therefore I am" is only an appearance provided by the brain, then truly nothing is certain. However, I do not think that this is what Damasio is hinting at. I believe he means to convey that everything you are is in fact material, and any transcendental idea one might have of the self, like for example the self as a soul which is a separate entity, is disregarded; the self is only that, which the brain as matter can produce.

II. FROM PROTO-SELF TO CORE SELF TO AUTOBIOGRAPHICAL SELF

Damasio argues that there are three kinds of self, one form of which is the *core* self, "a transient entity, ceaselessly re-created for each and every object with which the brain interacts."[87] This self arises in *core consciousness*, which provides us with a sense of self in the here and now.[88] A second form of self is the *autobiographical* **s**elf, which Damasio describes as the more traditional notion of self, linked to the idea of identity and depended on memories.[89] Thus, the first form of self lasts only for moments, whereas the second spans across a lifetime. Are both forms of self always there? The former must be, since it is related to any 'object' with which the brain interacts. The latter perhaps not so much, since it could be argued that it is not a part of imminent self-consciousness, but that is more of an 'image' in the sense that Damasio uses this term; one thinks of the autobiographical self by accessing memories as objects of thought, as knowledgeable content in consciousness. It is interesting that when Damasio discusses these two forms of self, he does not bring up much scientific theory postulation; he mostly philosophises about it in the lines of perennial philosophy. This is different when he describes the third form of self: the *proto-self*, which is the nonconscious forerunner of the two mentioned forms of self. "The deep roots for the self, including the elaborate self which encompasses identity and personhood, are to be found in the ensemble of brain devices which continuously and *nonconsciously* maintain the body state within the narrow range and relative stability required for survival."[90] These 'brain devices' regulate our life

[87] Damasio, *The Feeling of What Happens*, 17.
[88] The assumption that the 'core self' lasts for only moments, and that it is constantly recreated, is quite problematic, and it is an issue I will return to in section III of Chapter Four, "True materialism and Neuroscientific Materialism: Body and Mind".
[89] Damasio, *The Feeling of What Happens*, 17.
[90] Ibid., 22.

processes, like breathing, the beating of our heart and he maintaining of our internal milieu. Damasio describes these parts of the brain as not being 'free to roam at all', whereas other parts, for example those engaged in perception, are.[91] Although Damasio does not say this explicitly, it seems to follow that the *proto-self* stems from the parts of the brain which are unfree, engaged in automated management of the organism, whereas in regard to the 'newer' selves, parts of the brain are engaged that are free to roam. What he does make explicit is that the self is grounded in a collection of nonconscious neural patterns which regulate the body proper of the individual. When we look 'behind' individual singularity, argues Damasio, we find stability which relates to the self, for the self requires structural invariance to a certain extent, so it can provide continuity of reference.[92] I think that Damasio wishes to convey to his reader that the self is strongly related to the physical body of the individual, not through the body itself per se, but through the 'devices' in the brain which regulate the processes within the body (brain over body). The maintaining of stability which enables an organism to maintain its life is essential, and it might be a blueprint for what might become a self in the mind, suggests Damasio.[93] It could be countered that it is not the case that these devices necessarily entail the existence of a self in other organisms, but one can see how these two features of the human being are relatable in this common stability, stability which enables survival, and stability which enables self-consciousness.

> *The proto-self is a coherent collection of neural patterns which map, moment by moment, the state of the physical structure of the organism in its many dimensions.*[94]

There are different parts of the brain which make up the proto-self. These parts are all involved in regulating the processes within our bodies, but they have nothing to do with knowledge, perception, or language, and it is therefore that Damasio believes that the proto-self is not something we are conscious of. Another important aspect is that it does not occur in one place; it is a result of the interaction of neural and chemical signals among a set of regions. This also holds for the other forms of self which Damasio distinguishes. "Phrenological thinking must be resisted at all costs."[95] As we shall

[91] Damasio, *The Feeling of What Happens*, 21.
[92] Ibid., 134-135.
[93] Ibid., 136.
[94] Ibid., 154.
[95] Ibid., 154.

see, this is a widely accepted view in contemporary neuroscientific research.

I will now provide a short overview of the brain structures which Damasio describes as essential for implementing the proto-self. Damasio writes that his hypothesis about the proto-self could be tested by formulating predictions about the effects of damage to these structures, which would entail the loss of consciousness, and to comparable structures which are not required for implementing the proto-self, but which may or may not result in the loss of consciousness. A key structure is formed by a multitude of brain-stem nuclei. These function as regulators of body states and as mappers of body signals, as they signal the overall current body state, based upon the signals entering the brain through the spinal cord. The hypothalamus, together with the basal forebrain, holds a register of the state of the internal milieu, and it reacts to deviations, for example by initiating a change in the concentration of ions or hormones. Damasio writes that by maintaining a sort of register "[t]he hypothalamus contributes to the current representation of the body."[96] Does this mean that when we think of our body, or when we feel it, the hypothalamus plays a role in such a representation in the sense that it can tell us whether our internal milieu is in order? How does one feel, or think of, one's own internal milieu? These are not easy questions to answer, for the combining of first-person subjective experience with third-person scientific knowledge is something which is hard to achieve; a core problem especially when it comes to self-consciousness. Regardless, what is of importance in Damasio's theory, is that these required structures provide the roots for the self, the 'something-to-which-knowing-is-attributed', for they are key in providing the primary current body states; they are the first things in your brain which could give you a sense of self.[97]

The 'something-to-be-known', e.g. an object, is strongly related to 'something-to-which-knowing-is-attributed' in consciousness. Damasio tries to establish what the mechanisms are that the brain uses to represent the relationship between the object and the organism, which he describes as "the causal action of the object on the organism and the resulting possession of the object by the organism."[98] He argues that our organisms are changed by objects, in the sense that we internally construct an image

[96] Damasio, *The Feeling of What Happens*, 156.
[97] Ibid., 159.
[98] Ibid., 167.

of these objects, which is the exhibiting of a specific kind of wordless knowledge. The emergence of this knowledge in its simplest form is 'the feeling of knowing'.[99] His answer to how such knowledge is acquired, and why it comes to us firstly in the form of a feeling is: "*core consciousness occurs when the brain's representation devices generate an imagined, nonverbal account of how the organism's own state is affected by the organism's processing of an object, and when this process enhances the image of the causative object, thus placing it saliently in a spatial and temporal context*."[100] The generation of this account is the source of self in the act of knowing; through the becoming aware of one's own organism being altered by the processing of an object, one becomes aware of the existence of one's self. It could be argued that this form of consciousness, being aware of the alteration of one's own state, is in fact self-consciousness. But because of the second part of the hypothesis, which is much closer to the common sense conception of consciousness, enhancing and placing images of objects, it is clear that Damasio wishes to intertwine both the self and perceptual awareness of one's surroundings within his idea of consciousness; the self is a fundamental part of consciousness.

The nonverbal account, when it is flowing in your stream of thought, gives you reason to believe that you are conscious, because it exhibits the knowledge, or feeling, that your proto-self has been altered by the image of an object. Damasio argues that in your personal narrative of thought you are the protagonist in the act of knowing, and you feel your core self, transiently, because it renews with each new object that appears in thought, from outside or from inside through memory. It is you who does the experiencing. The main difference with the proto-self is that the core self is conscious, whereas the former is nonconscious. The basis for the former is also a feeling, which "arises in the re-representation of the *nonconscious proto-self in the process of being modified* within an account which establishes the cause of the modification."[101] Thus, the 'something-to-which-knowing-is-attributed', as the proto-self, is originally represented within an account of the modification by the object, which is then consciously represented *again* in core conscious and felt as a core self. It therefore seems that there are multiple layers of self active at the same time in Damasio's view, at least during states of core consciousness.

[99] Damasio, *The Feeling of What Happens*, 168-169.
[100] Ibid., 169.
[101] Ibid., 172.

"The first trick behind consciousness is the creation of this account, and its first result is the feeling of knowing."[102] Perhaps Damasio calls this a trick because it is the brain that tricks us into believing that we have consciousness and that we are self-conscious. The result of the trick is that it provides us with our core consciousness, with the feeling that it is you, who is the imaginer of objects. This thought might give rise to the philosophical anxiety that our brain provides us only with illusions, making the existence of an objective outer reality uncertain. Damasio tries to comfort us in an endnote, however: the nonverbal account, knowing, and self are not fictional or illusions. All living individual organisms, the objects, and relationships are consistent, systematic and widespread occurrences, argues Damasio, and because they respect a relative, be it non-absolute, truth, they are not fictional.[103] But what if this seeming consistency is yet another trick of the brain?

Even though all the moments in which we feel our existing selves are of a temporal nature, there is something which aids us in stringing together the facts about our existence: memory. Because of our capability to remember we are able to develop autobiographical memory, in which we hold who we have been both physically and behaviourally, and who we want to be in the future. Damasio describes it as aggregate memory which develops as life unfolds. "The real marvel, as I see it, is that autobiographical memory is architecturally connected, neutrally and cognitively speaking, to the nonconscious proto-self and to the emergent and conscious core self of each lived instance."[104] The emergence of the core self from the feeling of knowing, and the proto-self's relation with the body's internal milieu are 'enriched' by the display of facts from autobiographical memory. Thus, the autobiographical self depends on both core consciousness and continuous reactivations of autobiographical memories. The capacity to remember who you have been, and your reasoning ability, enable the existence of your autobiographical self. It does not require language. Animals like apes and dogs have an autobiographical self. They do not, however, 'possess a person', due to a lesser endowment of memory and reasoning ability, and because they lack language.[105] The basis of the autobiographical self is formed by the practically invariant proto- and core selves, the latter of which is founded upon the

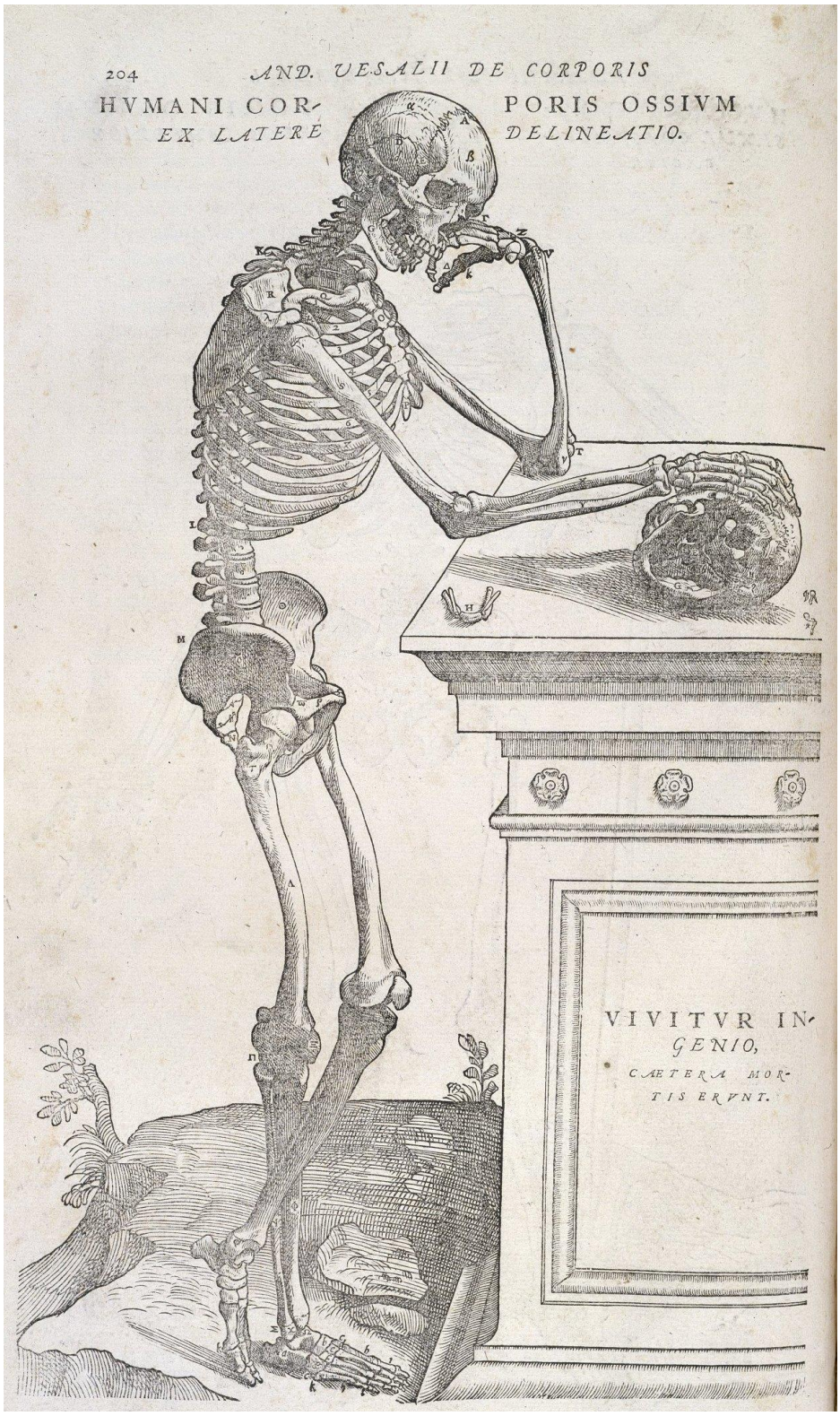
[102] Damasio, *The Feeling of What Happens*, 172.
[103] Ibid., 351.
[104] Ibid., 173.
[105] Ibid., 198-199.

former. These notions of self are thus strongly intertwined, physically and cognitively, but at the same time they seem layered and only partially dependent on each other.

Although Damasio has an elaborate theory on which areas in the brain are responsible for these notions of self, involving scientific findings, the way in which he describes them verge, I believe, to the manifest image of man-in-the-world, for it is not unthinkable that one could come up with a comparable theory on the self without the postulation of imperceptible entities, and principles pertaining to them. When brain areas and their functionality are not included, the positing of the existence of a proto-self based upon the representation of an internal milieu, a core self based upon a feeling of knowing occurring whenever the proto-self is modified by an object, and an autobiographical self based upon the former two in combination with memory, seems to be a rather manifest approach to self-consciousness, which could be placed within the Platonic tradition. I do not hereby mean to discredit Damasio's theory, for the scientific knowledge he brings up in support of his philosophical views on the self, as is the case when he discusses the neuroanatomical basis for the autobiographical self, and in the section in which he presents scientific evidence for a role of proto-self brain structures in consciousness, to name just two examples, is very convincing.[106] However, I do believe that the manifest aspects of his theory have the potential of being improved with the aid of the philosophical conceptions of self-consciousness presented in the next chapter. For example, it is not clear to what extent the three levels of self form a unity. I will return to this issue, and others, in the fourth chapter.

Damasio offers us a very particular view of self-in-the-world, one in which the scientific image and the manifest images are both apparent, and one in which the former does not seem to hold primacy over the latter. Based upon this manifest and neuroscientific account alone, it could be argued that our ordinary notions of mental states, like images and forms of self, do have a home at a high level of analysis in a sophisticate account of the mind, which makes the position of eliminative materialism rejectable. But this will not suffice, however, and to strengthen my case I will present five other case studies with which we might be able set additional steps towards a stereoscopic view.

[106] Damasio, *The Feeling of What Happens*, 219-222 and 236-260.



Engraving by Andreas Vesalius of a skeleton preoccupied by a skull, from *De Fabrica Humani Corporis* (1543).[107]

[107] Image from: Gerald M. Edelman and Giulio Tononi, *A Universe of Consciousness: How Matter Becomes Imagination* (New York: Basic Books, 2000), 13.

III. PHILOSOPHY, LANGUAGE, SCIENCE, AND THE SELF

I will now discuss *A Universe of Consciousness*, a book by Gerald Edelman, biologist, neuroscientist and philosopher, and Giulio Tononi, neuroscientist and psychiatrist. The book is described in *Nature* as an introduction to Edelman's ideas on consciousness, the explaining of which has become 'the Holy Grail of modern neuroscience'.[108] The authors discuss self-consciousness quite extensively from different perspectives. They are influenced, in some respects, by Antonio Damasio's work.

Edelman and Tononi make a distinction between primary consciousness, which is found in animals with brain structures that are similar to human brain structures, and higher-order consciousness, which is found in humans. Animals with primary consciousness, like apes, appear to have the ability to construct a mental scene, but they lack the semantic and symbolic capabilities that we have, and they have no true language. A mental scene is generated through the simultaneous processing of large amounts of diverse information, which has the purpose to direct behaviour.[109] We here find a different conception of thought. Whereas Damasio made use of the simpler concept of images in a flow of thought, Edelman and Tononi speak of a more complex 'mental scene', which seems comparable to an abstract form of thinking due to the collection of diverse information that is processed simultaneously.

Higher-order consciousness, argue Edelman and Tononi, minimally requires a semantic capability, and in its most developed form, it requires a linguistic capability. "*Higher-order consciousness (which flourishes in humans and presupposes the existence of primary consciousness) is accompanied by a sense of self and the ability in the waking state explicitly to construct past and future scenes*."[110] Animals with primary consciousness do not have a 'sense of self', but what this sense of self is exactly, is not defined. The mentioned ability seems comparable to Damasio's concept of the autobiographical self, the common-sense conception of self, but Damasio, however, does believe that certain animals have an autobiographical self. Primary consciousness, argue Edelman and Tononi, is required for higher-order consciousness as it provides the latter's foundations. This, in turn, seems comparable to Damasio's grounding of the autobiographical self on the proto-self. Edelman and Tononi

[108] Gerald M. Edelman and Giulio Tononi, *A Universe of Consciousness: How Matter Becomes Imagination* (New York: Basic Books, 2000), cover.
[109] Ibid., 102-103.
[110] Ibid., 102.

explicitly endorse Damasio's ideas on the proto-self. The earliest conscious dimensions and discriminations are concerned with the inner state of the body, reacting to its environment. Proprioceptive, kinaesthetic, somatosensory, and autonomic components are bodily functions of which we are only dimly aware, but which are fundamental for our bodily wellbeing. They provide a basis for a 'bodily based neural reference space', out of which memories based on signals from the outer world ('nonself') are elaborated. Through the development of linguistic skills, higher-order consciousness appears. The nameable self that is differentiable from within appears as man becomes conscious of his own consciousness. Higher-order consciousness and self-consciousness are thus rooted in primary consciousness and the proto-self.[111] In a later chapter, "Language and the Self", Edelman and Tononi primarily discuss their ideas on self-consciousness. "*In this chapter, we consider several issues of central human significance in a new light, relating our efforts to philosophy and to science itself, and offer insights into what we may and may not expect from a scientific view of consciousness*."[112] Let us see to what extent they provide a Sellarsian view of self-in-the-world.

Animals with primary consciousness can generate mental scenes, which are determined by their environment and by unconscious subcortical activity. They have biological individuality, but no true selves, for a true self is a self that is aware of itself, argue Edelman and Tononi. They link this self to memory, as they write that animals only have a "remembered present", and they have no concept of the past or the future. Only with semantic capabilities, which appeared in humans in the course of evolution, can a creature have these concepts. When the full linguistic capability based on syntax appeared in precursors of Homo sapiens, higher-order consciousness, mediated by a new type of memory, became highly developed, enabling consciousness of consciousness.[113] Furthermore, it is after the emergence of higher-consciousness, that the self can be constructed from social and affective relationships, write Edelman and Tononi.[114] The above is quite a philosophical picture of self-consciousness, which could be described as being aware of one's own conscious enabled by autobiographical memory, and it reminds us of Sellars' pre-historical description of man first becoming aware of himself in the world.

[111] Edelman and Tononi, *A Universe of Consciousness*, 173-175.
[112] Ibid., 193.
[113] Ibid., 193-194.
[114] Ibid., 197.

However, it is different in the sense that it is supported by a scientific image in which the concept of 'reentry' is a key element. It is described as a process which occurs in the interaction between a large number of groups of neurons. This interaction occurs rapidly and reciprocally. Neuron groups which have similar functions are often connected in this reciprocal way, and they respond simultaneously to certain stimuli. Thus, reentry is the process of signalling back and forth between groups of neurons along reciprocal connections. Edelman and Tononi believe that this offers the key to understanding how various functionally segregated properties of brain areas are integrated, despite the absence of a central coordinative area.[115]

The specific re-entrant connectivity between the brain systems for language and the already existing conceptual areas in the brain was a key step in the evolution of higher-order consciousness, state Edelman and Tononi, because it allowed reference to inner sates and objects or events by means of symbols. Through social interaction, the growth of a lexicon of these symbols was enabled in human beings, and this allowed for the discerning of a self within each individual consciousness. The authors support this idea with the following argument: "When narrative capabilities emerged and affected linguistic and conceptual memory, higher-order consciousness could foster the development of concepts of the past and future related to that self and to others."[116] Thus, the authors seem to believe that the development of the connectivity between language and conceptual thinking in the brain was essential for the emergence of the autobiographical self. They support this claim with theories on the evolution of the brain, as for example regarding the development of cortical and subcortical structures that are involved in phonological categorisation and memorisation of speech sounds, or with the emergence of the 'supralaryngeal space' engaged in producing the sounds that make up our voices.[117] These are examples of postulational theory formation and they fit within Sellars' concept of the scientific image of man, whilst they are used to support manifest ideas on self-consciousness.

Their scientific theory does not solely belong to the scientific image, for we can still discern philosophical assumptions that seem to have been taken for granted, but are not necessarily grounded in

[115] Edelman and Tononi, *A Universe of Consciousness*, 36 and 44.
[116] Ibid., 195.
[117] Ibid., 193-196.

scientific fact or postulation. For example, in no way is it clear *why* these symbols, which are indeed important for reference to inner states and the self in regard to others, are essential for the discrimination of a self within oneself. Couldn't a person know that she has a self, if she would have lived her entire life on a deserted island? What is also not ascertained scientifically, is why the development of concepts of the past and future related to the self and to others could only emerge after the development of narrative capabilities. A problem here is that it is not clear what the concepts of the past and the future are in this context, only that they are 'concepts of time past and time future'. But if we assume that they help in understanding memories and thoughts about the future, then language does indeed greatly extend the possibilities in regard to this understanding, for our capacity to understand is certainly strongly affected by transmissible knowledge, as many things are. It could be countered, however, that animals that have no narrative capabilities might also have their own concepts of the past and future related to some form of core self and others, especially animals with relatively high cognitive abilities, which could also be the case for the forefathers of Homo sapiens. And it must not be forgotten that there are other ways to communicate beside the verbal and narrative variant, which Edelman and Tononi do recognise when they speak of gestures belonging to a 'protosyntax', having developed into a syntax in humans.[118] We are quite certain that animals like elephants and great apes have the capacity to remember, which might entail that they have some sort of knowledge of what the past is, but it would, admittedly, be harder to argue that they are conscious of themselves living in the present and of having a past, even more so when it comes to the future. Still, it is not unthinkable that an animal might have some vague notion that there are things which are yet to come.

Edelman and Tononi present a manifest philosophical analysis of the self and language, when they discuss the internalist-externalist argument concerning the relation of subjectivity to the development of self. According to internalists there is an early subjective state in humans, i.e. as babies, which is followed by increasing differentiation of the self through social and linguistic interactions. According to externalists there is no early subjective state until language is acquired. It is only after the acquiring of enough language that the conceptual foundations of the self emerge. Edelman and Tononi

[118] Edelman and Tononi, *A Universe of Consciousness*, 196-197.

do not conform to one of the two, and instead argue that their notions of primary consciousness and higher-order consciousness allow the merger of the two views. The drive towards language in babies results in a rapid development of higher-order consciousness, a self-concept, and a notion of past and future, partly due to emotional exchange with their mothers, whereas this is not the case in other animals, which lack this drive. The train of thought that accompanies language is likely to be narrative and metaphorical, believe Edelman and Tononi.[119] This explanation does not tell us whether there is an early subjective state before language, only that their development goes hand in hand, thus I am not sure if we can truly speak of a resolution in regard to the internalist-externalist argument.

The authors conclude the chapter by discussing how their resolution can help us in understanding how qualia are affected by the emergence of the self. They argue that "an animal without semantic or linguistic capabilities lacks the symbolic memory that would allow it explicitly to relate its various qualitative experiences to a self. It also lacks the set of neural events that mediate that relationship by consciously linking past, present, and future."[120] It is because of these capacities that we can refine our discriminatory capacities to discern 'describable qualia' to a far greater extent than other animals. In humans, future acting based on previous experience is grounded in long-term memory and "the explicit memory of pain and pleasure phenomenally experienced by a linguistically grounded self", whereas for animals this is done solely with long-term memory, unconsciously.[121] Thus, again we come across a very manifest discussion of the self, one in which concepts like 'language' and 'qualia' are discussed in a very manifest way, supported by a scientific image, in which postulational concepts like 'neural events' are brought to the fore.

IV. BODILY SELF-CONSCIOUSNESS MEASURED

In this section, I will discuss a neurophysiological article on bodily self-consciousness by Hyeong-Dong Park and five other researchers.[122] This article is an example of a scientific approach towards better understanding self-consciousness. My aim is to make a comparison with the neuroscientific works

[119] Edelman and Tononi, *A Universe of Consciousness*, 197-198.
[120] Ibid., 199.
[121] Ibid., 199.
[122] Hyeong-Dong Park et al., "Transient Modulations of Neural Responses to Heartbeats Covary with Bodily Self-Consciousness," *The Journal of Neuroscience* 36(32) (2016): 8453-8460.

analysed above with respect to their two Sellarsian images of self-in-the-world. In their 'Significance Statement', the authors state that, although prominent views propose that internal signals from organs plays a crucial role in self-consciousness, which is a long-standing proposal found in philosophy and psychology, definitive experimental evidence supporting this idea is lacking, despite its recent impact in neuroscience. It is this evidence that the authors wish to provide with this article.[123]

> By combining electrical neuroimaging, analysis of peripheral physiological signals, and virtual reality technology in humans, we show that transient modulations of neural responses to heartbeats in the posterior cingulate cortex covary with changes in bodily self-consciousness induced by the full-body illusion. […] These neurophysiological data link experimentally the cortical mapping of the internal body to self-consciousness.[124]

According to the authors, in recent research the notion of bodily self-consciousness is referred to as the multisensory processing of bodily stimuli. In this research, the experimental manipulation of the perception of these multisensory stimuli have resulted in alterations in aspects of bodily self-consciousness, like self-identification with a body, or self-location of one's body. The authors present the findings of Damasio, among others, in regard to the role of the central nervous system in regulating the internal bodily milieu, providing continuity and stability for the organism and for self-consciousness. Other researchers have proposed the existence of functional links between interoceptive and exteroceptive signals, as between cardiac and visual signals, affecting the sense of self-identification with the body. "Based on these theoretical, behavioral, and clinical findings, we hypothesized that the neural processing of interoceptive signals would be linked to states of bodily self-consciousness as altered experimentally by the full-body illusion paradigm."[125] The 'the full-body illusion paradigm' is an essential aspect of the experimental and theoretical setup of the authors' research. In this setup, a person is synchronously stroked on the back, coupled with viewing stroking applied to a virtual body perceived through a head-mounted display, inducing increased self-identification with the presented virtual body compared with asynchronous stroking conditions.[126] In the edited volume, *The Embodied*

[123] Park, "Bodily Self-Consciousness," 8453.
[124] Ibid., 8453.
[125] Ibid., 8454.
[126] Ibid., 8453-8454.

Self, Adrian Smith suggests that it is problematic to uphold that this paradigm seems to support the idea that experience of the body as a whole involves representation of the body as a whole.

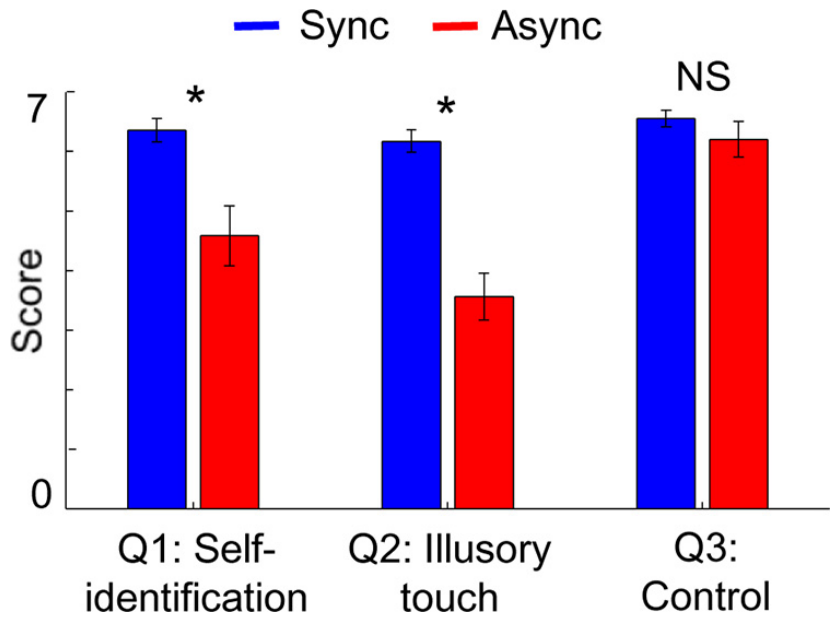
> Unfortunately, the question of whether or not there *are* full-body illusions is empirically under-determined, as putative full-body illusions are difficult to isolate from illusions involving composite parts that do not constitute a "full" or "whole" body. That is to say, a plausible alternative is that only representations of the body parts directly stimulated become subject to the experimentally induced bias, whilst other parts remain relatively (perhaps even completely) unaffected.[127]

This might also be the case in the experimental setup which Park and his colleagues used, for the participants were stroked only on their back, whilst wearing an EEG cap, and a head-mounted display (Oculus Rift Development kit).[128] Thus, the illusion may have only been of the back, which was the only part of the body that was subjected to stroking. It must be noted, however, that the authors use the word 'paradigm', a new paradigm perhaps which has not yet replaced an older one.

In order to test their hypothesis, the authors measured neural processing of cardiac signals through heartbeat-evoked potentials (HEPs), which they obtained by averaging electrophysiological signals time-locked to heartbeats. To investigate the relationship of HEPs with bodily self-consciousness, they tested (a) whether the HEP amplitude differs between two experimental conditions of the full-body paradigm, i.e. synchronous and asynchronous stroking inducing differential illusory states of bodily self-consciousness, and (b) whether such HEP amplitude modulations correlate with the questionnaire scores measuring the strength of illusionary states.[129] Using a 7-point scale (ranging from bottom-extreme to top-extreme) the following questions were judged by the participants: Q1 "How strong was the feeling that the body you saw was you?" Q2 "How strong was the feeling that the touch you felt was located where you saw the stroking?" Q3 "I felt my body [the participant's physical body] as usual, nothing changed."[130] It is somewhat problematic that these first two questions are quite

[127] Adrian J.T. Smith, "Comment: Minimal Conditions for the Simplest Form of Self-Consciousness," in *The Embodied Self: Dimensions, Coherence, and Disorders*, ed. Thomas Fuchs et al. (Stuttgart: Schatteuer, 2010), 38-39.
[128] An electroencephalogram (EEG) cap is used to measure electrical brain activity through electrodes placed along the scalp noninvasively.
[129] Park, "Bodily Self-Consciousness," 8454.
[130] Ibid., 8455.

suggestive, and that the experimental setup might indicate to the participants what the sought for answers are. The results of the questionnaire were that self-identification (Q1) and illusory touch (Q2) scores were significantly higher in the synchronous stroking condition. With regard to the control question (Q3) the results were deemed "not significant" with both results averaging between the scores of 6 and 7, the synchronous condition ending slightly higher on the scale. It would have been telling if the score would have been significantly lower during synchronous stroking; if it would have felt as if the virtual body was stroked, than we could certainly speak of self-identification and of a full-body illusion. The authors' conclusion that "these results demonstrate that the synchronous visuotactile manipulation induced changes in bodily self-consciousness", seems a bit strong, for self-identification with a virtual body due to illusory touch is not the same as being self-consciousness of one's own body, it might be more appropriately called 'bodily other-consciousness'.

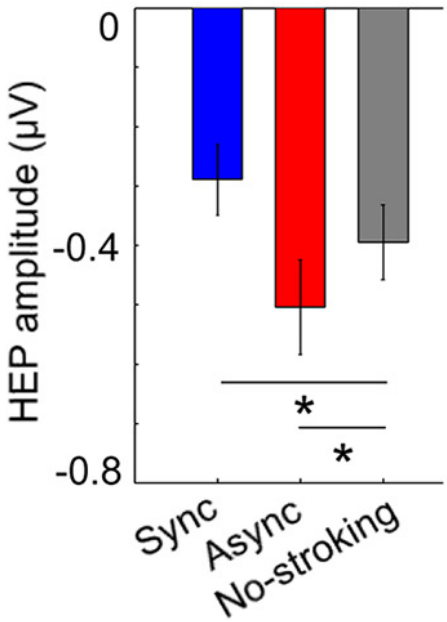


Questionnaire scores as represented in Park et al. *$p > 0.05$. NS, not significant.[131]

With the idea in mind that bodily self-consciousness is altered by synchronous stroking, the authors come to the conclusion that neural responses to heartbeats reflect bodily self-consciousness, based upon the measurements of the HEP amplitude, which differed significantly between the synchronous and asynchronous conditions over frontocentral regions of the brain (see image below).

[131] Park, "Bodily Self-Consciousness," 8456.

Although Park et al. convincingly show that the difference in EEG signals was really time locked to the heartbeat, and that the synchronous stroking induced the observed changes in HEPs, they do not prove that bodily self-consciousness is altered. It can be agreed upon that there is a correlation between neural responses to heartbeats and the subjective effects of the illusion, but this is does not give enough grounds to speak of "neurophysiological evidence supporting the proposed relationship between the brain's mapping of the internal body and self-consciousness, which to date has primarily been based on speculation, clinical studies, and behavioral studies."[132] It is granted that different amplitudes have been measured during the experiment, and that the authors have done an excellent job in locating the cortical sources, but the results only hint at the proposed conclusion in the form of scientific evidence. It must be noted that in the closing section of the article the authors themselves are more reserved about their findings. For example, they state that their results cannot demonstrate any causal relationship between HEP and changes in bodily self-consciousness, and they write that their findings *suggest* that neural processing of internal body states is a fundamental biological mechanism for the subjective aspect of conscious experience.[133]



Differences of the mean HEP amplitudes in Park et al.[134]

[132] Park, "Bodily Self-Consciousness," 8457.
[133] Ibid., 8457-8459.
[134] Ibid., 8456.

Let us now discuss this research in light of Sellars' theoretical framework. What is of great interest to this current investigation is that 'the proposed relationship between the brain's mapping of the internal body and self-consciousness', mentioned in the last citation can be easily placed within the manifest image of man. As I have argued above in the section on Damasio's work, apart from measurements and theoretical models about the brain, this is quite a philosophical thesis, for it is founded upon philosophical ideas about what self-consciousness is and on how this is related to the body. Any notion of Descartes' mind-body dualism seems to have completely evaporated in this view of man-in-the-world, for mind and body are both essential aspects in this theory on self-consciousness. Now, Park and his colleagues, who seem well-informed on neuroscientific theories on the self (Damasio is referred to several times, among others), seem to take the philosophical, or manifest aspect for granted, and they wish to provide scientific evidence with their findings. But as I have tried to show, this scientific image of man-in-the-world, is found to be somewhat wanting, for to speak of an alteration of bodily self-consciousness based upon the different answering of the two questions, "How strong was the feeling that the body you saw was you?" and "How strong was the feeling that the touch you felt was located where you saw the stroking?", when only the subjective experience of the self-identification and the illusion of touch were seemingly altered, goes too far. What is needed here, I think, is a more extensive philosophical explanation on how these two concepts are related to bodily self-consciousness. Although Sellars believes in the primacy of the scientific image, this case study shows that, sometimes, solely a scientific image of man is not enough. This also shows that, at least with regard to the concept of self-consciousness, philosophy still has a role to play. In the next section, I will analyse another experimental article, but with a different methodology, as it is a functional magnetic resonance imaging (fMRI) study.

V. NEURAL CORRELATES OF THE FIRST-PERSON PERSPECTIVE AS A CONSTITUENT OF SELF-CONSCIOUSNESS

In 2004, K. Vogeley and five co-authors investigated which brain regions are active in first-person perspective (1PP) visual perception and which are active in third-person perspective (3PP) visual perception, using functional magnetic resonance imaging, as they believe that taking the first-person perspective centred upon one's own body is constitutive for human self-consciousness, whereas taking

the third-person perspective is not.[135] Vogeley et al. provide a, albeit concise, philosophical justification in the introduction, primarily based upon Vogeley's earlier work.

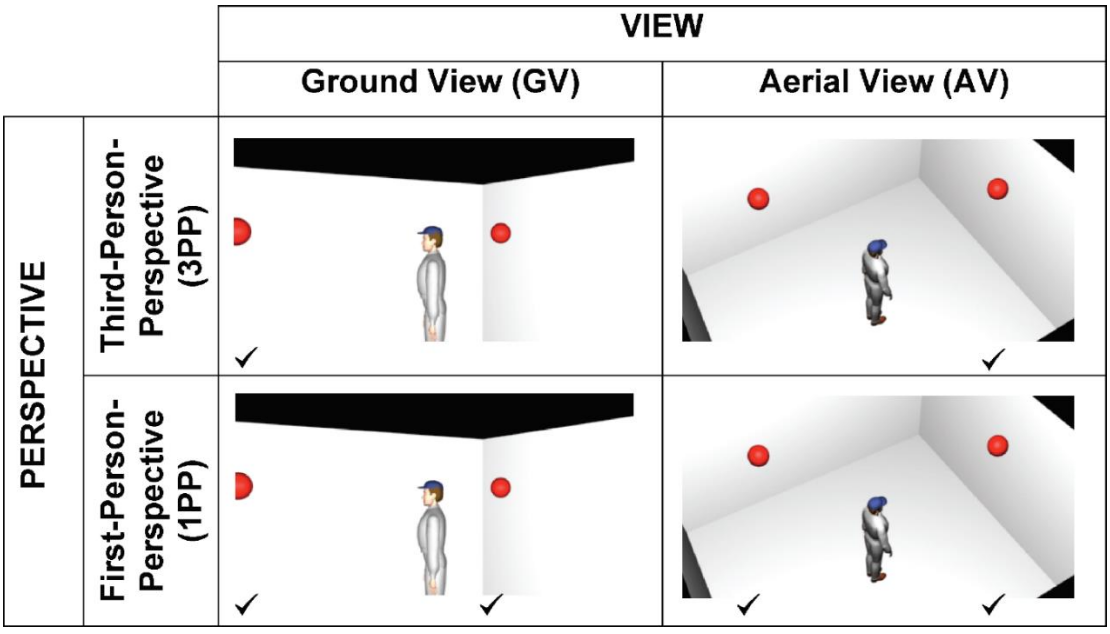
> Self-consciousness includes the consciousness of one's own mental states, such as perceptions, attitudes, opinions, and intentions to act. Representing and integrating such mental states into a common framework, which represents the integrity of our own mind, requires the ability to take a self- or first-person perspective (1PP) among other constitutive features, such as experiences of agency or transtemporal unity (Vogeley, Kurthen, Falkai, & Maier, 1999). 1PP can thus be considered as a basic constituent of a ''minimal self'' (Gallagher, 2000). It enables us to experience the subjective multimodal experiential space centered on our own body. This transient relationship between oneself and objects in the world is the key component of perceptual processes and thus the underlying basis of every cognitive process dealing with the content of these perceptions (Vogeley & Fink, 2003).[136]

Thus, we are presented with a partial definition of self-consciousness, as being aware of one's own mental states. Among its constituents there is the ability of taking the first-person perspective, which is deemed essential for the integrity of our own mind when it comes to the representing and integrating of our own mental states. As a basic constituent of a 'minimal self' it enables us to experience the world around us as standing in some sort of relation to our selves.[137] It is clear that the authors uphold that investigating the relation between perception and self-consciousness is fundamental for understanding self-consciousness, and rightfully so, for what is a self if it cannot be placed within an outer world with other objects? It would be the only thing which exists. *Cogito ergo sum*, and nothing else. I especially like the last sentence of the quote, for it entices further thinking about this temporal relationship between self and the world; is it really the key component of perceptual processes? It probably is not when the functionality of perception is regarded, but it would be if we look at one of the most fundamental values of perception to us; through it we gain a greater understanding of ourselves and the world around us, i.e. it gives meaning, and it could therefore indeed be the case that it is an

[135] K. Vogeley et al., "Neural Correlates of First-Person Perspective as One Constituent of Human Self-Consciousness," *Journal of Cognitive Neuroscience* 16:5 (2004): 817-827.
[136] Ibid., 817.
[137] I will return to Gallagher's concept of a "minimal self" in the final section of this chapter.

underlying basis in regard to cognitive processes dealing with the content of, or resulting from, perceptions. However, I cannot delve into the problems of perception in this essay, for it does not lie within its scope.

With this philosophical view of self-consciousness in mind, Vogeley et al. hope to unearth the neural mechanisms which are associated with the ability to differentiate between first-person and third-person perception, i.e. the processing of perception from oneself and from the viewpoint of another person. Eleven test subjects were placed within an MRI-scanner in which they had to perform a visuospatial task: they had to assess the number of red balls as seen from their own perspective and from an avatar's perspective from a ground view (GV) and from an aerial view (AV) (as shown below), in order to establish whether the taking of 1PP or 3PP results in disparate neurological activity. Through a handheld device, the participants could register the quantity of objects perceived by themselves or by the avatar (the man with the cap). By doing so, the mean reaction times and the percentage of correct answers of the participants could be analysed. By using an MRI-scanner, the authors investigated "whether (a) 3PP and 1PP rely on the same or different brain regions and (b) whether the neural mechanisms underlying 3PP and 1PP are modified by the view onto the scene."[138]



Two-factorial design in Vogeley et al.[139]

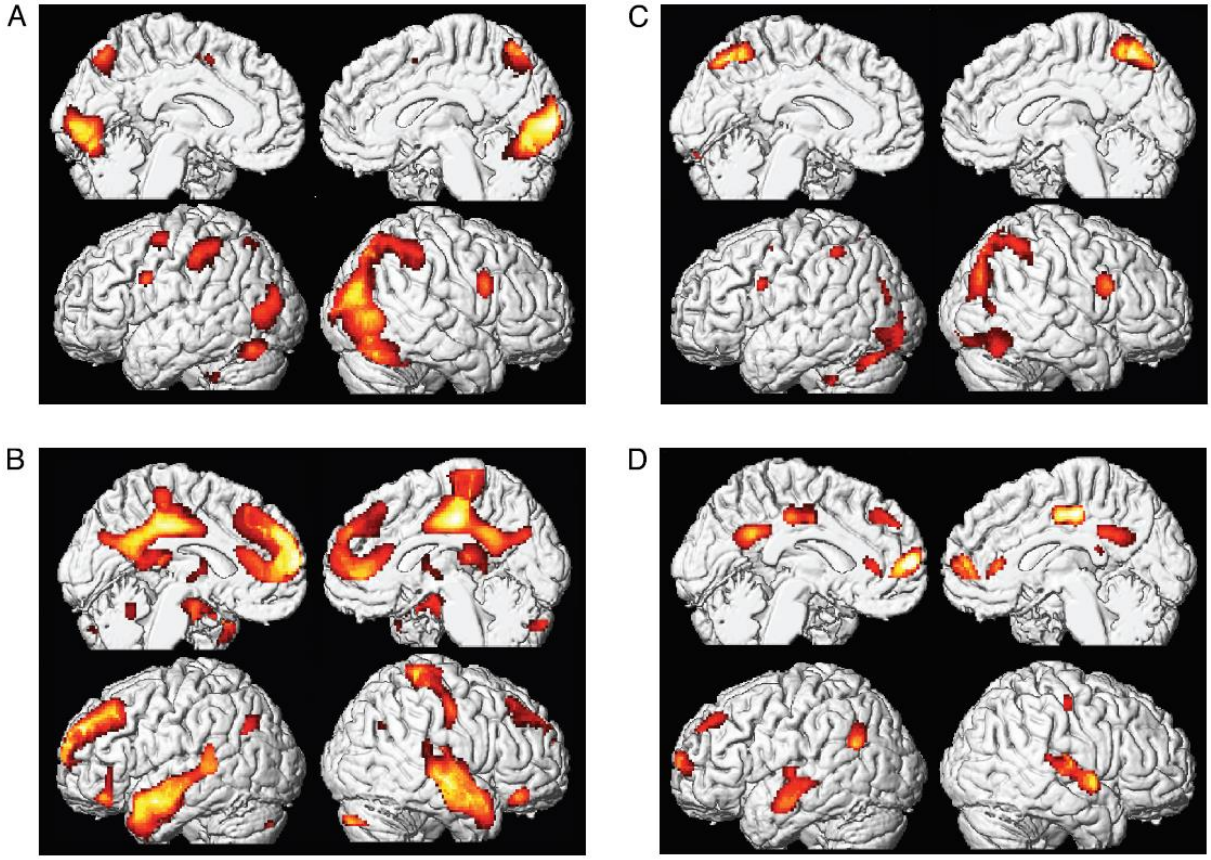
[138] Vogeley et al., "Neural Correlates of First-Person Perspective," 818.
[139] Ibid., 818.

When asked to assess the number of balls from a third-person perspective, participants needed significantly more time and made significantly more errors compared to 1PP. With respect to the time needed, there were no significant differences with respect to the viewpoint (GV and AV), whereas more errors were made during AV conditions. With the help of fMRI, the main effects in neural activity were measured, and the specific brain areas in which activity increase or decrease was found could be determined quite specifically, answering (a).[140] Vogeley et al. found that there is a substantial overlap of 3PP with the areas activated by all four conditions, which suggests an augmentation of neural activity in these brain areas, rather than the activation of additional areas. Neural activity during 1PP was found in different areas, which in turn shows a substantial overlap with deactivations of all four conditions, as can be seen in the image below. In regard to (b), differential neural activity during 'ground view' (GV > AV) could be observed in the right occipital pole, whereas there was no significant change in neural activity associated with the main effect of 'aerial view' (AV > GV).[141]

In the discussion section, the authors state that both viewpoints require egocentric cognitive processes, highlighted in the image below (A), which make use of joint neural mechanisms, e.g. areas involved in general processes of perspective taking and in the storing of spatial information. According to Vogeley et al., the data clearly demonstrate differential brain activations between taking 1PP and 3PP, with a recruitment of additional areas during 1PP, specifically mesial cortical and superior temporal cortical sites. The overlap with the deactivation pattern of all four conditions is explained as follows: these deactivated regions have been previously shown to exhibit task-related decreases in comparison to unspecific baseline tasks, meaning that they might represent a 'default mode of brain function' which includes first-person perspective taking as something that we always do, and which is apparently less active when a specific task is being performed, like taking a third-person perspective, which in this case involves the translocating of the egocentric point of reference onto the avatar's position in the visualisation. Building upon the work of others (Greicius, Raichle, and others), Vogeley et al. argue that the 1PP condition could be taken as a 'state of self' on the phenomenal level, as it appears as a 'default

[140] For example, activations common to all conditions were perceived in the occipito-parietal and frontal regions bilaterally, and in the left cerebellum
[141] Vogeley et al., "Neural Correlates of First-Person Perspective," 818-819

brain state' on the neural level. "As our study focusses on a simple visuospatial task, the theoretical framework of the concept of a "core self" (Damasio, 1999) as a constituent of human self-consciousness also seems to apply for self-related visuospatial processes."[142] Once again we come across Damasio, who seems to have provided an important theoretical and philosophical framework in *The Feeling of What Happens*, which is referred to by quite a few neuroscientists. This is interesting, for it once again shows that the scientific image of self-consciousness, which neuroscientists are attempting to build up step by step with intriguing research like this, cannot do without a manifest image of self-consciousness supporting it.



Neural correlates of perspective taking in Vogeley et al. (A) and (B) respectively represent activations and deactivations common to all four conditions. (C) and (D) respectively represent the main effects of 3PP and 1PP.[143]

[142] Vogeley et al., "Neural Correlates of First-Person Perspective," 823.
[143] Ibid., 822.

In their conclusion, Vogeley et al. state that the results of their study reveal neural correlates of first-person perspective as a particular constituent of human self-consciousness. Taking this perspective appears to be a key component of human self-consciousness, the authors believe, together with at least the experience of transtemporal unity and the experience of agency.[144] This conclusion is telling, for here we find a clear example of a postulational theory, i.e. the neural correlates of first-person perspective, which can be placed in the scientific image of man from a Sellarsian perspective, but which at the same time rests upon philosophical and thus manifest assumptions. For to argue that postulational neurological entities are a particular constituent of human self-consciousness, still does not truly tell us what this self-consciousness is, merely that certain areas of the brain are involved in its constitution, which of course is no small feat, and which might possibly bring us closer to a purely scientific understanding of what self-consciousness is. However, it still seems to be a strange idea that this could ever be achieved without any form of philosophical reflection or introspection. The eliminative materialist might want to work towards a future in which manifest ideas about self-consciousness are dramatically reduced, but as this case study one again shows, the concept of self-consciousness is inherently manifest. In the next two sections, we will move away from the direct experiment, and focus on literature studies based on experiments and clinical studies to further establish why the eliminative materialist will likely never achieve her goal.

VI. SOCIAL COGNITVE NEUROSCIENCE: SELF AND OTHER

In their article, "Shared representations between self and other: a social cognitive neuroscience view", Jean Decety and Jessica Sommerville argue that self-awareness and agency are integral components for navigating a common representation network between self and other.[145] They have collected evidence from developmental science, social psychology, and neuroscience in support of such a common representation network, which works at both the computational and neural levels.[146] I will focus on the neuroscientific evidence, which includes evidence from clinical neuropsychology.

---

[144] Vogeley et al., "Neural Correlates of First-Person Perspective," 819 and 821-824.
[145] Jean Decety and Jessica A. Sommerville, "Shared representations between self and other: a social cognitive neuroscience view," *TRENDS in Cognitive Science* 7:12 (2003): 527-533.
[146] Interestingly, the authors make use of terms like 'coding' and 'computational processing' when discussing neural mechanisms and processes.

Decety and Sommerville start from the assumption that social transactions are a key aspect of human survival. These transactions involve the ability to distinguish yourself from others, and the ability to identify with others. From this assumption the authors jump to their main argument.

> In this paper we argue that the self is a multi-dimensional construct that relies on a distributed neural network that encompasses shared self–other representations. This network is predominantly right-hemisphere based and includes prefrontal, posterior temporal and inferior parietal areas. Rather than considering this network as a single module, we view it as a collection of interconnected regions that are essential for the subjective experience of a 'self'.[147]

To call the self a 'multi-dimensional construct' provides food for thought, is it implies that it is an artificial entity, something that is manmade as it were. One could argue this, for the picture of the self that has been established in this paper so far, by looking at what neuroscientists can tell us about it, could indeed be called a construct; an abstract entity based upon postulational theories and manifest ideas. The authors state that the self *relies* on a network. Although they believe that they can determine the location of this network quite specifically, they do not go as far as to say where the self is located, only that it is an essential collection of regions which enables our personal experience of a self. Decety and Sommerville state that their view of the self relies heavily on the concept of shared representations between self and others, which in turn rests on the idea that the self cannot be understood by an individual independently of a conceptualisation of others, placing the authors on the externalist side of the internalist-externalist debate, although it depends on the more general category of social interaction, instead of language. They believe that "one potential role that cognitive neuroscience can play in the study of the self is to help conceptually define the different dimensions, aspects and characteristics of the self and to help address the potential separability or relatedness of each component part of selfprocessing."[148] This sounds like a very philosophical way of approaching the study of the self, but it stands at odds with the main point of the article, which is the suggestion of a shared neural network in a certain area of the brain; the role seems manifest and the methodology seems scientific. I will return to this issue in the fourth chapter. Below, I have reproduced an image in which the authors have

[147] Decety and Sommerville, "Shared representations between self and other," 527.
[148] Ibid., 528 (Box 1).

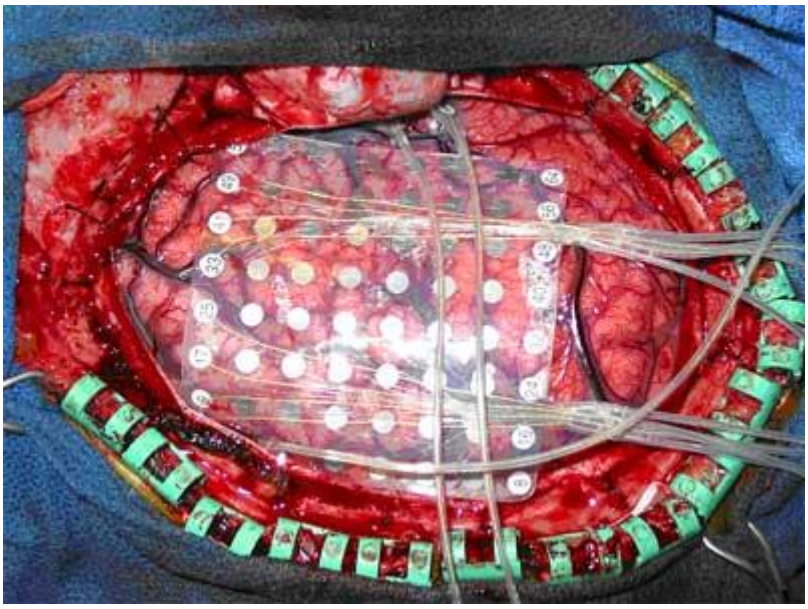
conceptualised and delineated the self into a multi-dimensional construct. As we can see, it is conceptualised as something which encompasses conscious manifest mental activities. The question therefore arises whether this is a delineation of consciousness, rather than of self-consciousness.

Dimension	Aspect	Characteristics
Levels of awareness	Aware	Appears early, primarily perceptual and embodied
	Self-aware	Appears later, based on mental representations
Type of awareness	Agency	Awareness of volition and ownership
	Distinctiveness	Awareness of uniqueness
	Personal continuity	Awareness of continuity through time
	Reflection	Awareness of awareness
Contents of awareness	Physical	Physical features
	Active	Action capabilities
	Psychological	Traits and values
	Social/relational/collective	Social roles and membership, reputation, relationship to others

"Conceptualizing the self" table in Decety and Sommerville.[149]

Decety and Sommerville argue that clinical cases show that the right hemisphere is important for representing the self, as damage to it may cause deficits in various aspects of 'self-processing', like face recognition or the feeling of one's own body. They write about people who could not recognize their own faces, or could not recognize certain limbs as belonging to them. It might also cause deficits in autobiographical memory and self-evaluation. Some people with damage to the right frontal lobe suffer from confabulation; they produce and possess fabricated and incorrect memories about themselves. Other research has shown, which is in line with Vogeley et al.'s article, that there is selective activation of specific brain areas when subjects viewed their own faces among other faces, and also during the presentation of personal autobiographical memories versus impersonal statements, all suggesting a right-hemisphere based network for representing the self. Interestingly, when researchers applied direct electrical stimulation to the right inferior parietal lobe in neurological patients (which involves a physically invasive procedure, as can be seen in the image below), an out-of body experience was induced, which is described as "the experience of dissociation of self from the body".[150]

[149] Decety and Sommerville, "Shared representations between self and other," 528 (Box 1).
[150] Ibid., 529-531.



Electrodes placed over the surface of the exposed right frontal, temporal, and parietal lobes.[151]

Decety and Sommerville underline that the right hemisphere plays a predominant role in the way that the self is connected to the other, with a particularly important role for the inferior parietal cortex for distinguishing self from other, and with an essential role for the prefrontal cortex in coordinating and contrasting cognitive representations of self and other. Furthermore, there seems to be a partial overlap between self-processing and the processing of others; the ability to represent our own thoughts and those of others are closely tied together and may have similar origins within the brain, even though, as we have seen, there is selective activation of specific brain areas when test subjects are asked to perform specific tasks. Questions for future research in regard to neuroscience that are posed by the authors are: "In which specific ways can neuroscience and clinical neuropsychology help tackle different levels of the self[?] Can cognitive neuroscience help to generate a more parsimonious description of the self? What are the limitations of these methodologies?"[152] The first question will be a tough nut to crack, as the different levels of the self seems to encompass consciousness in its entirety as the authors describe it. I shall try to provide an answer to the second question in the fourth chapter of this thesis, as we must still look into more philosophical ways of describing self-consciousness. I hope to have been answering

[151] Arthur J. DiPatri, Jr. and Tord D. Alden, "Surgical Treatment of Epilepsy in Children," from *The Child's Doctor*, Spring 2005, accessed January 26, 2017, https://www2.luriechildrens.org/ce/online/article.aspx?articleID=103.
[152] Decety and Sommerville, "Shared representations between self and other," 532.

the third question to some extent throughout this chapter, in the sense that the limitation of the analysed methodologies is that a philosophical framework is still required for explain what self-consciousness is. In the following and last section of this chapter I will analyse an article which might help with bridging this chapter and the next.

VII. THE PHILOSOPHICAL SELF IN THE COGNITIVE SCIENCES

In "Philosophical conceptions of the self: implications for cognitive science", Shaun Gallagher proposes that several recently developed philosophical approaches to the self might be beneficial to the exchange of ideas between philosophy of mind and the other cognitive sciences.[153] The article is a concise literature review, and it is therefore that Gallagher has decided two focus on two groups of approaches that are focussed on aspects of the self, namely the 'minimal' self and the 'narrative' self.[154] I will focus on the concept of the minimal self, because it is in line with the two philosophical conceptions discussed in the next chapter. Gallagher believes that the approaches that he discusses "promise the best exchange of ideas between philosophy of mind and the other cognitive sciences."[155] Clearly, this is a man who is in favour of bridging the gap between philosophy and science; not many would call philosophy of mind a science, at least not in the common sense of the English word, nor would Sellars regard it as such.

The minimal self is the most basic kind of self imaginable, devoid of any inessential features; the immediate experience of the self in the present, thus only accessible to immediate self-consciousness without the need of any perception of the world. It is further defined as "phenomenologically, […] a consciousness of oneself as an immediate subject of experience, unextended in time."[156] Gallagher approaches the notion of a minimal self with the 'immunity principle', which is described as "immunity to error through misidentification relative to the first-person pronoun."[157] This entails that when someone makes use of the first person pronoun when she refers to herself, there is no possibility that

[153] Shaun Gallagher, "Philosophical conceptions of the self: implications for cognitive science," *Trends in Cognitive Sciences* 4:1 (2000): 14-21.
[154] In the concluding remarks Gallagher refers to Damasio's *The Feeling of What Happens*, when he discusses the interrelations between the minimal ('core') self and the narrative ('autobiographical') self. Gallagher, "Philosophical conceptions of the self," 20.
[155] Gallagher, "Philosophical conceptions of the self," 14.
[156] Ibid., 15.
[157] Ibid., 15.

she refers to another person, only to herself, as the immediate self is referred to as "I", as a subject of immediate and non-observational experience which does not involve a perceptual or reflective act of consciousness.[158] Interestingly, research on schizophrenia by J. Campbell and others has shown that there are counterexamples that seem to form exceptions to this principle. Schizophrenic patients who suffer from 'thought insertion' seem convinced that some thoughts do not belong to themselves, but that they are the thoughts of others inserted in their minds, whilst for 'normal' people there can be almost be no doubt that they are the ones who are having their own thoughts, and not the thoughts of others. Thus, the principle is violated in the sense that whereas any other person would refer to herself when speaking about her thoughts, these patients do not; they make an error through misidentification relative to the first-person pronoun. Gallagher justly writes that it may be argued whether Campbell is correct in his claim that this is a counterexample, for the principle is violated in a rather peculiar way; these people do not refer to other people's actual thoughts as belonging to themselves, which would be a true violation of the principle. Immunity to such an error could only be breached if a person could truly read someone else's mind, and then also truly believe that these thoughts are her own. Nevertheless, Gallagher believes that the implications of this analysis are quite productive, for Campbell's argument implies "that a scientific explanation of schizophrenic phenomena such as thought insertion might also count as a scientific explanation of how the immunity principle works."[159] Is this a counterexample to the primacy of the scientific image, as a scientific image of (schizophrenic) man-in-the-world is used to better understand a manifest image of man-in-the-world? Perhaps not, as Gallagher, arguing that the identification of failing mechanisms at the cognitive or neurological level could indicate the mechanism responsible for the normal immunity to error, in the same paragraph states that "[t]his insight moves us from the conceptual and often abstract arguments of philosophy to the more empirical inquiries of neuropsychology and neurophysiology."[160] It is interesting that it is quite common in the neurosciences to bring to light certain neurological mechanisms by studying defunct ones.[161] In the fourth chapter I

[158] I have here summarised Gallagher's discussion of L. Wittgenstein's and S. Shoemaker's ideas on the first-person pronoun.
[159] Gallagher, "Philosophical conceptions of the self," 16.
[160] Ibid., 16.
[161] Ibid., 15-16.

will discuss how this differs from most philosophy, in which 'normal' minds are usually the object of study.[162]

Gallagher also discusses robotics and artificial intelligence and the minimal self, and concepts of the self by famous philosophers like Galen Strawson, David Hume, Daniel Dennett, and Paul Ricoeur. Because this thesis is not concerned with AI, and because we have yet to delve into the philosophy of mind on self-consciousness proper, I will not discuss these aspects of the review here. I do, however, wish to present Gallagher's closing words, as they touch upon the thesis which I have set out in the introduction of this essay.

> In this review, I have tried to show that philosophical ideas about the self can be aligned with, and can inform, current ideas in cognitive science. I also believe that philosophers can learn about the nature of the self from psychologists, neuroscientists and other cognitive scientists. Thus, collaborative efforts between philosophers and scientists promise to open up more subtle and sophisticated avenues of research, which will define more fully the concept of the self.[163]

In this chapter, I have tried to show that neuroscientific ideas about the self are still inherently aligned with, and reliant on, current and older ideas in philosophy. So, to return to a question that I posed in the first chapter; manifest man survives in the synoptic view of the neuroscientific image of man to the extent that philosophical conceptions of what the self really is, are still required for investigating the nature of the self and its material aspects.[164] I also believe that philosophers can learn about the nature of the self from neuroscientists, especially when the self is regarded as something which is material and as something which can be located within certain regions of the brain. In the fourth chapter, I will discuss whether collaborative efforts between philosophers and scientists can open up more subtle and sophisticated avenues of research, whether they can inform each other, and whether we can define more fully the concept of self-consciousness if we conjoin the manifest and scientific images of self-in-the-world into a stereoscopic view by returning to many of the examples presented in these case studies.

Certain manifest aspects of self-consciousness are necessary conditions for understanding the

[162] See, in chapter four of this essay, "V. The Abnormal and the Normal – the Human and the Primate".
[163] Gallagher, "Philosophical conceptions of the self," 20.
[164] The question was: 'To what extent does manifest man survive in the synoptic view of the neuroscientific image of man?'

possibility of human cognition in general, and these ideas can function as philosophical principles upon which a neuroscientific and stereoscopic image of self-consciousness can be built. Our common-sense manifest ideas of self-consciousness certainly have a home in the neuroscientific image of self-in-the-world, and it has become clear that it would be problematic for the eliminative materialist to argue for the primacy of the scientific image and the undoing of the manifest image with regard to the concept of self-consciousness. However, we have not yet established why certain manifest aspects of self-consciousness are necessary conditions for understanding the possibility of human cognition in general, and this is something which I hope to achieve in the next chapter by discussing two intricate accounts on self-consciousness.

Chapter Three
Philosophical Conceptions of Self-consciousness

Although bookcases full have been written about self-consciousness, especially during the last two centuries, I have selected two theories on self-consciousness. The first theory is by Sebastian Rödl as presented in his book *Self-Consciousness*, as it is a recent book with a properly expounded thesis, and, as I will show in the fourth chapter, it is a suitable candidate for providing the neurosciences with a manifest philosophical foundation of self-consciousness.[165] Furthermore, "the book can be read as an attempt to recover and rejuvenate the achievement[s] of the German Idealist tradition."[166] I am in favour of such a rejuvenation, especially with regard to Immanuel Kant's ideas, which will be extensively discussed throughout this chapter, as it is in Rödl's work, in order to shed a different light on Rödl's theory, with the underlying intention to provide several possibilities for working towards a stereoscopic view in the next chapter. I shall mostly focus on Kant's ideas on the self and self-consciousness as presented in the *Critique of Pure Reason*, and it is the second theory of self-consciousness proper that I will discuss in this chapter.[167]

Rödl's linguistic, or logical, approach to self-consciousness is based on the rational, in the sense that he assumes the existence of *a priori* mental principles and functions, and on the empirical, in the sense that he involves both non-receptive, spontaneous experience, and receptive experience, whereas Kant's approach could be called more rational, as he tries to establish what can and cannot be said about self-consciousness *a priori*, also with regard to pre-experiential mental principles and functions.[168] As we have seen in the previous chapter, neuroscientists approach the investigation of self-consciousness empirically, often grounded on Damasio's empirical philosophical system of consciousness.[169] It is therefore that I will present Kant's ideas beside Rödl's, in order to establish two philosophical

[165] Sebastian Rödl, *Self-Consciousness* (Cambridge, MA: Harvard U.P., 2007).
[166] Ibid., dust jacket.
[167] Immanuel Kant, *Critique of Pure Reason*, trans. Paul Guyer and Allen W. Wood (Cambridge: Cambridge U.P., 2000).
[168] Notwithstanding that Kant discusses the role of experience extensively, e.g. in demarcating rational from experimental psychology of the soul. Rödl rejects transcendental idealism and 'all hitherto existing materialism flawed by its empiricism' in favour of a 'true materialism'.
[169] It is empirical in the sense that Damasio mostly approaches self-consciousness with knowledge acquired through perception, whereas Kant, in contrast, mostly approaches self-consciousness from within thinking itself.

conceptions of self-consciousness that are both rational and empirical, enabling us to offer suggestions for possible syntheses with neuroscientific conceptions of self-consciousness in the next chapter. I believe that a thorough analysis of these two philosophical conceptions might prove useful for bridging the gap between rational, manifest philosophy, and neuroscientific research, which is more empirical by nature.[170] In this chapter I will place less emphasis on the ways in which philosophical conceptions of self-consciousness might belong to a scientific or manifest image; it will become clear that the conceptions discussed belong to the latter category in Sellars' framework, for there is little or no scientific theory postulation to be encountered. This primacy of the manifest image is problematic with regard to the attainment of a stereoscopic image of self-in-the-world, and it is an issue that we will turn to in the next chapter.

I. KNOWLEDGE OF ONESELF, OR THE ORIGINAL APPERCEPTION

> Self-consciousness is the nature of a subject that manifests itself in her thinking thoughts whose linguistic expression requires the use of the first person pronoun, "I". […] An inquiry into self-consciousness, then, is an inquiry into a form of knowledge, which is knowledge of oneself as oneself.[171]

In *Self-Consciousness*, Sebastian Rödl inquires into self-consciousness thought as reference to something real, i.e. thinking of oneself as an object, and he tries to gain a better understanding of this form of knowledge (knowledge of oneself as oneself) through corresponding forms of predication, which are forms of knowledge, as referring to an object and predicating a concept of it bear a certain unity.[172] The understanding of forms of reference through corresponding forms of predication, and the disproval of other claims, is Rödl's line of approach in his book.[173] He argues that belief and action,

[170] I believe that neither the rational nor the empirical should hold primacy, but that a combination of the two deserves primacy. Also, I must note that Rödl himself seems quite sceptical about the role that neuroscience might have to play when it comes to investigating (self-)consciousness. When he discusses the study of happenings in nerves and the brain to understand animal consciousness, he writes that "[i]n truth, the latter study is barred from so much as making contact with its alleged subject: the inner of the animal." It is possible that he would say the same with regard to the inner of the human being. Sebastian Rödl, "Education and Autonomy," *Journal of Philosophy of Education* 50:1 (2016): 90.
[171] Rödl, *Self-Consciousness*, vii.
[172] Ibid., vii.
[173] In this chapter I will mostly discuss Rödl's conclusions of his complex and analytic arguments and refer to the relevant page numbers to enable the reader to delve into them.

distinguishable as two kinds of thinking, are such to be known by their subject in a first person way, and it is therefore that a theory of self-consciousness is also a theory of action, belief, and knowledge. Rödl presents this idea as a central thought of the German Idealist tradition, i.e. "that the philosophical study of action and knowledge must be pursued as part of an inquiry of self-consciousness."[174] This central and arguably essential thought is under-represented in contemporary philosophy, at least, this is what Rödl sets out to prove in his book. I hope I can contribute to the rejuvenation of a small part of that tradition. Namely, its starting point: transcendental idealism.[175]

Rödl's principle claim is that first person knowledge of action and belief is *not* receptive; first person knowledge of acts of thought is spontaneous, which entails that this kind of knowledge is identical with its object: "my knowing first personally that I am doing such-and-such is the same reality as my doing it, and my knowing first personally that I believe that such-and-such is the case is the same reality as my believing it."[176] This kind of knowledge springs from thinking itself. We are provided with reasons to believe and act conforming to a rational order that is inherent in all of us. Explaining an act or belief in terms of conformity with a rational order, is a form of explanation which Rödl calls 'the source of self-consciousness'. "It is in virtue of the unity of explaining why one is doing something and showing it to be good, and of explaining why one believes something and revealing it to be true, that action and belief are known in a first person way. This is the nexus of self-consciousness and reason."[177] An important aspect of Rödl's theory is that he adheres to the position of 'true materialism'. His philosophy is truly materialist in the sense that it represents spontaneity, i.e. self-consciousness, as a material reality; self-consciousness is solely material, and this material is self-conscious.[178] This is Rödl's thesis in a nutshell. In order to make more sense of it, we shall take a closer look at the main body of the text in the following sections. But before we start with the analysis proper, I will shortly introduce another theory on self-consciousness. It is the theory by Immanuel Kant (1724-1804), the

[174] Rödl, *Self-Consciousness*, viii.
[175] A key thought in Kant's transcendental idealism is that there is a difference between phenomena, objects of human cognition as appearances, and noumena, objects as things in themselves (*Dinge an sich*), which are not objects of human cognition. For an insightful introduction to Kant's transcendental idealism see: http://www.iep.utm.edu/kantmeta/#H4.
[176] Rödl, *Self-Consciousness*, ix.
[177] Ibid., ix.
[178] Kant describes the spontaneity of cognition as the faculty for bringing forth representations itself. See page 79 of this essay.

philosopher who started the German Idealist tradition, and whose works have proven to be relevant in philosophy of science to this very day.

> The **I think** must **be able** to accompany all my representations; for otherwise something would be represented in me that could not be thought at all, which is as much as to say that the representation would either be impossible or else at least would be nothing for me. That representation that can be given prior to all thinking is called **intuition**. Thus all manifold of intuition has a necessary relation to the **I think** in the same subject in which this manifold is to be encountered. But this representation is an act of **spontaneity**, i.e., it cannot be regarded as belonging to sensibility. I call it the **pure apperception**, in order to distinguish it from the **empirical** one, or also the **original apperception**, since it is that self-consciousness which, because it produces the representation **I think**, which must be able to accompany all others and which in all consciousness is one and the same, cannot be accompanied by any further representation. I also call its unity the **transcendental** unity of self-consciousness in order to designate the possibility of *a priori* cognition from it.[179]

I have selected this passage as it is essential for understanding what Kant believes to be the function of self-consciousness. Whereas Rödl describes self-consciousness as a form of knowledge, i.e. knowledge of oneself as oneself, Kant describes it as the original apperception, or the mental comprehension which a person has of oneself, that accompanies all representations, which can be representations of objects in the world, but also representations of *a priori* intuitions like space and time. This entails that every thought is accompanied by self-consciousness, always connecting the thought to the subject thinking it, which is essential for the experiencing of reality as it grounds the link between the subject and all objects of possible experience. "The effect of an object on the capacity for representation, insofar as we are affected by it, is **sensation**."[180] Our understanding processes the raw material of sensation, which is acquired through sensibility, as experience, which in our thinking becomes a representation that is always accompanied by the "I think". Self-consciousness could be regarded as knowledge, which stems from spontaneity, but for Kant it seems to be something different, as it functions as an *a priori* ground for gaining knowledge in any form. It is possible that Rödl's use of

[179] Kant, *Critique of Pure Reason*, 246-247 [B 131-132].
[180] Ibid., 155 [A 19-20/B 34].

the term ‘knowledge’ to understand self-consciousness might provide more clarity, or better usability, as Kant’s conception of self-consciousness is somewhat ambiguous; it seems to be something which is always there when we are conscious, but we are unable to truly grasp what the self is as a thing in itself from within our own thinking.[181] We might come closer to understanding what the self is as a physical thing or process, however, with the aid of neuroscience, i.e. through experience, from without our own thinking. Throughout this chapter, I will set Rödl’s and Kant’s views side by side in order to establish which aspects of both works lend themselves best for a fruitful interaction with the neurosciences.

II. FIRST PERSON THOUGHT, ACTION, AND THE “I”

Rödl inquires into self-consciousness through its linguistic expression, and he poses the question on how one refers with “I”. Recognizing that “I” can never refer to an object acquired through the outer senses, even if the object is oneself, entails that in every relation a subject can have with an object, the “I” always refers to the self, which is spontaneous ‘knowledge from the inside’. And because we know things as ‘other’ than “I” through sense perception, unmediated first person knowledge cannot be perceptual.[182] What then, exactly is this unmediated first person knowledge, or ‘knowledge from the inside’? “First person knowledge is knowledge one has not by perceiving but by being its object.”[183] This is still rather vague, but it becomes clearer in light of the primacy of thought. Sensation is excluded as the bedrock of self-consciousness as looking into the nexus that a subject bears to her sensations presupposes an understanding of the subjectivity of the subject, thus thought about thought, rather than thoughts about sensations, and this is what Rödl bestows primacy upon. He discerns two kinds of thought: thought about practical thought, i.e. thought about action, or acts of the will, and thought about theoretical thought, i.e. thought about belief, or acts of the intellect. Reasoning about what to do and what to believe, which is thought about thought, gives rise to first person knowledge of action and belief; it is knowledge from reflection, the form of which Rödl designates as self-consciousness. Actions and

[181] Kant seems mostly interested in arguing what we can say about the self by demarcating what we cannot. He does not provide a clear-cut definition of self-consciousness in its entirety, because there is much that we cannot know about it through experience.
[182] Rödl, *Self-Consciousness*, 1 and 7-10.
[183] Ibid., 10.

beliefs are acts of spontaneity, and hence they are known spontaneously.[184] "They are the kind of thing that *is* its subject's knowledge of it. This explains why it is, and what it means, that the "I do" must be able to accompany all my actions and the "I believe" all my beliefs."[185] The last sentence is concurrent with the first sentence of the quote by Kant presented on page 66, as such thoughts are representations.

In "The Paralogisms of pure reason", Kant does not make a distinction between two kinds of thought about thought, but rather discusses the concept "I think", which he believes is a transcendental concept, i.e. a concept the cognition of which is occupied with our *a priori* concept of this object, revolving around the question what it could be before possible experience.[186] This concept is the vehicle of all concepts according to Kant, as it serves to enact all thinking as belonging to consciousness, whilst it also serves to distinguish two kinds of objects as representations: "**I**, as thinking, am an object of inner sense, and am called "soul." That which is an object of outer sense is called "body""[187] Making one's own representation, "I think", the object of one's thoughts, is due to the faculty of inner sense, and it might be argued that it is thus *experienced*, but Kant believes that this kind of 'inner perception' is nothing beyond the mere transcendental apperception "I think", which functions as the spontaneous source of all synthesis in thinking, and that is always there as it is required for the cognition of the empirical in general; inner experience in general and its possibility, built on the proposition "I think", thus belongs to the investigation of the possibility of every experience, which is transcendental, and not empirical.[188] This ties in nicely with the following statement by Rödl: "First person knowledge is knowledge one has not by perceiving but by being its object", as "I think" refers to the original apperception that is inherent in all thinking and is inseparable from our consciousness.[189]

We shall now look into Rödl's ideas on the subjectivity of practical thought, or action. The question that Rödl seeks to answer is "how I know that I am doing something when I know that by doing it."[190] Rödl argues that there is a unity between thought and movement; if I move my arm intentionally,

[184] Rödl, *Self-Consciousness*, 10-14.
[185] Ibid., 14.
[186] Kant does demarcate pure reason from practical reason in the *Critique of Pure Reason* and in the *Critique of Practical Reason*, but self-consciousness is primarily discussed within the context of pure reason.
[187] Kant, *Critique of Pure Reason*, 412 [A 342/B 400]
[188] Ibid., 132-133 [A 11-12/B 25], 257 [B 152-153] and 411-412 [A 341-343/B 399-401].
[189] Rödl, *Self-Consciousness*, 10.
[190] Ibid., 17.

then this action is a thought that is simultaneously a movement. An intentional action being a movement is thus a conclusion of reasoning.[191] Now the question that arises here is 'how do I know what to do?' Where does this intentionality spring from? Rödl offers the following view on what he believes is *the order* to which the question refers. The relevant unity here is not the unity of desire, but a unity of 'infinite ends'. Whereas all-things-considered judgments, from the theory of calculation from desire, are used at specific times and can differ per moment, making it an unsuitable candidate for a grounding principle of progress of an action, infinite ends are time-general. [192] An example of an infinite end that Rödl gives is 'health'; it is an end which most of us, at least, are always pursuing and which is at times achieved, when feeling healthy, and which at other times is not, as when feeling ill. The usage of antibiotics when suffering from a bacterial infection is explained by the infinite end of wanting health, for example, as this end contains the whole temporal extension of actions it explains, i.e. the daily intake of medication. Thinking that taking antibiotics is something to do upon understanding that it manifests the infinite end of wanting health, gives proper ground for thinking what to do.[193]

"Infinite ends are the order that defines the question what to do. More precisely, the order of this question is an *objective unity of infinite ends*, which we shall call a *practical life-form*."[194] The totality of infinite ends that an individual may have, exhibit her practical life-form in unity, according to Rödl. A tricky aspect of this is that at certain times infinite ends may conflict, for example as is the case when one must lie for the sake of health, whilst the conflicting infinite end is 'to be honest'. The point here is that as the subject ranks the infinite ends at a specific time, she makes the decision and gives primacy to either one on which she bases this decision, and making such a ranking of infinite ends represents a changeable state, placed in the realm of time-specific desires, thus unfit as an answer to the general question 'what to do'. On the contrary, the objective unity of infinite ends, the practical life-form, defines the question 'what to do' in a time-general manner. The thought that my taking in of

[191] Rödl, *Self-Consciousness*, 19. This is an Aristotelian idea, the origin of which can be traced back to Aristotle's *Nicomachean Ethics*, which was written around 350 BC.
[192] Rödl presents this view after analysing why calculation from desire does not arrive at an action due to its altering temporality, and after exhibiting why Donald Davidson's principle of continence does not define the question what to do with a similar argument. The principle of continence urges us to "perform the action judged best on the basis of all available relevant reasons." Donald Davidson, "How is weakness of the will possible?," in *Essays on Actions and Events* (Oxford: Oxford U.P., 2001 [1980]), 41.
[193] Rödl, *Self-Consciousness*, 34-38.
[194] Ibid., 38.

antibiotics represent an action as exemplifying an infinite end, i.e. health, presupposes a grasp of a unity of infinite ends, believes Rödl, because an action manifests an infinite end if and only if it manifests the practical life-form that includes this end among other ends. Thus, it presupposes that I understand that my wanting health is a similar end as wanting to be honest. The ends gathered in a practical life-form may differ per subject, but in all subjects the unity of ends transpires from any of them; an individual always represents an action as manifesting her practical life-form. Practical life-forms may differ, but people may also share them, which may be established through reflection, thus enabling community of practical thought. The self-conscious subject's thoughts that represent an action as exemplifying an infinite end are ultimately grounded on all her infinite ends, a system of forms, whose unity makes up the subject's practical life-form, the logical form of the system's principle.[195] How then does the answering of the question 'what to do' in thought, result in an action?

"[W]e shall argue that an action *is* its subject's answer to this question; it is the conclusion of her practical reasoning. From this follows […] that there is a special way of knowing one's own actions."[196] Answering the question 'why one performs a certain action intentionally' comes in the form of an 'action explanation', which, argues Rödl, always contains the first person reference of the acting subject.[197] Practical reasoning arrives at an action, and it can be explained by the subject as something which she has done because she thought that she should have done it, whilst simultaneously this action explanation represents her action as resting on thought. Thus, the system of answers, i.e. the system of infinite ends, to the question 'what to do' of practical reasoning, is also the system of answers to the question 'why' of action explanation. In an action explanation, the subject thinks that she does something, because it is 'good' to do so. This functions as the cause of doing something and as the ground of someone's thinking that one should do it. There is thus a unity of action explanation and practical reasoning (which concludes in a movement), according to Rödl, as they are one act of mind.[198] I will try to elucidate the above with an example: I think that I should take my antibiotics as I want to get rid of an infection, thus I perform the required actions. If I would then be asked to explain 'why' I

[195] Rödl, *Self-Consciousness*, 38-44.
[196] Ibid., 44.
[197] Rödl develops this idea under the concept of 'the nexus of practical reasoning and action explanation'.
[198] Rödl, *Self-Consciousness*, 44-48 and 56.

took the medication, I could give a complicated explanation about how antibiotics work and how I understand them to counter a bacterial infection, but ultimately it would be grounded on the infinite end of health, as I believe that wanting health is something which is good. The question is how this is all related to self-consciousness.

> Actions fall under a form of explanation such that she who is acting thinks what she is doing to be done. We will see that this entails that, if a movement can be explained in this way, its subject is able thus to explain it. From this it follows in turn that she expresses action explanations by a first person pronoun. When she does, she states knowledge not from observation, but from practical reasoning, which way of knowing sustains her first person reference. In this manner, self-consciousness, action, and practical reasoning are internally related.[199]

First person knowledge about why one is doing something does not come from the senses, but, as we have seen, it comes from thought. This knowledge comes not from observing what one as an object is doing, but from ascertaining what to do as a subject, that is, from within. There is no possibility of me misidentifying the object that is me when I know that I am doing something by concluding that I should do it; these are unmediated first person thoughts which are one and the same reality, i.e. thought about doing something and doing it. If we consider thinking about doing something as an action concept, as doing something's sense depends on action explanation, then this concept should be regarded as something that can be acquired, according to Rödl. If it is acquired by an individual, then she has the power to know in a way that underwrites first person reference. It is thus described as an ability which functions as a source of power to think and say "I".[200]

If I am to interpret Rödl and Kant, and bring their views together, then I would say that the acquiring of an action concept would be realised with experience through apperception in an individual. Kant believes that apperception is the ground of the possibility of the categories, i.e. the twelve pure concepts of the understanding applying to all objects of intuition (of quantity, of relation, of modality, and of quality), which represent the synthesis of the manifold of intuition, insofar that manifold has unity in apperception. Thus, that everything that we can perceive, is processed through these categories

[199] Rödl, *Self-Consciousness*, 55.
[200] Ibid., 55-63.

in apperception to cognizable thought, which would include the acquiring of an action concept to the extent that experience is required for such an acquirement. "Self-consciousness in general is therefore the representation of that which is the condition of all unity and yet is itself unconditioned."[201] Thus, the thinking "I", or the soul as Kant calls it, "cognizes the categories, and through them all objects, in the absolute unity of apperception, and hence cognizes them through itself."[202] It is therefore that the "I" cannot be cognized as an object itself, as it is the very part of us that allows us to cognize objects at all.[203] So, whereas Rödl speaks of a power which functions as a source of power to think and say "I", Kant believes that the source of power to think, and not so much say "I", in no way lies in such an acquired ability as the "I think" must be able to accompany all my representations in the transcendental unity of self-consciousness *a priori*; the thinking "I" has no need of any underwriting as it grounds all thoughts for Kant, whereas for Rödl experience seemingly strengthens the purely rational aspects of self-consciousness. Thus, in my interpretation, Rödl affirms pure reason, and brings into play the world of experience as essential for thinking of oneself in his analysis, whereas Kant considers the purely rational aspects of self-consciousness, before bringing into play the world of experience, which is essential for human cognition nonetheless. In section IV, "True Materialism, or the "I" as a Self-Subsisting Being or Substance," I shall return to this issue.

III. BELIEF AND FIRST PERSON THOUGHT, A PARALLEL NEXUS

> [W]e shall expound a parallel nexus of first person thought and belief. Believing something is answering the question what to believe. Therefore I can know that I believe something by determining what to believe. When I know that I believe something in this way, I have unmediated first person knowledge; I do not need to recognize a person in order to tie the belief to myself.[204]

As first person thought and belief has a parallel nexus with first person thought and action according to Rödl, I shall only discuss the most important similarities and differences. As Rödl sought for a high form of temporality for determining the will, which he found in the form of infinite ends and not in

[201] Kant, *Critique of Pure Reason*, 442 [A 401].
[202] Ibid., 442 [A 401-402].
[203] Ibid., 212 [A 79-80/B 105-106] and 442 [A 401-402].
[204] Rödl, *Self-Consciousness*, 65.

'desire', he also seeks a high form of temporality for determining the intellect, the order of theoretical reasoning, which cannot be found in sensation, as an all-things-considered judgement based on sensations represents a changeable state; what I believe today at a certain time based on certain sensations through inference, I do not necessarily need to believe tomorrow, as other sensations might affect my theoretical reasoning through inference then. Believing something, however, is a *temporally unlimited act*, according to Rödl; at the time that I believe something, any limit of its duration is accidental to this believe and it does not have a proper end. If I believe that something is true at a specific time, e.g. that antibiotics help in combating an infection, at that moment I believe this to be true at all times. Here we find a difference with practical reasoning, as it is perfectly logical to think that it is right to do something at a certain moment, and wrong at another.[205]

Answering the question 'what to believe' thus poses no temporal limit, and when I determine what I believe by reference to perception, I represent my believing as an act of a power to gain knowledge. This power is an *infinite ground*, according to Rödl, as it functions as a foundation of beliefs that exhibits my possession of this power. A power of knowledge is comparable to a practical lifeform, as they exhibit a necessary unity in our believing and acting. Whereas the practical life-form can be signified as "*a unity of actions that manifest this life-form and are ordered among themselves as means to ends*", with what is 'good' as the order of practical reasoning, the power of knowledge can be signified as "*a unity of beliefs that manifest this power and are ordered among themselves as ground to consequence*", with what is 'true' as the order of theoretical reasoning.[206] The 'true' and the 'good' are both formal concepts, in the sense that they designate the way in which a proposition is represented in an answer to the question 'what to believe', and to the question 'what to do', respectively. As we have seen, the system of answers, i.e. the system of infinite ends, to the question 'what to do' of practical reasoning, is also the system of answers to the question 'why' of action explanation. Similarly, the system of answers, i.e. the system of infinite grounds, to the question 'what to believe' of theoretical reasoning, is also the system of answers to the question 'why' of belief explanation. The main argument of belief explanation is comparable: "[T]hat which explains why someone believes something in the

[205] Rödl, *Self-Consciousness*, 74-79.
[206] Ibid., 84.

intended way is something from which she concludes that it is something it is right to believe."[207] And lastly, in a belief explanation, the ground on which one thinks it is proper to belief something, is also the *cause* for one's believing it; my believing something to be true and my thought that it is something it is right to believe, argues Rödl, are the same reality; belief and justified knowledge spring from the same source that is our self-consciousness.[208]

The inner nexus of self-consciousness, belief, and theoretical reasoning, is comparable to the nexus of self-consciousness, action, and practical reasoning; giving an explanation on why I believe something, I do not expound knowledge from observing what I believe, but rather from *ascertaining* what to believe. This spontaneous knowledge defines the first person reference of a subject to beliefs, according to Rödl. Now, if I believe something because I have perceived it, then the causality of this explanation contains my representation of this causality, as I believe and thus know it to be true. A power of receptive knowledge thus sustains a certain form of explanation; it is a formally represented order with believes conforming to it. And, a formally represented order is a source of self-consciousness, as it sustains causality of thought. This is the same for practical life-forms. Rödl supports this claim with the argument that there is identity of subject and object in unmediated thought, which satisfies the formula of first person knowledge: I, as subject, give the explanation by being its object, and there is no possibility for a different identity judgment; I cannot misidentify *my* believing with a supposed believing of another object.[209] Rödl once again speaks of 'acquiring' in this context, which usually stand for 'gaining possession of' or 'gaining through experience', but not of a concept, but of a power: "Acquiring the power of theoretical thought and belief is acquiring the concept of belief and with it the power of first person thought."[210] Rödl presumably refers to the acquirement of this ability during human evolutionary development. A parallel can be drawn here with Sellars' coming to being of the manifest image, as this was the moment in which man first became aware of himself as man-in-the-world. The

[207] Rödl, *Self-Consciousness*, 89.
[208] Ibid., 82-92.
[209] Ibid., 95-98 and 106-107. As we have seen in the previous chapter, schizophrenic patients who suffer from 'thought insertion' do seem to misidentify their believing with a supposed believing of another object or subject. We will return to this issue in chapter four of this essay in section V, "The Abnormal and the Normal – the Human and the Primate".
[210] Ibid., 103.

power of theoretical thought of belief seems a requisite for this new level of awareness in which first person thought plays a fundamental role.

Let us now consider a section of Kant's writing which relates to believe and experience, to the "I think" and objects, and to the self as object. Kant explains the limits of the proposition "I think" in regard to objects as follows: all modes of self-consciousness in thinking, which are the different ways of being conscious to oneself of the intuition of oneself as determined in regard to the function of thought as accompanying all categories as their vehicle (simply put, the "I" accompanying all thoughts), are *functions* which do not provide an object in thought, and thus also not present the self as an object to be cognized. Thus, the consciousness of the *determining self* (comparable to what might be called the *ascertaining self*), which is fundamental for all thinking, can never be an object of thought as it is *prior* to all thought, whereas, conversely, the consciousness of the *determinable self* is an object of thought, in the sense that it is an object of inner intuition, acquired in thinking *a posteriori*, an object which is believed to be the self as it is ascertained by the formal self.[211] To draw a, hopefully somewhat clarifying, parallel: it might be argued that the determining self is what Rödl calls spontaneous knowledge, whereas the determinable self is what Damasio would call the autobiographical self. Regarding the relation between the self and other objects, Kant argues that it is an analytic proposition when you say that you would distinguish your own existence, that of a thinking being, from all other things outside this existence, including your own body, as other things are always those things which are distinguished from yourself.

> But I do not thereby know at all whether this consciousness of myself would even be possible without things outside me through which representations are given to me, and thus whether I could exist merely as a thinking being (without being a human being). Thus through the analysis of the consciousness of myself in thinking in general not the least is won in regard to the cognition of myself as object. The logical exposition of thinking in general is falsely held to be a metaphysical determination of the object.[212]

[211] Kant, *Critique of Pure Reason*, 445 [B 406-407].
[212] Ibid., 446-447 [B 409].

How does this take on self-consciousness compare to Rödl's? It follows from this quote that there is *no* identity of subject and object in unmediated thought, at least not *a priori*. Rödl, however, argues that there is identity of object and subject, as it is "I" that believes that such and such is the case, resulting from my experiencing this. Thus, only through the analysis of the consciousness of myself in thinking in general after perception of the outer world, can I have a cognition of myself as object in Rödl's view. The last sentence from the quote by Kant implies that Descartes' *cogito ergo sum* can only refer to the existence of the thinking *subject*, as it cannot function as the metaphysical determination of the existence of myself as an object, which can only be determined with experience of what is outside our existence as solely thinking beings, always without the determining self as part of this object, and perhaps only in a non-metaphysical and manifest manner. This leads us to the question what the self then *is*, or can be, as an object, and it is the question that we shall now turn to.

IV. TRUE MATERIALISM, OR THE "I" AS A SELF-SUBSISTING BEING OR SUBSTANCE

> It is the principle thought of German Idealism that self-consciousness, freedom, and reason are one. In our manner of developing it, this thesis is part of a truly materialist theory of self-consciousness.[213]

The formally represented orders, a power of receptive knowledge and a practical life-form, are key to understanding Rödl's conception of self-consciousness as they both define a concept of justification, epistemic and practical, which shape the rational being; all acts and beliefs are justified by reason, and they always fall under a formally represented order. Me thinking about my acts and beliefs automatically places this thought under this order in unmediated first person thought. It is therefore that a formally represented order is both an order of reason, and an order of self-consciousness.[214] "Reason and self-consciousness are two sides of a coin; the coin is a normative order internal to and represented by the acts that it governs."[215]

When reflecting on this idea, Rödl writes that Kant (and Hegel) see reason as the sole order of self-consciousness, whereas he himself believes that there are two orders, which are both orders of reason and self-consciousness. The order of the will (practical life-form) and the order of the intellect

[213] Rödl, *Self-Consciousness*, 105.
[214] Ibid., 108-109.
[215] Ibid., 109.

(power of receptive knowledge) have not been presented as being the same thing, which entails, Rödl explicitly admits, that he lacks an account of the unity of theoretical and practical reason and of the unity of the subject of action and belief.[216] This is a somewhat troubling gap, a gap on which Rödl comments that he cannot fill it within his treatise. It is clear why Rödl makes the demarcation between these two orders, as believing, which happens in the mind after bodily perception, and acting, which happens in the mind with bodily muscular action, both relate to self-consciousness in their seemingly particular ways, the former pertaining to infinite grounds, and the latter pertaining to infinite ends. But it could be counter-argued that believing and acting, and infinite grounds and infinite ends, are in fact the same things that do not exist without reason, and that the unity of theoretical and practical reason is found in the self-consciousness of the subject. Reason is the absolute prerequisite basis of all thinking and acting, from which the unity of the subject of action and belief automatically follows; there is no gap. We have seen all the parallels between these two forms of thinking. Why does Rödl not make the step of seeing them as synonymous? Perhaps it is because of the intermingling of experience in his account which might obscure what happens in the mind? I do not think that it could be easily disproved that my acting based on an infinite ground because it is good and my believing this ground to be true because I have ascertained it, ultimately stem from the same source: reason. Furthermore, it could be argued that the division of the order of reason is unnecessary as it undermines the unity of consciousness.

In the *Critique of Pure Reason* Kant upholds that the understanding brings the manifold of given representations under the unity of apperception, uniting them in my self-consciousness, as all representations given in the intuitions of space and time belong to me. The principle of the unity of apperception, 'the supreme one in the whole of human cognition', is the necessary synthesis of the manifold given in an intuition, always related to me in self-consciousness.[217]

> The supreme principle of the possibility of all intuition in relation to sensibility was […] that all the manifold of sensibility stand under the formal conditions of space and time. The supreme principle of all intuition in relation to the understanding is that all the manifold of intuition stand under conditions of the original synthetic unity of apperception. All the manifold representations of

[216] Rödl, *Self-Consciousness*, 109-110.
[217] Kant, *Critique of Pure Reason*, 247-248 [B 133-136].

> intuition stand under the first insofar as they are **given** to us, and under the second insofar as they must be capable of being **combined** in one consciousness; for without that nothing could be thought or cognized through them, since the given representations would not have in common the act of apperception, **I think**, and thereby would not be grasped together in a self-consciousness.[218]

Thus, according to Kant, everything I perceive with my senses is in space and time, as formal human-imposed conditions, limiting and shaping my representations of the outer world, and the idea that I can form thoughts about everything that I can perceive in space and time, is something that I am conscious of *a priori* in regard to the manifold of intuition of the representations given to *me*, and only me. These representations form *one*; they are mine and no-one else's. All representations that I can have in life stand under this unity, synthesised in my understanding, forming thought that is always connected to my self. The "I think" forms the act of apperception. The question that is to be answered is whether the unity of the subject of action and belief follows from this. I would say that it does, if actions are indeed based on infinite ends: these ends cannot be anything else but the result of the combining of the manifold representations of intuition, as they are principles based upon the believe of what is true; my wanting health is not only something of which I think that it is good to do, but it is also something which I have ascertained to be true time and time again, as I feel better when I am healthy compared to when I am ill. Action and belief ultimately rest on the principle of the unity of apperception.

Rödl offers us a kind of causality which links acts of the will with acts of the intellect. As we shall see, this brings us closer to closing the gap between theoretical and practical reason. It is the concept of freedom which is characteristic for the causality of thought sustained by a formally represented order, at least, if you take the existence of free will for granted. The idea of freedom "is the idea of a kind of determination; it signifies a kind of causality, intelligibility, and reality: a causality of thought, an intelligibility that passes through an order of reason, and a reality that is self-conscious."[219] Rödl presents Kant's vision on freedom, and he sets forth Kant's claim that being autonomous is the same as being under laws of reason; the autonomy of the will is its own law, as a law that governs thinking and a as a law which springs from, and constitutes, self-consciousness. The nexus of self-

[218] Kant, *Critique of Pure Reason*, 248-249 [B 136-137].
[219] Rödl, *Self-Consciousness*, 112.

consciousness and freedom underlies first person reference as unmediated first person knowledge comes from spontaneity.[220] It is this spontaneity which provides us with free will, as it is from spontaneity that thought springs. The faculty of thinking provides its own causality, with the principle of the unity of apperception providing it with its essence: self-consciousness. In this context, it might be helpful to cite one of the most famous passages of the *Critique of Pure Reason* more extensively, as it tells us what spontaneity is, and as it bridges the gap between rationalism and empiricism.[221]

> If we will call the **receptivity** of our mind to receive representations insofar as it is affected in some way by **sensibility**, then on the contrary the faculty for bringing forth representations itself, or the **spontaneity** of cognition, is the **understanding**. It comes along with our nature that **intuition** can never be other than **sensible**, i.e., that it contains only the way in which we are affected by objects. The faculty for **thinking** of objects of sensible intuition, on the contrary, is the **understanding**. Neither of these properties is to be preferred to the other. Without sensibility no object would be given to us, and without understanding none would be thought. Thoughts without content are empty, intuitions without concepts are blind.[222]

Both the inner working of our minds, the rational, and the outer working of our bodies in the form of sensibility acting on the receptivity of our mind, the empirical, are required for the thinking of possible objects, and neither of these holds primacy as both are required for our subjective apprehension of objective reality, i.e. the reality we experience as a representation of the outer world, which is given to us in the manifold of sensibility. The equation of rationalism with empiricism is fundamental if we are to work towards a stereoscopic view, especially when we try to come to suggestions for bringing together Kantian and neuroscientific ideas on self-consciousness in the synthesis that will take place in the next chapter. I uphold that if we wish to understand self-consciousness to the fullest extent, we can't merely philosophise about it and argue what can and cannot be said purely from thinking about it, as there is a material and thus empirical aspect to self-consciousness as well; our brain as part of our body.[223] We shall now see how Rödl's true materialism relates self-consciousness to matter, and in the

[220] Rödl, *Self-Consciousness*, 110 and 118-120.
[221] It could even be argued that it closes this gap through an equation.
[222] Kant, *Critique of Pure Reason*, 193-194 [A 51/B 75].
[223] I must stress that I express my own philosophical position here; Rödl and Kant might not be inclined to give primacy to such a stereoscopic image of self-consciousness.

next chapter we shall see how we can connect this to neuroscientific materialism.

Rödl upholds that me knowing something or me acting, is equal to exercising my power of belief or intentional action, without the room for a further power of being aware of this, as these powers simultaneously have the function of performing acts of knowing. This kind of first person knowledge is spontaneous as it does not spring from sensory affection, but from the mind itself, which could be described in a Kantian way as springing from the understanding. Rödl believes that from this assumption it would be wrong to conclude that the object of first person knowledge cannot be a material reality, knowable only empirically, as if it is nothing other than thinking. He is inspired by Karl Marx's theses on Feuerbach in rejecting this idea.[224]

> All hitherto existing materialism is flawed by its empiricism: it conceives of material reality exclusively as an object of intuition, or as to be known receptively. Idealism shares this flaw, wherefore it develops spontaneity ("die tätige Seite") in contrast to the material. According to Marx, true materialism reveals spontaneity and its knowledge to be of, and thus to be, a material reality. Our account of self-consciousness aspires to being materialist in this way.[225]

This does not seem to fit the formula "without sensibility no object would be given to us, and without understanding none would be thought", as the object is not given in sensibility but in understanding itself, and we have seen that, according to Kant at least, this is not possible when it comes to the determining self, as the consciousness of the *determining self* can never be an object of thought, as it is *prior* to all thought, whereas, conversely, the consciousness of the *determinable self* is an object of thought, in the sense that it is an object of inner intuition, acquired in thinking receptively *a posteriori*.[226] It must be noted, however, that also in Rödl's theory no particular object would be given to us without sensibility. He rather wishes to convey that sensibility and understanding are part of the same material reality. We must therefore ascertain what form of self Rödl is speaking of when it comes to the object of first person knowledge.

In order to present his own materialist philosophy, Rödl expounds G.E.M. Anscombe's essay

[224] Rödl, *Self-Consciousness*, 121-122.
[225] Ibid., 122.
[226] Kant, *Critique of Pure Reason*, 445 [B 406-407].

"The First Person", as he believes that she is working towards a true materialism in her paper.[227] From it, he deduces that there can be nonreceptive knowledge of a material substance, the comprehension of which comes from an analysis of the form of predication that characterises first person thought. "First person knowledge is knowledge I have of an object by being that object."[228] It does seem that the knowledge Rödl is speaking about concerns the determining self, as he is speaking about *a priori* nonreceptive knowledge, and it simultaneously concerns the determinable self, as first person thought is presented as an object of inner intuition. Rödl upholds that when you have a thought about an object other than yourself, then an object is singled out from a manifold of intuitions, but when you have an "I"-thought there is no manifold as the object of this thought can only be yourself as a material substance.

He also provides us with his own account of how first person thoughts pertain to a material substance, by arguing that when someone thinks a first person thought representing movement, the thought "I am doing A", she applies a material substance concept to herself, and the unity of action explanation and practical reasoning function as the source of this spontaneous knowledge of movement.[229]

> I have spontaneous knowledge of the kind of substance I am, the kind of substance that the concept designates that is contained in my first person thoughts that represent my intentional actions. I know that I fall under this concept not by perceiving a substance that falls under it, but by being a substance that falls under it, or, shorter, by falling under it. An acting subject is a material substance of a kind such that she knows what kind of substance she is by being a substance of that kind.[230]

The last part of this argument seems somewhat circular, an issue to which I will return shortly, but Rödl's claim is clear: we know from spontaneity that we are a material substance through acting. An "I"-thought representing movement, results in the physical movement of a seemingly temporal and spatial substance that cannot be anything but your material self. It seems that this first person knowledge in no way refers solely to the determining self, or solely to the determinable self, but it refers to these

[227] G.E.M. Anscombe, "The First Person," in *Mind & Language: Wolfson College Lectures 1974*, ed. Samuel Guttenplan (Oxford: Clarendon Press, 1975), 45-65.
[228] Rödl, *Self-Consciousness*, 124.
[229] Ibid., 123-131.
[230] Ibid., 131.

forms as part of a larger concept: the human being in its entirety. Thus, from Rödl's view, it would be a mistake to distinguish between two forms of self. Now, how does Rödl's true materialism compare to Kant's transcendental idealism?

In these paragraphs, I will focus particularly on the "Paralogisms of Pure Reason" rewritten for the second edition, in which the proposition "I think" (taken problematically) is discussed. Kant states that "in every judgment, **I** am always the **determining** subject of that relation that constitutes the judgment. However, that the I that I think can always be considered as **subject**, and as something that does not depend on thinking merely as a predicate, must be valid – this is an apodictic and even an **identical proposition**; but it does not signify that I as **object** am for myself a self-**subsisting being** or **substance**."[231] It would thus go too far to believe that you can have any spontaneous knowledge of the kind of substance you are, as "I-thoughts" only provide you with your determining subject. A fundamental difference between the two theories, which must be addressed, is that Rödl discusses the thought that is movement, whereas Kant discusses representations of the understanding, without discussing possible simultaneous movements.

With regard to the seeming circularity in Rödl's reasoning, i.e. an acting subject being a material substance of a kind such that she knows what kind of substance she is by being a substance of that kind, the point that can be brought up is, paraphrasing Kant, that the subject of movement cannot obtain a concept of itself as an object of movement; for in order to think this, it must take its pure self-consciousness, which is just what is to be explained, as its ground.[232] The explanation is thus not given by simply equating the subject with the object. Furthermore, being a substance does not entail that you know what kind of substance you are. For Rödl, the acting subject knows she is an acting object that is a substance through experience and spontaneous knowledge, but for Kant, the acting subject knows she is an acting object that is a substance solely through experience. The fundamental difference is thus that in Rödl's view the subject and object are the same thing apodictically, whereas in Kant's view the subject

[231] Kant, *Critique of Pure Reason*, 445-446 [B 407]. Kant makes an interesting remark directly after this last citation: "The latter goes very far, and hence demands data that are not encountered at all in thinking, and thus (insofar as I consider merely what thinks as such) perhaps demands more than I will ever encounter anywhere (in it)." This data might indeed not be encountered in thinking, but it might be encountered when the brain is studied empirically.
[232] Ibid., 453 [B 422].

and the object are only inferred to be the same thing without any apodictic certainty.[233] It must be stressed that Kant would not reject our existence as an object, as he upholds that the transcendental idealist is an empirical realist, in the sense that a reality is appropriated to matter as it is to appearances, a reality which is immediately perceived.[234] However, I do believe that he would reject true materialism's claim that an "I"-thought on its own is sufficient for grounding the subject as an object with substance. The object that is our body is an object in space of which I can have a representation, but which I do not truly know as a thing in itself. The object that we might believe to be "I", which to Rödl is more than mere belief, represented through inner sense in time, is also an appearance. Kant underwrites that although these two appearances are wholly distinct, they are not thought of as different things, as it is in Rödl's true materialism, who believes them to be truly one and the same. "The **transcendental object** that grounds both outer appearances and inner intuition is neither matter nor a thinking being in itself, but rather an unknown ground of those appearances that supply us with our empirical concepts of the former as well as the latter."[235]

Thus, it seems that if we would adhere to Rödl's position, on the one hand, we would perhaps attribute too much certainty to the physical reality of the concept of self-consciousness, in the sense that the apodictic certainty of the equation of subject and object is too strong a claim, as our perceived reality, in thinking, is not reality as it truly is (as *Ding an sich*); it is only reality as it appears to us. And on the other hand, if we would adhere to Kant's position we might struggle with the philosophical anxiety of not knowing what the object that grounds all our thought truly is, as it is not matter, nor a thinking being in itself. We will have to see whether these two philosophical positions, true materialism or transcendental idealism, can be combined with a neuroscientific approach, and, vice versa, we will have to see whether we can establish an altered philosophical view on self-consciousness with the aid of neuroscientific theories. In the next section, we will consider Rödl's truly materialist account of receptive knowledge to see whether the material self-consciousness reality he is proposing can be

[233] This is a crucial difference between true materialism and transcendental idealism that will be taken into account in the next chapter.
[234] Kant, *Critique of Pure Reason*, 427 [A 371].
[235] Ibid., 431 [A 379-380].

fleshed out further, and whether we can align it with, or further separate it from, Kant's transcendental account.

V. RECEPTIVE KNOWLEDGE, OR THE REALITY OF OBJECTS GIVEN IN PERCEPTION

Because of the similarity between Rödl's account of receptive knowledge with his account of practical knowledge as expounded in the previous section, I shall highlight only those aspects that strengthen the accounts when beheld together. This entails that I will not focus on his critique on contemporary epistemology, 'sharing in the empiricism of all hitherto existing materialism'.[236]

> We shall depart from this empiricism and develop a truly materialist account of receptive knowledge. According to it, the sensory relationship with an object by which it is known from receptivity is itself known from spontaneity. […] Knowing from receptivity that something is the case is *an act of a self-conscious power*, a power whose acts are known by their subject from spontaneity.[237]

We have seen earlier that, according to Rödl, a power to gain receptive knowledge is formally represented, as a subject of a belief is in a position to explain her belief by ascertaining what to believe by perceiving something. A person's knowledge of acts of this power of knowledge is the same reality as the acts of this power, i.e. receptively knowing that something is the case is the same reality as the knowing that you receptively know. "I know *what I perceive* from receptivity, *by virtue of a receptive nexus with the object* I thus know; but I know *that I perceive* what I do from spontaneity, *by virtue of the object* of my knowledge."[238] This spontaneous knowledge is identical with its object, "I", and is of oneself as oneself, as Rödl nicely puts it, and it is different from receptive knowledge, which is of an independent object in space.[239] Thus, although these two forms of knowledge are interrelated in the sense that they rely on the same power and occur simultaneously, they are presented by Rödl as two separate forms of knowledge, the former is knowledge which provides us with self-consciousness, whereas the latter provides us with knowledge about the outer world. The former seems to have primacy,

[236] I.e. the view in which objects can only be known through the senses.
[237] Rödl, *Self-Consciousness*, 133-134.
[238] Ibid., 144.
[239] Ibid., 143-145.

as receptive knowledge springs from it. When it comes to practical knowledge, the link with material reality is easier to grasp, as the thought that is simultaneously movement results in the physical movement of the body and gives the self its 'substance'. What, then, is the link that Rödl provides when it comes to receptive, or theoretical knowledge?

"The notion of a power – upon which the idea of fallibility follows – is contained in the notion of a material subject of action and knowledge."[240] The concept of receptive knowledge has its reality within us as it figures in first person thoughts. These thoughts are knowledge of material fact. "They represent a material reality, the reality of a self-conscious subject of receptive knowledge. […] The concept of a power of knowledge describes a reality that is such as to be known from spontaneity."[241] And thus it seems that Rödl assumes that the reality of a self-conscious subject is material, based upon the idea that we human beings are material objects in space and time that are substances, and the idea that knowledge from spontaneity includes and is included in its object, non-empirically. Arguably, the idea of the self as a material subject of action is better grounded than the idea of the self as a material subject of knowledge, particularly because the former presupposes the latter in Rödl's philosophy and because actions are part of the perceivable outer world, which is not necessarily the case vice versa. Let us return to Kant to see why the equation of subject with object through belief is problematic.

Things in space and time, according to Kant, like ourselves as objects, can only be *perceived*, given through empirical representation.[242] The pure concepts of the understanding, of which the categories are part of, are applied to empirical intuitions. The categories in particular serve only for the possibility of empirical cognition. "The categories consequently have no other use for the cognition of things except insofar as these are taken as objects of experience."[243] Now, the reality of a self-conscious subject of receptive knowledge known from spontaneity cannot be acquired solely through experience in Rödl's philosophical system, as knowledge from spontaneity must spring from our mind simultaneously. It is therefore inappropriate to apply pure concepts of understanding to the subject, like

[240] Rödl, *Self-Consciousness*, 153.
[241] Ibid., 162 and 163.
[242] This argument is false in Rödl's view. He tries to disprove it by arguing that the knowledge you have of your self as object is spontaneous knowledge and not knowledge through perception.
[243] Kant, *Critique of Pure Reason*, 255 [B 147-148].

substance, as if it is an object, for the subject is not something that can be experienced through perception. On the contrary: the subject is the experiencer, it applies pure concepts of the understanding to grasp the world around it, always standing in connection with itself.

When comparing physiology of inner sense, 'the doctrine of the soul', with physiology of the object of outer sense, 'the doctrine of bodies', Kant notes that although both are concerned with cognizable appearances, i.e. the self and material objects respectively, the former has nothing abiding in the sense that its appearances are only in time and not in space; it does not provide proper cognition of the determinable object, as it is not experienced in space as objects of outer experience are. It only gives cognition of a change of determinations, believes Kant.[244] Kant is here speaking of the changes of the determinable self. "For in that which we call the soul, everything is in continual flux, and it has nothing abiding, except perhaps (if one insists) the I, which is simple only because this representation has no content, and hence no manifold, on account of which it seems to represent a simple object, or better put, it seems to designate one."[245] Even though this might *seem* so, Kant upholds that this "I" is a mere form of consciousness, which only accompanies representations like the soul as a simple object. The thought of the soul as an object does in no way make the soul an object; it is merely a thought, and it is not anything of which can be said with certainty that it has substance.

"Why do we have need of a doctrine of the soul grounded merely on pure rational principles? Without doubt chiefly with the intent of securing our thinking Self from the danger of materialism."[246] Kant proposes that he has done just that: one the one hand, if all matter would cease to exist then all thinking and the existence of thinking beings would be wiped out too, but on the other hand, if the thinking subject is to be taken away, then the whole corporeal world would disappear as well, as it is nothing more than an appearance given through the sensibility of the subject. This is not to say that the corporeal world would cease to exist, but without a subject to experience it, there would be nothing but matter, and no cognition about this matter in any form.

[244] Kant, *Critique of Pure Reason*, 432 [A 381].
[245] Ibid., 432 [A 381]. Paul Guyer and Allen Wood remark the following in footnote *b* on the same page: "Kant later indicated that "perhaps (if one insists)" was to be omitted (E CLXIII p. 49; 23:50)."
[246] Kant, *Critique of Pure Reason*, 433 [A 383].

> Thereby of course I obviously cognize this thinking Self no better as to its properties, nor can I have any insight into its persistence, or even the independence of its existence from whatever transcendental substratum of outer appearances there may be, for this is just as unknown to me as the self is. But since it is likewise possible that I may find cause, drawn from somewhere else than mere speculative grounds, to hope for an existence of my thinking nature that is self-sufficient and persisting through all possible changes of my state, much is still won if, through the free confession of my ignorance, I can nevertheless repel the dogmatic attacks of a speculative opponent, and show him that he can never know more in which to deny my expectations about the nature of my subject than I can in order to hold to them.[247]

One again, one might be struck by the unknowability of the self in Kant's writing, but I believe that he is right in arguing this. The self does indeed seem to be something which is ungraspable in thinking about thinking about the self, and I am not certain whether it is philosophically advantageous if the self as subject is believed to be an object, even though it is hard to uphold that, ultimately, the self is not something physical which occurs within the physical confines of our body. By presenting this quote, I do not mean to depict Rödl as a 'speculative opponent', that would go way too far, and as we shall see in the next chapter, his materialist account of self-consciousness has much potential if it is be used in a stereoscopic view of self-consciousness. There are problems that arise when such manifest images of self-in-the-world are to be placed alongside and aligned with neuroscientific images of self-in-the-world, and these will be discussed. What these manifest images have already proven, however, is that certain manifest aspects of self-consciousness, like the determining self and self-consciousness as spontaneous knowledge, are essential for understanding the possibility of human cognition in general, as they show how human beings as conceptually and abstractly thinking organisms are related to the world around them, their selves always connected to their thoughts and representations of this world.

Kant notes that he might find cause for a self-sufficient and persisting existence of our thinking nature elsewhere. This 'elsewhere' might very well be the neurosciences. The persistence of thinking nature through possible changes is likely to be researched, and its self-sufficiency might just be

[247] Kant, *Critique of Pure Reason*, 433 [A 383-384].

established by looking at the physical functioning of the brain in relation to the rest of the body.[248] Can the neurosciences be the 'elsewhere' when it comes to the enrichening of philosophy of mind? Now is the time to explore this question and others, as we have arrived at the final chapter of this essay on self-consciousness.[249]

[248] This form of self-sufficiency was probably not the kind which Kant had in mind, as he argues in the context of the existence of a soul, but it surely is of interest in light of the aims of this thesis.
[249] Due to the scope of this essay I cannot discuss the last chapter of Rödl's book, in which he discusses 'the second person'.

Chapter Four

Working Towards a Stereoscopic View of Self-Consciousness

Eliminative materialism should be rejected in favour of a stereoscopic view of self-consciousness. A view in which neither the scientific, nor the manifest image should overwhelm the other. In this chapter, I shall offer some suggestions that might contribute to a more uniform image of self-in-the-world, based on that which has been discussed in the previous chapters. I shall focus on Damasio's theory, as it plays an important role in much neuroscientific research on self-consciousness, and on Rödl's and Kant's philosophical theories, as there are quite some connections that can be made between their theories and neuroscientific research, enabling possibilities for a stereoscopic view.

But before we can delve into these matters, I must discuss an important tension which has to be considered if we are to establish any suggestions for a stereoscopic view. On the one hand, the neuroscientific image of self-in-the-world already presents us with a stereoscopic view, but one in which the scientific image overwhelms a somewhat 'naïve' manifest image, in the sense that certain philosophical assumptions are taken for granted by a diverse group of researchers, whilst these assumptions rest on shaky foundations. Emphasis is placed on the scientific aspects of this image, but, as we have seen, a proper manifest image is essential when approaching the subject of self-consciousness.[250] On the other hand, the philosophical images of self-in-the-world that I have presented seems to provide little or no space for a scientific image; they are not stereoscopic views at all. Rödl believes self-consciousness is material, but he does not provide us with the means to better understand this material aspect of self-consciousness, nor does he seem to believe that the inner of the human being can even be investigated empirically by the neuroscientist. Kant believes self-consciousness is transcendental, as it is not something which can be truly understood from within thinking itself, or through empirical investigation; self-consciousness is a function which is a fundamental *precondition* for the functioning of our understanding. Thus, both authors seem to leave little space for the neurosciences to fill in any blanks, and this is something that I wish to amend in this chapter.

[250] It is likely that there are neuroscientists who would disagree with this claim, but I hope that this essay entices them to revisit their views.

I will try to set steps towards a stereoscopic view of self-consciousness with the following comparison as the central thought of this chapter: thinking about thinking from within, in which the experiencer, and that which is experienced as the experiencer are two different things (Kant) vs. thinking about thinking from within, in which the experiencer and that which is experienced are one and the same thing (Rödl) vs. thinking about thinking from within, and from without by empirically investigating the nervous system (neuroscience). With this comparison in mind, let us look at some examples of how philosophy and neuroscience can complement each other. I will start by discussing the possibilities of a more unified philosophical foundation of self-consciousness based upon the different levels of self that we have come across.[251]

I. THE UNITY OF SELF-CONSCIOUSNESS: LEVELS OF SELF

According to Damasio, we should discern three kinds of self: the proto-self, the core self, and the autobiographical self.[252] In this section I argue that it is better to speak of two levels of self, or of one level of self, as a more unified account of self-consciousness provides a sounder philosophical foundation upon which to ground neuroscientific and philosophical research. I will first make some connections with Kant's philosophy, and afterwards I will show how these findings tie in with Rödl's views.

We should disregard the autobiographical self as a concept that represents a form of self-consciousness proper, and we should subsume it under another concept. The autobiographical self is linked to the idea of identity, it is dependent on memories, and it is not usually a part of immediate self-consciousness; it is only during the moments that we think about who we are, by accessing memories, that it is knowledgeable content in consciousness. This is thus a temporarily limited concept that one can have about oneself. You are self-conscious whilst thinking about the past you, in the sense that it is you who is having these thoughts. The "I think" is connected to these thoughts, as it is to any other form of thought. In the previous chapter I suggested that Kant's determinable self might be what Damasio

[251] I wish to stress that most of this chapter should be regarded as a conglomerate that I have procured from the contents of the previous chapters. I thus present ideas here which do not directly reflect the opinions of the authors that I have presented; they are mixtures of my personal interpretations of their works.
[252] Damasio, *The Feeling of What Happens*. See, in chapter two of this essay, "I. The Problem of Consciousness" and "II. From Proto-self to Core Self to Autobiographical Self".

would call the autobiographical self, but in fact, the autobiographical self is only a part of the broader concept of the determinable self. The determinable self, is the self we are conscious of as an object of thought, as an object of inner intuition. This does not only involve the self that is based on memories, but it can involve any form of idea we can have about the self. It can be the idea of self established through introspection, or the idea that it is a feeling that you have of yourself which is always there, and it can indeed be the idea of self based upon who you have been in the past, or a combination of these ideas. Thus, the autobiographical self as a level of self can be subsumed under the category of the determinable self, which is one of the two levels of self that Kant describes, the other being the determining self.

But what about the two other forms of self-consciousness that Damasio describes? I shall now offer a suggestion on how these forms of self-consciousness can be unified within my own adaptation of Kant's concept of the determining self. The core self is the transient entity which is constantly recreated for each and every object with which the brain interacts, and the proto-self is its nonconscious forerunner. The latter is not free to roam at all, and it is composed of the brain devices that regulate our internal milieu. The former is free to roam and it depends on the stability provided by the latter, which provides the roots for the self, the 'something-to-which-knowing-is-attributed'. We have seen that there are multiple layers of self active at the same time, as the state in which you are conscious of the fact that your proto-self is being altered by the perception of an object, is felt as the core self, as it you who is the one that is experiencing. This is unsatisfactory. The layeredness and partial dependency of these levels of self lack a clear unity, as it is unclear where one level of self ends, and where the other begins; we cannot be conscious of this proposed layeredness, nor can we be self-conscious of a self which is nonconscious. Furthermore, the ceaseless recreation of the core self is ontologically problematic. I will further discuss these issues below and in section III "True materialism and Neuroscientific Materialism: Body and Mind".

I propose that the core self and the proto-self should be regarded as both belonging to one and the same entity: the determining self. The something-to-which-knowing-is-attributed, and the feeling of knowing, can easily be connected to the "I think" that must accompany all representations. However, the concept of 'the feeling of knowing' is somewhat problematic, as knowing is not usually described

as a feeling in neuroscience or in philosophy; they are two different concepts which stand on their own. Perhaps it is better to simply call it knowledge, as Rödl does, or to call it 'the original apperception' as Kant does, which is the most basic and most fundamental mental comprehension a person has of oneself that accompanies all thought. From within our thinking there is not much more that can be said about it, as this form of self-consciousness is a precondition for human condition in general. This is the transcendental aspect. However, there is a material aspect to it as well, which is not transcendental at all, but which can be investigated empirically. The brain regions which regulate our internal milieu may very well provide the roots for the determining self, as they are always active, and thus might be a part of this knowledge, or comprehension, which accompanies all representations. At least, they are very suitable candidates as brain regions which are involved, among others, in the *acquirement* of the original apperception.[253] It is of no use to demarcate between the proto-self as something nonconscious, and the core-self as something that we are conscious of, as the latter relies on the former, and as the former seems to provide us with consciousness through the latter. As experiencing subjects, we are not able to say which brain areas we are conscious of and of which areas we are not, as we cannot directly perceive or introspect brain function. Both these forms of self should therefore be regarded as part of the determining self. The determinable self is what we make of that self by thinking about our selves, which is an abstraction. From this, we can deduce that the most fundamental level of self is the determining self, and that the determinable self is secondary, as it is one form of thought among other forms of thought. Let us now turn to Rödl's philosophy. In it, we find an even more unified concept of self-consciousness, one in which it would be a mistake to discern even between these two Kantian forms of self.

As we have seen in the previous chapter, Rödl argues that there can be nonreceptive knowledge of a material substance, in the sense that first person knowledge is knowledge that you have of an object by being that object. When you have an "I"-thought, the object of this thought can only be yourself as a material substance. Thus, this first person knowledge, which is self-consciousness, refers not solely to the determining self, or to the determinable self, but to the human being in its material entirety. This is

[253] That is, the acquirement of this function during human evolutionary development.

the ultimate unity of subject and object, but is it a workable unity? It seems that the autobiographical self plays no role in this way of looking at self-consciousness, as spontaneous knowledge is always something which an individual has in the present. I think that Rödl would therefore disregard the concept of an autobiographical self as belonging to the inquiry into self-consciousness. However, Rödl's idea of self-consciousness can be united with Damasio's ideas of the core and the proto-self, which could be described as the consciousness of spontaneous knowledge, or simply as self-consciousness, which is always there during the processing of receptive knowledge. The something-to-which-knowing-is-attributed cannot only be connected to the "I think" that must accompany all representations, but also to spontaneous knowledge. The difference between these approaches is that Damasio and Kant are somewhat vague about what this 'something' or mental comprehension exactly is, whereas Rödl upholds that it is our material selves in their entirety.

A more unified account of self-consciousness provides us with a less clouded philosophical and neuroscientific basis for stereoscopic research. We could speak of one fundamental level of self, the determining self, and of one secondary level that includes the autobiographical self as part of the determinable self, or we could speak of one level of self which includes every immediate form of self-consciousness, but excludes the autobiographical self, as the spontaneous knowledge which accompanies the believes one could have about one's past self, is always in the here and now, which is self-consciousness proper. The former excludes the immediate form of self-consciousness insofar it involves the self as an object, whereas the latter includes this form, but excludes the autobiographical self as it is non-spontaneous. The researcher who would like to set forth a stereoscopic view could, if she so wishes, choose between these two unified accounts for approaching self-consciousness. I believe that both approaches are viable options. But before we can take a closer look at the options for further integrating the philosophical with the neuroscientific to see why, we must first reflect on the problems that one is faced with when working towards a stereoscopic view of self-consciousness.

II. BRIDGING THE SELLARSIAN GAP

In the first chapter of this essay I presented Sellars' theoretical framework of the two images of man-in-the-world: the manifest image and the scientific image. The philosophies of Kant and Rödl are part of

the manifest image, as they can be grouped under the perennial philosophy of man-in-the-world. In their philosophies, the manifest image is promoted as real, and that which is unexperienceable is not. They do not include the specific type of scientific reasoning in their writings belonging to the scientific image; the postulation of imperceptible entities, and principles pertaining to them. It does indeed seem that their manifest images are presented as adequate accounts for understanding the reality of self-in-the-world. Sellars argues that in a manifest image science is merely used as a tool to better understand specific aspects of the reality of man in the world, but when it comes to the concept of self-consciousness, both Kant and Rödl seem to not even allow science this modest role, as they do not seem to uphold that the self is one of these aspects. Thus, from Sellars' point of view, these philosophers indeed seem to refine and endorse a misconceived and dominant manifest image with regard to the concept of self-consciousness, which does not accommodate the scientific image at all. This is problematic for our Sellarsian project, as both images are to be blended in a stereoscopic view in which neither image should overwhelm the other if we are to adhere to Sellars' aims.[254] Hence for our project to proceed, we must find a middle way between the conceptions by Kant and Rödl, in which the manifest image is dominant, and the eliminative materialist's position, in which the scientific image is dominant.

How can the dialectical gap between the manifest image and the scientific image of man-in-the-world be bridged when it comes to something that is as philosophical as the concept of self-consciousness? We have seen that most of the neuroscientific ideas on self-consciousness that I have presented, are philosophical at their core. The starting point of neuroscientific enquiry always seems to require a manifest idea on what self-consciousness is, and by starting from this idea, postulational theories based upon imperceptible entities and principles pertaining to them are presented as belonging to a dominantly scientific image. This leads me to believe that there is room for a manifest philosophical foundation combined with scientific inquiry in a stereoscopic image, even though Kant and Rödl themselves might not support such a route of enquiry. In the following sections I will show some possibilities of how philosophical contemplation about the seemingly ungraspable can be aligned with scientific postulation. Self-consciousness is an abstract concept that is not easily defined, and therefore

[254] Sellars, *Science, Perception, and Reality*. See, in chapter one of this essay, "II. The Manifest Image – Leaping Away from the Original Image".

it could benefit from both philosophical reflection and empirical investigation.

It is interesting that Sellars grants a slight primacy to the scientific image, as the manifest framework lacks an external point of view with which to defend its reality, and as Sellars upholds that the scientific image is the more intelligible, or more truthful account, of what there is, whilst it is reliant on the manifest image for its methodological foundation.[255] When it comes to the concept of self-consciousness, however, it is much harder to argue that the scientific image provides a more intelligible, or more truthful account, than it does for e.g. the concept of gravity, as it is such a manifestly human concept. Nonetheless, the external point with which to defend its reality can truly strengthen a stereoscopic view of self-consciousness, as the neurosciences might be able to provide evidence for the material reality of self-consciousness from without the manifest framework, even though it is such a manifest concept. Because the concept of self-consciousness is surrounded by uncertainty, there is much more that we might still learn about it, especially from a stereoscopic point of view. It is therefore that we must investigate the possible cohesion between philosophy and neuroscience. I believe that here is a strong connection between these two disciplines which can help us in gaining a greater understanding of the workings of the mind.

In the following two sections I will try to present several ideas on how we can work towards a stereoscopic view in which the philosophical conceptions of self-consciousness by Rödl and Kant function as philosophical foundations for neuroscientific research, and in which this neuroscientific research could tell us more about the material reality of self-consciousness from an external and empirical point of view. First, I will offer some suggestions on how Rödl's true materialism can be united with neuroscientific materialism within the context of the dualism of body and mind. And secondly, I will look at some possibilities for the unification of transcendental idealism with neuroscientific research.

[255] Sellars, *Science, Perception, and Reality*. See, in chapter one of this essay, "IV. The Scientific Image – The Unification of Entities and Principles," and "V. The Clash of Images – The Fusion of Worlds".

III. TRUE MATERIALISM AND NEUROSCIENTIFIC MATERIALISM: BODY AND MIND

As became clear in the previous chapter, Rödl upholds that spontaneous knowledge is identical with its object, "I", and is of oneself as oneself. This is knowledge which provides us with self-consciousness, and which is knowledge of material fact representing a material reality; the subject being an object. And, as we have seen in chapter two, Neuroscientists also investigate self-consciousness as part of a material reality, as they try to establish which areas or devices of the brain are responsible for our self-consciousness. However, in their writing we have not often come across proper philosophical principles that are as far-reaching as Rödl's with which to ground their research. Damasio's account is the best-defined theory, but it lacks unity, and it does not provide a proper ontological framework for defining what the core self really is. His argument that it is ceaselessly recreated for each object the brain interacts with, gives it a weak ontological status, for something that is constantly recreated also ceases to exist at each moment of recreation. From something to nothing, and from nothing to something. This makes it unfathomable, and therefore a troublesome starting point for the neuroscientist. It is also at odds with the seeming stability that is provided by the proto-self, which is nonconscious, and thus hard to relate to something that is supposed to provide us with a form of consciousness. This strengthens me in my conviction that we should look for a more unified account of self-consciousness, in which the proto-self and the core-self should be brought together. With the help of Rödl's theory we can provide a stronger ontological basis for a unified account of self-consciousness that is fitting for neuroscientific research, and vice versa, we can strengthen Rödl's philosophy by providing more empirical grounds for his materialist account of self-consciousness. Of course, I do not uphold that I can singlehandedly offer the ultimate synthesis; I can only provide some suggestions, which might later be worked out in greater detail.

If self-consciousness is nonreceptive knowledge of a material substance, and if everything is material, then self-consciousness is to be found in the material entity that we call the human being. Rödl does not specify whether that which provides us with self-consciousness is to be found in a certain part of the human being, or whether it is the human being in its entirety that is self-consciousness thought. The latter option seems the less likely one, as, for example, our nails and hairs are not directly involved

in sentience, or in reason, as there are no nerves to be found in them, only in the connecting tissue. It could be countered that they are part of sentience due to the rational acts than can be performed with them, but to say that we have control over our nails and hair, and that we can make use of them, does not entail that these inert parts of our bodies rational or sentient themselves; they merely contribute to sentience and reason through their possible uses and functionalities. The part of the body which is the most likely candidate for providing us with the most fundamental aspects of self-consciousness, is the brain. However, the brain does not operate on its own, and in normal circumstances it requires the functioning of a large part of the human body to be able to function properly. A brain without a body would not be provided with the nutrients and oxygen necessary for its functioning, nor would it have any sensory input whatsoever. This is related to Kant's idea that "[w]ithout sensibility no object would be given to us, and without understanding none would be thought."[256] A body without a brain would be lifeless, as the necessary regulation of its internal milieu would be absent, and as it would receive no impulses or hormones with which its state would be altered. It thus seems that the brain is as fundamental for the body, as the body is for the brain. Of course, many other parts of the body are essential for the proper functioning of the human being, and they are reliant on each other, but the brain plays a special role as it is the part which steers the other parts, whilst being reliant on them. The brain provides the human being with reason. However, we should regard all the parts that the human being is made of as belonging to one material entity without giving primacy to the brain (brain over body) or the body (body over brain), which is something that Rödl's theory entices us to do. The human being functions properly only in its entirety, and it is self-conscious of itself being a material substance in its entirety. The part of the body which provides us with this self-consciousness can be investigated empirically, even though as subjects we cannot be certain that it is the provider of self-consciousness, as it is not experienced as such. The determining self cannot be determined by ourselves, but it can be studied from without.

The problem that the neuroscientist must tackle, if she would choose to adhere to Rödl's philosophical position, revolves around the determining of the brain areas that are responsible for this form of knowledge which gives us our self-consciousness. If we return to chapter two, we can see that

[256] Kant, *Critique of Pure Reason*, 193 [A 51/B 75].

material knowledge is not a concept that neuroscientists usually refer to when discussing self-consciousness, but there are several possible connections that I will now discuss. As we have already treated Damasio quite extensively, we shall now turn to Edelman and Tononi. In their work, they speak of a 'sense of self', and of the evolutionary development of linguistic capabilities enabling self-consciousness, which seem unfit for unification with Rödl's broader concept of knowledge. Their theory of 'reentry', however, does seem to offer a possible connection with the concept of material knowledge. The rapid and reciprocal interaction between a large number of groups of neurons, which underlines the integration of functionally segregated properties of the brain, fits in well with non-receptive spontaneous knowledge which occurs simultaneously with the processing of receptive knowledge.[257] The example of me moving my arm intentionally as a conclusion of reasoning giving me self-conscious knowledge, must surely involve many brain regions that are active at the same time, presumably without a clear-cut central coordinative area. This example nicely shows how body and brain are to be studied as being two parts of one material entity. This can also be seen in the research by Park et al., as they suggest that their neurophysiological data experimentally link the cortical mapping of the internal body to self-consciousness, and that the neural processing of internal body states is a fundamental biological mechanism for the subjective aspect of conscious experience.[258] The idea of bodily self-consciousness supports the claim that it would be more fruitful to disregard the demarcation between the proto-self and the core self, and instead use a more unified concept, like Rödl's concept of spontaneous knowledge, as the processing of sensory input from our body by our brain seems so fundamental for our self-consciousness.

Vogeley et al.'s description of self-consciousness can also be united with Rödl's theory to a certain extent, as they deem the ability to take a first-person perspective necessary to represent and integrate mental states into a common framework. Taking the first-person perspective (1PP) results in first-person knowledge. For Rödl, the relation between perception and self-consciousness is also fundamental, but taking a third person perspective (3PP) also results in first-person knowledge, as the

[257] Edelman and Tononi, *A Universe of Consciousness*. See, in chapter two of this essay, "III. Philosophy, Language, Science, and the Self".
[258] Park, "Bodily Self-Consciousness," 8453-8460. See, in chapter two of this essay, "IV. Bodily Self-Consciousness Measured".

believe that I obtain during such an exercise, i.e. the assessment of the amount of red balls, is accompanied by the spontaneous knowledge that it is I who does the believing. The overlap with the deactivation pattern of all four conditions supports this claim. Even though neural activity in the regions of the default mode of brain function (1PP) decreases when performing a specific task like taking a 3PP, these areas are still active, and are thus able to continuously provide us with first-person knowledge, i.e. with our self-consciousness.[259]

The description of the self by Decety and Sommerville cannot be united with Rödl's, as they argue that it is a multi-dimensional construct, whilst Rödl upholds that the self is our knowledge of material fact representing a material reality, which in no way is a multi-dimensional construct. Nevertheless, Decety and Summerville's claim that a collection of interconnected brain regions is essential for the subjective experience of a 'self', is something that all neuroscientists that I have discussed seem to uphold, and it could be investigated further by taking a stereoscopic perspective.[260] Rödl's philosophy seems a good candidate for a manifest image to support such research, as many brain areas are presumably active when it comes to first person spontaneous knowledge, which involves perception, action, and reasoning. Decety and Summerville posed the question whether cognitive neuroscience can help to generate a more parsimonious description of the self, which they answered to some extent by stating that "one potential role that cognitive neuroscience can play in the study of the self is to help conceptually define the different dimensions, aspects and characteristics of the self and to help address the potential separability or relatedness of each component part of selfprocessing."[261] I do not believe that without philosophy, cognitive neuroscience can generate a more parsimonious description of the self. To separate aspects of the self into components is problematic, as it may lead to conceptual confusion without proper philosophical analysis; the conceptual defining of dimensions, aspects, and characteristics of the self is a very philosophical endeavour, because it can only be done from within our own thinking, and not by empirically investigating the brain from without.

[259] Vogeley et al., "Neural Correlates of First-Person Perspective," 817-827. See, in chapter two of this essay, "V. Neural Correlates of the First-Person Perspective as a Constituent of Self-Consciousness".
[260] Decety and Sommerville, "Shared representations between self and other," 527-533. See, in chapter two of this essay, "VI. Social Cognitive Neuroscience: Self and Other".
[261] Ibid., 528 (Box 1).

Gallagher's description of the minimal self as the most basic kind of self imaginable is in line with Rödl's self, as both concern the immediate experience of the self in the present, and as the 'immunity principle' applies to both.[262] They differ in the sense that Gallagher believes that the self is accessible without the need of any perception of the world, whereas Rödl's conception of self always seems to involve perception in one form or another, as both belief and action require perception for spontaneous knowledge to unfold.[263]

I hope to have shown that although the philosophical position of true materialism has its limitations, it also provides possibilities for synthesis. Let us now look at an alternative: transcendental idealism. I believe that the limitations of this position might also be diminished with the aid of neuroscientific research. The locating and delimiting of self-consciousness is very important in this regard. Another key aspect that is involved, which might benefit the neurosciences, are the bounds of sense, as Kant shows that we are limited in what we can know about the world and ourselves. These limitations can more strictly demarcate what it is that we write about when we discuss self-consciousness in our research, not only as neuroscientists, but also as philosophers.

IV. TRANSCENDENTAL IDEALISM AND THE NEUROSCIENTIFIC OBJECT: LOCATING AND DELIMITING SELF-CONSCIOUSNESS

I ended the previous chapter with the question whether the neurosciences can be the 'elsewhere' when it comes to the enrichening of philosophy of mind, as Kant suggests that he may find cause to hope for an existence of thinking nature that is self-sufficient and persisting, drawn from elsewhere than solely from speculative grounds. I believe that it is safe to say that over two-hundred years later we still do not know whether we can speak of (1) self-sufficiency of thinking nature, or of (2) persistence of thinking nature through all possible changes of one's state, but that science has brought us closer to establishing grounded knowledge on these issues. Before we discuss transcendental idealism proper, we shall shortly discuss these two matters.

The persistence of thinking nature through changes of one's bodily and mental states are

[262] The immunity principle is described as "immunity to error through misidentification relative to the first-person pronoun." Gallagher, "Philosophical conceptions of the self," 15.
[263] Ibid., 14-21. See, in chapter two of this essay, "VII. The Philosophical Self in the Cognitive Sciences".

extensively researched. Strokes and brain injuries can, for example, drastically alter someone's brain functions, and change the way in which a person experiences reality, or interacts with the outer world. We are pretty sure that thinking stops once one is dead, as any form of brain activity is absent; there seems to be a strong correlation between thought and brain activity. The research that is done with people whose brain 'deviates' from the 'normal and healthy' brain, as the manner in which they think also deviates, is very relevant for this essay. I will return to the topic of the abnormal and the normal in the next section. When Kant wrote about thinking nature that is self-sufficient, he presumably referred to speculation about the self-sufficiency of the soul, but the self-sufficiency of our self-consciousness is something that we can do more than merely speculate about. Self-consciousness is *not* a self-sufficient entity that can exist without the brain or the body. However, self-consciousness *is* self-sufficient in the sense that it is the living body and brain that form self-consciousness.

Any attempt at uniting transcendental idealism with contemporary neuroscience will quickly lead to the following paradox: how can we establish which brain areas are responsible for the determining self, whilst we cannot know what the determining self truly is? In other words, what can we find out empirically, about that which is transcendental and rational? Self-consciousness, according to Kant, is a function of thought, as the "I" is the vehicle that accompanies all categories, and thus all representations. It is an *a priori* ground for gaining knowledge in any form. It precedes thought, and it cannot be an object of thought. But whilst the determining self cannot be an object of thought, I argue that the brain as part of the body can be investigated as an object which somehow houses this function, whilst maintaining the philosophical position of transcendental idealism.[264] I believe there are at least two viable options, which tie in with each other, for doing so.

One option would be to further look into the 'default mode of brain function'. As every possible thought is accompanied by self-consciousness, it is highly likely that it always involves a form of brain activity that is always there. Presumably, it is even active when we sleep, as dreams are also thoughts in which we seem to have some sense of self; these thoughts are usually connected to the subject thinking them. The other option, which might be strongly connected with the former, would be to try to locate

[264] I.e. my own interpretation and representation of Kant's transcendental idealism.

the determining self by further looking into the brain areas that are responsible for the regulating of our internal milieus. These areas have a constant necessary interaction with the rest of our body. Such a fundamental brain function does not only provide the necessary stability for our survival, but it could also provide the necessary stability for self-consciousness, as was argued by Damasio. These areas might be involved in providing the function of thought that is the original apperception, always connecting the body with the brain, and our representations with the "I think". And, perhaps through the process of re-entry, these areas might be connected with other parts of the brain which provide us with reason in general, making us self-conscious beings and truly fully aware of ourselves. Thus, the mental comprehension we have of ourselves would then be directly related to our bodies in their entirety, and not only to our brains.

With regard to what has just been suggested as a possible avenue for further research, it could be countered that Kant would never have allowed such a blatant examination of that which is transcendental. Kant might have indeed rejected my endeavour as foolish in his time, but by investigating the brain areas that are responsible for the determining self with present-day technology, one does not present the determining self as an object of thought within pure reason, but as an object of empirical scientific enquiry, in the sense that only its physical location is investigated, whereas not much more would be offered about what this self-consciousness is as a *Ding an sich*. Thus, we would not exceed the limitation posed by Kant that the determining self can never be an object of thought, as it is prior to all thought, because we would solely postulate its existence in brain matter as a starting point, whilst upholding Kant's philosophical conception of it, and from there start an empirical investigation of the brain matter which might be involved in its occurrence in us as human beings. The delimiting of self-consciousness is very important in this regard, and it is a point we will now turn to.

In the *Critique of Pure Reason* Kant makes a clear demarcation between the knowing subject and the knowable object, and he upholds that the knowing subject is limited in her sensory and cognitive capabilities, and that the outer world that we perceive and think of is a limited representation of this world. A knowable object in the outer world can only be known by us to the extent that our sensory and cognitive abilities allow us to, and thus we cannot know it as a *Ding an sich*, i.e. as it truly is. For example, we can behold a chair, but we do not perceive the chair in its totality, as we do not see every

single atom this chair consists of, let alone the even smaller particles. This idea, of which I have only given a very narrow synopsis, has been very influential since Kant, and I believe that it can still be very important for philosophy of science today. Delimiting what we can know about certain objects is highly important for trying to establish what is merely speculative, what is postulational, and what is manifest. To give some examples: the persistence of the soul after death is a speculative idea, the existence of electrons is a postulational idea (as they are not immediately perceivable, whilst scientific inquiry has shown them to be manipulatable entities), and the existence of the human being is a manifest idea. Throughout this essay we have encountered different conceptions of what kind of idea the existence of self-consciousness is, and we have encountered them in all the three forms that I have just mentioned. Kant has tried to show which aspects of self-consciousness can be known by us and which cannot, and I believe that this aspect of his philosophy can also be useful in a stereoscopic view, in the sense that we can establish what can and cannot be said about the brain as an object of scientific inquiry, as part of the self-conscious human being.

One of the principles that should be kept in mind by the scientist and the philosopher at all times in a Kantian-inspired stereoscopic view, is that they are limited in their sensory and cognitive capabilities. Although it is commonly upheld by philosophers and scientists that scientific knowledge is the most objective form of knowledge, one should never forget that this knowledge is procured by humans and only understood by humans. Knowledge in any form is thus always something which is procured and shared by the human lifeform, and our sensory and cognitive capabilities shape and delimit it. This is not to say that these boundaries are fixed, as we produce tools with which we can perceive things differently, like for example electron microscopes with which we can 'see' molecules, and as we develop postulational theories, with which we can transgress the boundaries of sense in the search of new knowledge. Even though it is a noble pursuit to strive for the obtainment of absolute knowledge, in which we would understand the universe in its all-encompassing totality as a *Ding an sich*, this can never be achieved. And this is also the case when it comes to self-consciousness.

The determining self, that which provides us with self-consciousness, or spontaneous knowledge, is a very human conceptualisation of something which is transcendental. Kant has made it clear in the "Transcendental dialectic" in the *Critique of Pure Reason* why it is impossible to ascertain

the existence of God, as the idea of God is transcendental; we are unable to experience the existence of God with our human capabilities.[265] As the determining self is also something which cannot be experienced in space and time by the subject personally, we can ascertain that it is also, in part, a transcendental idea. In part, because self-consciousness is also something which is manifest to us; we are aware that we are conscious of ourselves, but we do not truly know how this is so. It is thus a different sort of idea than the purely transcendental idea of the existence of God. I believe that we can and should have this philosophical conviction and still try to study self-consciousness from without, as we have quite convincing ideas about the material reality of self-consciousness in the human being as an object. Thus, whereas the idea of God is completely transcendental, the idea of self-consciousness is also in part manifest, and this enables us to come up with postulational theories about it. Scientific inquiry into the metaphysical can be very fruitful, as long as philosophical limitations and principles are taken into account.

V. THE ABNORMAL AND THE NORMAL – THE HUMAN AND THE PRIMATE

Now that we have taken a closer look at two options for further integrating the philosophical with the neuroscientific, it is time to look at two other possible avenues of research which might be helpful for the procurement of a stereoscopic view of self-consciousness. First, we will discuss some possibilities for investigating the functioning of the 'deviant' brain in comparison to the 'normal' brain by returning to some of the examples from chapter two. And secondly, we will discuss some of the possibilities provided by analysing the difference in brain function between humans and other primates.

In the two primary philosophies that I have discussed in this essay we have not encountered any form of demarcation between individual human beings; Rödl and Kant both seem to discuss their philosophy of mind with the general human being in mind. They try to understand self-consciousness from the perspective of a normally functioning human being, from a perspective that can be understood by other people who think in a similar manner, or who might be convinced by the philosophers that their thinking functions in the manner as the philosophers describe it. Both authors discuss principles of

[265] In particular, see chapter III, "The ideal of pure reason" in: Kant, *Critique of Pure Reason*, 551-589 [A 571/B 599-A 642/B 670].

thought, and the conditions of self-consciousness, but they do not write about what happens when these principles are transgressed, or what happens when these conditions are not met, whilst it is not unthinkable that there are people whose thinking processes might deviate in such ways. People who might think in a different manner are not necessarily included in Kant's and Rödl's analyses, whilst we might learn a lot from them.

The research on schizophrenia as discussed by Gallagher is of great interest for the philosopher and the neuroscientist.[266] The example of schizophrenic patients who suffer from 'thought insertion' seemingly violates the 'immunity principle' defy Rödl's concept of spontaneous first person knowledge, and Kant's concept of the determining self. It defies the former's concept because the patient is convinced that she has the thoughts of others which are inserted into her mind, which indicates that the spontaneous knowledge that should accompany this believe does not refer to the first person, but to a third person. This knowledge is thus not self-conscious, but 'other-conscious'. In light of Rödl's philosophy the question that rises is whether this should then be considered as thought, as it does not refer to the first person, which is necessary for the proper functioning of reason, or if this knowledge is simply wrongly interpreted first person knowledge. It defies the latter's concept because the "I think" does not seem to accompany all representations, as some representations are accompanied by a "she/he thinks" as another form of apperception. These inserted thoughts are experienced as the thoughts of others, and the original apperception does not seem to function as their category; these thoughts are not connected to the self. It is hard to place oneself in the mind of a schizophrenic patient, and to try to imagine how such an experience might be. What we can do, however, is compare their brain activity with the help of experiments and fMRI-scans, and see whether they have a different default mode of brain function compared to non-patients, and whether there are indications that there is a limited form of reentry perceivable, whilst maintaining a philosophically grounded idea on what self-consciousness is. There is research done which points in this direction. For example, a recent MRI-study by Josefina Castro-Fornieles et al. has shown that over a period of two years people with early-onset schizophrenia present a larger gray matter decrease in several brain areas compared to the control group, presumably

[266] Gallagher, "Philosophical conceptions of the self," 14-21. See, in chapter two of this essay, "VII. The Philosophical Self in the Cognitive Sciences".

affecting IQ and working memory.[267]

Another interesting example is presented by Damasio, who writes about the man who stopped interacting with Damasio as a person. This man started randomly looking and moving about in the room, and he did not respond to his name. Damasio argued that "[n]eurologically speaking, he had an absence of seizure followed by an absence automatism, two among the many manifestations of epilepsy, a condition caused by brain dysfunction."[268] Damasio interpreted this phenomenon, an impairment of consciousness, as a transition to a mind deprived of the sense of self. I think that he might be right in his interpretation, as in the theories that I have presented, consciousness and self-consciousness go hand in hand. I believe that the 'sense of self' is comparable to spontaneous first person knowledge, or the original apperception, as all three of these concepts are the most fundamental aspects of self-consciousness in their respective theories. Therefore, it might be worthwhile to investigate patients suffering from this specific form of epilepsy. If it would be possible to make an fMRI-scan of a patient during such an episode of brain dysfunction, then it would be interesting to compare it with an fMRI-scan of the default mode of brain function of said patient, and with the MRI-scans of the default mode of brain function of persons who do not suffer from epilepsy for control.

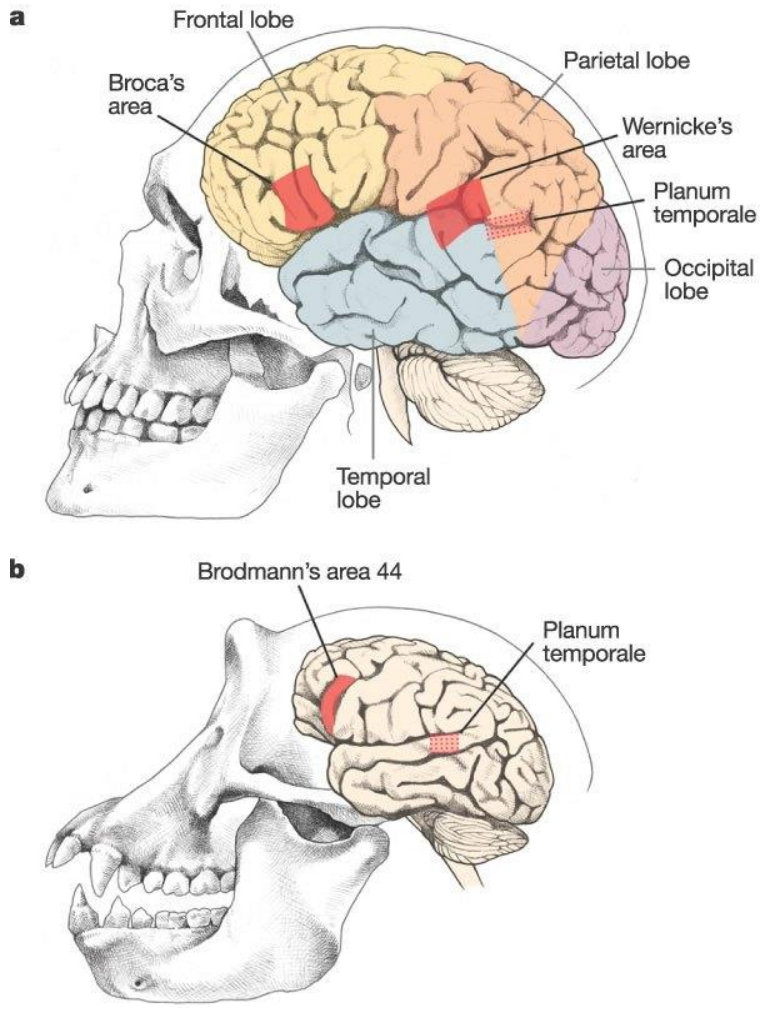
A different area of interest for the philosopher and the neuroscientist investigating self-consciousness, is the functioning of cognition in animals. Primate cognition is of particular interest, as we share certain brain structures with other primates. In 2010 Amanda Seed and Michael Tomasello, specialists in evolutionary psychology, published an article in which they state that in the past 30 years research into primate cognition has flourished. "It has established beyond a reasonable doubt that many of the most basic skills of human cognition are already possessed by our living primate relatives".[269] For example, most researchers would agree that all primates perceive the same physical world of objects in space, as we share the same basic perceptual organs and brain organizations. Not only is there general agreement among researchers that all primates are able to experience the same basic phenomenal world,

[267] Josefina Castro-Fornieles et al., "Gray matter changes and cognitive predictors of 2-year follow-up abnormalities in early-onset first-episode psychosis," *European Child & Adolescent Psychiatry*, accessed July 18, 2017, doi:10.1007/s00787-017-1013-z.
[268] Damasio, *The Feeling of What Happens*, 6.
[269] Amanda Seed and Michael Tomasello, "Primate Cognition," *Topics in Cognitive Science* 2 (2010): 408.

but "there is also little disagreement that they also go beyond this to cognitively represent absent objects, spatial relations, and some categories and quantities."[270] Whereas some other researchers uphold that the main difference between humans and other primates seems to be that humans have the unique ability to reason about abstract relations, both physical and social, Seed and Tomasello argue that the difference in cognitive abilities provide humans with an 'entrée into culture'.[271]

As it is already very hard to establish what self-consciousness is for humans, it is even harder to establish whether animals are self-conscious, let alone in what way. Although self-consciousness is not mentioned as one of the most basic skills of human cognition by Seed and Tomasello, we have seen in the theories of Damasio, Kant, and Rödl that self-consciousness is the cornerstone of human consciousness. The question that arises, whether other primates have some form of self-consciousness, is therefore not that far-fetched, especially if we consider that we share certain brain structures. The similarities, especially in shape, and dissimilarities, especially in size, between the human brain and the chimpanzee brain can be seen in the image below. The mirror self-recognition test, developed by Gordon G. Gallup Jr., indicates that certain great apes, like chimpanzees, have some sort of self-awareness. "After prolonged exposure to their reflected images in mirrors, chimpanzees marked with red dye showed evidence of being able to recognize their own reflections."[272] After 80 hours of exposure to a mirror, Gallup anesthetized several chimpanzees and marked them with red dye. After awakening, the marked portions of the skin were touched spontaneously four to ten times within thirty minutes by the chimpanzees, whereas they were touched only one time during normal conditions. The time during which the chimpanzees watched themselves in the mirror increased by a factor of more than four whilst marked.[273] This seeming self-awareness could perhaps be a form of self-consciousness in chimpanzees.

[270] Seed and Tomasello, "Primate Cognition," 410.
[271] Ibid., 408, 410, and 413-414.
[272] Gordon G. Gallup, Jr. "Chimpanzees: Self-Recognition," *Science* 167:3914 (1970): 86.
[273] Ibid., 86-87.



A representation of the human brain (a) and of the chimpanzee brain (b).[274]

It might be interesting to investigate whether primates, like chimpanzees, also have some sort of default mode of brain function, and if it is in some way comparable to the human default mode of brain function. This could hint at the existence of self-consciousness in other primates, and it could entice researchers to do more comparative research on this topic. Other brain areas of interest are the brain-stem nuclei which regulate the internal milieu, as they might provide the roots for the self, the 'something-to-which-knowing-is-attributed' in Damasio's theory, or the roots for Kant's determining self as I have suggested above. Comparable structures are presumably found in all primates, and thus they might be suitable candidates for comparative research as well. Rödl and Kant would probably not support such research, as self-consciousness is a fundamental part of human reason, and therefore not

[274] Sean B. Carroll, "Genetics and the making of Homo sapiens," *Nature* 422 (2003): 852.

something which is to be found in animals. They would not likely regard the seeming self-awareness in chimpanzees as self-consciousness, as primates do not exhibit the same sort of rationality that humans do. Furthermore, it is impossible for us to experience and to think like a primate, as this is not a part of our phenomenal world, thus we can never establish with any certainty how primates are conscious or self-conscious. However, what we can observe as part of the phenomenal world is that primates have comparable brain structures, and that they exhibit certain behaviour which is vaguely reminiscent of human behaviour. Could this provide the neuroscientist with sufficient grounds for doing comparative research?

The possibilities that I have presented for a stronger stereoscopic view in this chapter are merely suggestions, but I hope to have shown that there are indeed many possibilities for further research in which both philosophy and neuroscience play an important role, and in which neither discipline should overwhelm the other. They should only strengthen each other, and broaden our horizons. Although at first sight Kant and Rödl seem to leave little room for the neurosciences to fill in any blanks, there are quite some possibilities for filling them in. And vice versa, the theories on self-consciousness by neuroscientists could benefit from a more unified philosophical foundation, as provided by Rödl and Kant.

Conclusion

Certain manifest aspects of self-consciousness are necessary conditions for understanding the possibility of human cognition in general, and these ideas can function as philosophical principles upon which a neuroscientific and stereoscopic image of self-consciousness can be built.

This is the thesis we set out with in the introduction, and now the time has come to return to our most important findings pertaining to this thesis and to reflect on them. By doing so, we can see how the separate arguments in the four chapters are connected to each other, and how they ultimately contribute towards the substantiation of this thesis.

Through the analysis of the different conceptions of self-consciousness within Sellars' theoretical framework of the images of man throughout this essay, we have now arrived at the culmination of the following threefold argument. The first part of the argument is that although neuroscientists present their findings on the brain areas which are deemed to be responsible for our experiencing of self-consciousness within a scientific image of man-in-the-world, in the sense that they make use of postulational entities, and principles pertaining to them, their ideas on what self-consciousness is, are still inherently manifest, in the sense that they are philosophical, 'sophisticated common sense' ideas which are quite easily placed within the tradition of perennial philosophy. This entails that it is wrong to uphold that "our common-sense understanding of psychological states and processes is deeply mistaken and that some or all of our ordinary notions of mental states will have no home, at any level of analysis, in a sophisticated and accurate account of the mind" when it comes to the concept of self-consciousness, for we human beings are the only interpreters of the workings of the mind, and it does not suffice to simply say that we are all mistaken, for all knowledge might ultimately be disproved, or at least thought of otherwise, in a later age, be it scientific or not.[275] Thus, to return to an important question posed in the first chapter: manifest man survives in the synoptic view of the neuroscientific image of man.

The argument presented above is supported by my rejection of the eliminative materialist view.

[275] William Ramsey, "Eliminative Materialism," *Stanford Encyclopedia of Philosophy*, accessed July 31, 2017, https://plato.stanford.edu/entries/materialism-eliminative/.

This view purports that "[i]t is just (non-trivially) false that there are any sensations. The term 'sensation', for all its apparent referring function, fails to denote anything, just as 'unicorn' does."[276] The term 'sensation' does denote something, it denotes "the effect of an object on the capacity for representation, insofar as we are affected by it".[277] Without the raw material of sensible sensations, given to us through our sense organs, our understanding would have nothing to experience; we would not know that there is an outer world. Without sensations, there would be no objects given to us, and thoughts without content are empty. However, there can be no empty thoughts, for nothing cannot be something, whilst each thought is a representation of something, and each representation is accompanied by the original apperception, the "I think", the vehicle of all concepts. It is therefore that a 'sensation' as a concept is to be regarded as a manifest aspect of self-consciousness that is fundamental for understanding the possibility of human cognition in general; without sensation, there would be nothing to connect your self to, and without sensation there would be nothing to cognize at all. The same cannot be said with regard to unicorns, and therefore we should reject the eliminative materialist view.

The second part of my argument is related to the result that the philosophical conceptions of self-consciousness are utterly manifest images of man-in-the-world. These two great thinkers have provided us with very elaborate accounts on what they think that self-consciousness is, but despite their philosophical completeness and uniformity, they are found to be wanting; they lack a scientific image, an external point of view with which to defend the reality of the concept of self-consciousness. If we are to make use of their philosophies in a stereoscopic view, we should accommodate the scientific image to such an extent that neither image holds primacy. Immanuel Kant's transcendental idealistic account teaches us that there are two kinds of self; the determinable self and the determining self. The former is the self that we can be consciousness of as an object of thought, but it does not provide us with self-consciousness proper, and the latter is the self that is a precondition for having objects of thought in a coherent fashion, always connecting the self to each representation, providing us with self-consciousness proper, but it cannot be an object of thought. This form of self-consciousness is a fundamental part of human cognition, and it is thus a necessary condition for understanding the

[276] Lycan and Pappas, "What is eliminative materialism?," 150.
[277] Kant, *Critique of Pure Reason*, 155 [A 19-20/B 34].

possibility of human cognition in general. Sebastian Rödl's true materialist account provides us with the most unified conception of self-consciousness. In his account self-consciousness is spontaneous knowledge that is identical with its object, "I" as a human being in its entirety, and it is knowledge of oneself as oneself. All our acts and believes are accompanied with this spontaneous knowledge, always connecting our material selves to these acts and believes. Thus, also in this explanation, the manifest aspect of self-consciousness as spontaneous knowledge that one has of oneself, is a necessary condition for understanding the possibility of human cognition in general, as acts and believes would otherwise not be connected to one's self, and thus be random, ununderstood thoughts, which could not function as a part of proper human cognition.

The third and final part of the threefold argument is that the philosophical views which I have outlined in the third and fourth chapters can function as philosophical principles of science in neuroscientific research, and that neuroscientific findings can bolster philosophical ideas on self-consciousness, which might ultimately lead to a truly stereoscopic image of self-in-the-world. Not only are Rödl's and Kant's conceptions of self-conscious more unified and bulwarked accounts than neuroscientific conceptions of self-consciousness, i.e. the manifest aspects of these neuroscientific accounts (Damasio's in particular), but they also lend themselves quite well as philosophical foundations upon which a stereoscopic view can be built, if and only if we disregard their author's intentions, and appropriate their accounts as I have done in the fourth chapter. Rödl's true materialism can be adhered to by the neuroscientist, if she would uphold that she is investigating material knowledge that is knowledge of oneself as oneself. The neuroscientist might be able to improve Rödl's account by better defining the matter that is self-conscious, which would be quite helpful as Rödl does not really explain what this matter exactly is or how it functions. Kant's transcendental idealism could be adhered to by the neuroscientist, if she would uphold that the determining self, which provides us with self-conscious, cannot be an object of thought as it precedes thought, and that not much more can be said about it other than the things that Kant has when it comes to its functioning as a fundamental aspect of human cognition. However, she can empirically investigate that part of the body which houses the function of self-consciousness to try to better determine its physical location, whilst maintaining that she is limited in her sensory and cognitive capabilities and shall never truly understand self-consciousness as a *Ding*

an sich. Furthermore, a transcendental idealist is an empirical realist, and it is therefore that a Kantian-inspired stereoscopic view might prove to be fruitful, despite the seeming paradoxicality regarding the transcendental and the empirical aspects of self-consciousness.

Certain manifest aspects of self-consciousness are necessary conditions for understanding the possibility of human cognition in general, and these ideas can function as philosophical principles upon which a neuroscientific and stereoscopic image of self-consciousness can be built. The neuroscientist should decide which philosophical position starting point she uses in any form of research on self-consciousness, but I do believe that the Rödlian and Kantian syntheses that I have presented in the last chapter are viable options for working towards a fruitful stereoscopic view of self-in-the-world. The philosopher should decide whether she adapts the results of neuroscientific investigation into her research on self-consciousness, but I do believe that she could strengthen her account with such knowledge. Neuroscientists and philosophers should embrace each other's writings, and work together towards a stereoscopic view of self-consciousness.

Bibliography

Anscombe, G.E.M. “The First Person.” In *Mind & Language: Wolfson College Lectures 1974*, edited by Samuel Guttenplan, 45-65. Oxford: Clarendon Press, 1975.

Bennet, M.R. and Hacker, P.M.S. *Philosophical Foundations of Neuroscience.* Hoboken: Wiley, 2003.

Carroll, Sean B. “Genetics and the making of Homo sapiens.” *Nature* 422 (2003): 849-857.

Castro-Fornieles, Josefina, Bargalló, Nuria, Calvo, Anna, Arango, Celso, Baeza, Immaculada, Gonzalez-Pinto, Ana, Parellada, Mara, Graell, Montserrat, Moreno, Carmen, Otero, Soraya, Janssen, Joost, Rapado-Castro, Marta, and de la Serna, Elena. “Gray matter changes and cognitive predictors of 2-year follow-up abnormalities in early-onset first-episode psychosis.” *European Child & Adolescent Psychiatry*. Accessed July 18, 2017. doi:10.1007/s00787-017-1013-z.

Damasio, Antonio. *The Feeling of What Happens: Body, Emotion and the Making of Consciousness*. London: Vintage, 2000.

Davidson, Donald. “How is weakness of the will possible?.” In *Essays on Actions and Events*, 21-42. Oxford: Oxford University Press, 2001 [1980].

Decety, Jean and Sommerville, Jessica A. “Shared representations between self and other: a social cognitive neuroscience view.” *Trends in Cognitive Sciences* 7:12 (2003): 527-533.

Edelman, Gerald M. and Tononi, Giulio. *A Universe of Consciousness: How Matter Becomes Imagination*. New York: Basic Books, 2000.

Fletcher, Anthony. *Gender, Sex, & Subordination in England 1500-1800*. New Haven: Yale University Press, 1999.

Gallagher, Shaun. “Philosophical conceptions of the self: implications for cognitive science.” *Trends in Cognitive Sciences* 4:1 (2000): 14-21.

Gallup, Gordon G. Jr. “Chimpanzees: Self-Recognition.” *Science* 167:3914 (1970): 86-87.

Kant, Immanuel. *Critique of Pure Reason*. Translated by Paul Guyer and Allen W. Wood. Cambridge: Cambridge University Press, 2000 [1781].

Lycan, William G. and Pappas, George S. "What is eliminative materialism?." *Australasian Journal of Philosophy* 50:2 (1972): 149-159.

Park, Hyeong-Dong, Bernasconi, Fosco, Bello-Ruiz, Javier, Pfeiffer, Christian, Salomon, Roy and Blanke, Olaf. "Transient Modulations of Neural Responses to Heartbeats Covary with Bodily Self-Consciousness." *The Journal of Neuroscience* 36(32) (2016): 8453-8460.

Rödl, Sebastian. "Education and Autonomy." *Journal of Philosophy of Education* 50:1 (2016): 84-97.

Rödl, Sebastian. *Self-Consciousness*. Cambridge, MA: Harvard University Press, 2007.

Seed, Amanda and Tomasello, Michael. "Primate Cognition." *Topics in Cognitive Science* 2 (2010): 407-419.

Sellars, Wilfrid. *Science, Perception, and Reality*. London: Routledge and Kegan Paul Ltd, 1963.

Smith, Adrian J.T. "Comment: Minimal Conditions for the Simplest Form of Self-Consciousness." In *The Embodied Self: Dimensions, Coherence, and Disorders*, edited by Thomas Fuchs, Heribert C. Sattel, and Peter Henningsen, 35-42. Stuttgart: Schatteuer, 2010.

Vogeley, K., May, M., Ritzl, A., Falkai, P., Zilles, K. and Fink, G.R. "Neural Correlates of First-Person Perspective as One Constituent of Human Self-Consciousness." *Journal of Cognitive Neuroscience* 16:5 (2004): 817-827.

Websites

DiPatri, Arthur J. Jr. and Alden, Tord D. "Surgical Treatment of Epilepsy in Children." *Lurie Children's Hospital*. From *The Child's Doctor*, Spring 2005. Accessed January 26, 2017. https://www2.luriechildrens.org/ce/online/article.aspx?articleID=103.

McCormick, Matt. “Immanuel Kant: Metaphysics.” *The Internet Encyclopedia of Philosophy*. Accessed March 3, 2017. http://www.iep.utm.edu/kantmeta/#H4.

Ramsey, William. “Eliminative Materialism.” *Stanford Encyclopedia of Philosophy*. Accessed July 31, 2017. https://plato.stanford.edu/entries/materialism-eliminative/.