From Body to Self - Towards a Socially Enacted Autonomy
With Implications for Locked-in Syndrome and Schizophrenia

PhD Thesis

Miriam Kyselo

submitted August, 2012
defended November, 2012
summa cum laude
Institute of Cognitive Science, FB 08
University of Osnabrück

Reviewers:
Sven Walter, Ezequiel Di Paolo, Evan Thompson

Contents

Abstract

Embodied approaches to cognition consider themselves as alternatives to a brain-bound view of cognition. They decisively emphasize that the brain is not the minimal basis for cognition, but that the body plays a crucial role as well. But what do we actually mean by "the body" and to what extent is it a necessary condition for cognition? Is bodily action equated with movement? Is the human body just a biological phenomenon? How is it related to the human self and sociality?

This thesis explores these questions by confronting embodied cognitive science with Locked-in Syndrome (LIS), a case of global paralysis, which despite the lack of voluntarily bodily action seems to leave the patient cognitively intact. I suggest that LIS poses a challenge to embodied cognitive science putting into question our basic assumptions on what it means to be a human cognitive system. A body without movement and a self whose connection to the social sphere is radically impoverished – how can we make sense of this? LIS challenges the concepts by which we describe the structure and various dimensions of cognition and it invites us to make explicit the background epistemology and general perspective through which we relate the different aspects of cognition.

First, I provide an overview of the philosophy of cognitive science, from the orthodox perspective up to recent embodied cognitive science. I then introduce and clarify the enactive approach, an integrative framework for cognitive science that also serves as the epistemological basis of this thesis. Based on the different states of LIS I formulate a challenge to embodied cognitive science and discuss how the sensorimotor, functionalist and phenomenological approach to embodiment account for it.

The discussion casts doubt on the assumption that a body has mainly to do with movement and it reposes the question how tool-use figures in cognition. It also brings to attention the dimension of bodily subjectivity and raises a much-neglected issue in recent cognitive science: the role of the body in social interactions. I show that these approaches to embodiment entail restrictive or loose notions of the body and are not fully able to account for cognition in LIS.

I formulate a proposal for an enactive concept of the body integrating aspects from the sensorimotor and phenomenological approach to the body. I defend the idea that the enactive approach is the best framework in embodied cognitive science to counter the challenge posed by LIS and BCI. However, since embodied cognitive science entails an individualistic perspective not fully taking into account that humans are embedded in a social environment the question how the body matters in social interaction can also not be resolved from an enactive perspective on the body.

In the last part of this thesis I thus propose transcending the level of individual embodiment. I make suggestions for elaborating on the enactive notion for the cognitive system (autonomy) from a social perspective. I propose to conceive of human mind in terms of a network that is based on the enaction of social processes of distinction and participation. Based on this notion I show how we arrive at an understanding of the human cognitive system which could ultimately account for the basic challenge posited by LIS – the clarification and interrelation of the concepts of body, self and sociality. In the last chapter I provide support for the plausibility for this proposal by applying it to another empirical context, namely psychiatry. What we think about the nature of human mind sets the ground for our thinking about breakdowns and what happens in cases when it does not work. I explore possible implications of the concept of socially enacted human autonomy for mental disorders in general, and for schizophrenia in particular.

Preface

In 2005, during my philosophy studies in Germany, I got to know the work of Thomas Metzinger who was at that time the philosopher in Germany known to combine philosophy with the cognitive sciences. Since I was interested in the philosophy of mind and its linkages to the natural sciences and practice I wrote to him to get some advice on where or how to continue my research endeavors. He invited me to join the MIND group, an international group of young brilliant researchers coming from both sides – philosophy and cognitive science. Becoming a member of that group laid the ground for my academic career and for the topic of this PhD thesis. I am very grateful to Metzinger and the members of the group for giving me the opportunity to connect with the post-graduate academic world, meeting fantastic researchers from all over the world and sharing my ideas from early on.

At one of the MIND meetings I met Mike Beaton, thanks to whom I got to know about the M.A. program in the philosophy of cognitive science at the University of Sussex in Brighton (UK) that I joined in 2006. It was during my studies at Sussex that I first came in contact with enactive and extended theories of cognition. My first reading of *The Embodied Mind* by Varela, Thompson and Rosch made a great impact on my way of thinking about cognition and it inspires me to this day. It was at about that time, at another meeting of the MIND group in Frankfurt, that I heard about patients with complete locked-in syndrome (LIS). The idea that there are people trapped in an immovable body, their eyes forever closed with no way of expressing themselves shocked me deeply. But it also immediately raised my research interest. I wondered how a theory that assumes that cognition is the result of an ongoing embodied engagement with world would account for locked-in syndrome – given that these patients could not move. Should enactivism not imply that they cease to be cognitive? I wrote a term-paper in which I began to explore this question. It was entitled "No action, no consciousness? Challenging the enactive approach to cognition". Clearly, the present thesis mirrors this question in essence. However, it took a detour too to get back to it. During my studies at Sussex I was also intrigued by Andy Clark's work on extended cognition and in my research proposal for the PhD I attempted to combine it with the enactive approach to cognitive science. While I took enactivism to provide the basis for a dynamical approach to the self, I was convinced that extended cognition could serve as a lens for understanding language and tool-use.

With the goal of developing an extended and enacted approach to the self and applying it to language-based technologies I thus began to work as a research assistant with Sven Walter at the Institute of Cognitive Science in Osnabrück. In 2008, I started my PhD under his supervision. I am deeply grateful to Sven for accepting my proposal, which for philosophy in Germany at that time seemed rather exotic if not crazy. No one seemed to have heard about or discussed extended cognition, let alone enactivism. Sven Walter gave me the opportunity and the freedom to explore both extended and enactive approaches to cognition by letting me co-teach a lecture on extended cognition and hold a seminar on enactivism in the summer and winter semester of 2008 and 2009. We had many fruitful and intense discussions and it was a great pleasure to see some of them quickly lead to co-authored publications.

I learned a lot from Sven and I admire him for his rigorous, sharp and critical mind but also for his pragmatic and down-to-earth attitude that I found supportive on many occasions. In this vein I would also like to thank him for his feedback on earlier versions of this thesis, for having helped me find financial support for my project and for remaining supportive even after I changed the focus of my PhD topic in turning more properly towards enactivism.

It was in summer of 2008 at the workshop on enactivism and social cognition in Battle (UK) that I began a frequent dialogue and discussion of enactive ideas with people from the Sussex enactive group. Their work has inspired me to come back to my original research question of how recent cognitive science can account for the impact of LIS on a patient's life. They have helped me to foster my own perspective on cognition and thus the (enactive) grounding for my thesis.

I am very grateful to Ezequiel Di Paolo for having accepted to be my second supervisor. Ezequiel has been a great mentor and this thesis would not be what it is without him. He helped me shape and deepen my thinking. I would like to thank him for countless inspiring and enactive discussions, his detailed and continuous feedback and his kind encouragements. I admire Ezequiel's visionary mind, his enthusiasm and his ability to get to the essence of the most complex issues. He has been and he still is one of the greatest sources of inspiration for me.

I would like to thank Evan Thompson for accepting to be my external reviewer and for giving me feedback on earlier parts of my thesis. Evan's work has made a big impact on me – it set up my mind for thinking about cognition in the enactive way.

I am deeply indebted to the following people for many insightful discussions, for being an inspiration and for having supported this work by providing continuous feedback and sharp criticism: Hanne De Jaegher, Mike Beaton, Markus Eronen, Bertram Lomfeld, Elena Cuffari, Rudolf Müllan, Marek McGann, Thomas Fuchs, Marieke Rohde, Frank Schuman, Jason Clark and Michael Schüler.

My deep gratitude also goes to Hanne De Jaegher who is a great colleague and important inspiration to my thinking. Hanne's encouragements and advice have helped me many times.

I want to thank the following people from the Institute of Cognitive Science in Osnabrück: Peter Bosch and Carla Umbach, the directors of the PhD program, for their support and for encouraging exchange and discussion among the PhD students, Beate Eibisch and Marion Schmitz for their help in organizational and administrative terms, Kerrin Jacobs, Sascha Alexejenko, Konstantin Todorov, Ilaria Serafini, Ulas Türkmen and Saskia Nagel for being great colleagues and friends.

I could not imagine spending the years and especially the last months of my PhD without my friends. I am grateful for their patience, their love and moral support and for having put up with my weirdness during the thesis writing process but also in general. The essence of this PhD also relies on what I have learned in my relations with them. I would like to thank especially Katja Knodel for being a loving friend and for being there whenever I lose track. You help me see the world when I cannot do it on my own.

I would like to also express in detail why I am grateful to my other friends, but I prefer to tell you that in person and here just say that I would not be the person I am and that I would have been not able to write this thesis without you: Franziska Hillig, Priska Herger, Hanne De Jaegher, Nadine Kretzschmar, Michael Schüler, Jean-David Boucher, Stephanie Stein, Paul Hofmann, Johannes von Billerbeck, Joscha Bach, Lea Christiansen, Melanie Köhn, Loni Huber and Matthias Vincent.

My PhD studies have been financially supported by generous grants from the research training group "Adaptivity in Hybrid Cognitive Systems", University of Osnabrück (2009-2011) and the Marie Curie Initial Training Network, "TESIS: Towards an Embodied Science of InterSubjectivity" (FP7-PEOPLE-2010-ITN, 264828) (since 2012). The PhD program in Cognitive Science has provided funding for participation in many conferences and workshops. I am grateful.

Thanks to everybody who supported me but I forgot to mention here.

Introduction

> [I]t is not enough to say that the mind is embodied; one must say how. (Edelman, 1992, p.15)

The movie "The Diving Bell and the Butterfly" tells the story of the journalist Dominique Bauby whose life changes dramatically as he suffers a stroke in a car accident. The stroke causes a global paralysis of his body, leaving him unable to move except for the blinking of his left eye. Bauby is diagnosed with Locked-in Syndrome (LIS), a neurological condition in which a patient's body is severely paralyzed, inhibiting voluntary muscle movement. LIS almost completely prevents him from communicating verbally and non-verbally, placing a distance between himself and his social world.

The movie and book are based on a true story written by Bauby himself. It is the report of his experiences with the condition, a creative mind, a butterfly, locked into his rigid body, the diving bell. Bauby selected, with the blink of his eye, letters from the alphabet that was continuously read to him by an assistant and thus composed letter by letter and word by word a fascinating and harrowing testimonial of what it is like to be locked-in. It allowed him to cross the abyss that his rigid body had ruptured between him and the world of others. The condition had affected almost his entire body and therewith his abilities to move and to interact bodily with the world. His former self was deeply affected and he found himself at a distance to the people surrounding him. His body was no longer available to him while at the same time his creativity and the will to express and show himself to others remained pristine or grew even stronger.

The present dissertation is explicitly and implicitly guided by the case of Bauby and other persons with LIS. It is motivated by the assumption that the radical character of LIS urges us to look more closely at aspects of our existence we have always taken for granted or thought to have understood thoroughly. LIS should be considered of crucial importance for cognitive science. It poses a general challenge because it puts into question our basic assumptions on what it means to be a human cognitive system. A body without movement and a self whose connection to the social sphere is radically impoverished – how can we make sense of this? LIS challenges the concepts by which we describe the structure and various dimensions of cognition and it invites us to make explicit the background epistemology and general perspective through which we relate the different aspects of mind.

In this thesis I outline various ways in which cases of global paralysis are relevant for cognitive science, but I focus in particular on the question of how they might help to reassess our understanding of the body and what it means to have a body in order to act in the world. Recent frameworks in cognitive science generally labelled as "the embodied approaches" are considered as alternatives to a brain-bound view of cognition. They decisively emphasize that the brain is not the minimal basis for cognition, but that the body plays a crucial role as well. But how does an embodied cognitive science account for a case of global paralysis, which despite the lack of voluntarily bodily action seems to leave the patient cognitively intact? What do we actually mean by "the body" and to what extent is it a necessary condition for cognition? Does it only relate to our capacities for movement and action? What is meant by "action" and does it have to be realized by a body in movement? Is the human body just a biological phenomenon?

Related questions arise when we consider the impact of the paralysis on the patient's self. At least we intuitively grasp the personal drama that the loss of one's bodily movements must cause the person. But what is the role of the body for one's self to begin with, what would be the

impact of an impaired body on the self and what role might facilitating technologies, such as brain-computer interfaces (BCI), play in this context?

One of the most terrifying aspects of LIS is that due to the severe restrictions that can even lead to a complete cessation of muscular movement the patient is disconnected from the world of others. This raises a pressing issue so far widely neglected in cognitive science. Humans are not isolated subjects, but deeply embedded in a social environment, constantly having to negotiate our existence in a world of other people. How does bodily impairment affect us as social beings? How can embodied cognitive science accommodate that the fact that humans live not only in a world of objects but in a social world? How are self and body affected by sociality?

These are the leading questions that motivate the present thesis.

One of my main assumptions is that addressing these questions and thus accounting for the challenge posed by cases of LIS requires a holistic framework for mind, i.e. a framework that offers a basic background theory for cognition and allows us to intelligibly synthesize different strands and methodologies of researching cognitive phenomena. As I will defend throughout the thesis, I take the so-called enactive approach to the mind as the most promising candidate for providing such a framework. Throughout the thesis I draw on the enactive methodology to explore my main hypothesis, which is that the role of the body in human cognition can ultimately be assessed only against the background assumption that human beings are primarily social systems that do not only engage with a world of material objects, but interact with a world of other subjects.

This thesis provides a new look at the human mind bringing to light an important issue for embodied cognitive science: the body-body social problem, i.e. the question how embodiment is affected by sociality. It uncovers a way in which this problem can be resolved in terms of enactivism and it thereby makes three major conceptual contributions. Firstly, it refines and re-conceptualizes the notion of bodily action and offers an enactive proposal for the body that integrates other recent approaches to embodiment. Secondly, a new approach to human cognition in terms of socially enacted autonomy. At the heart of this proposal lies a conceptual innovation – the notion of the social processes of participation and distinction – that captures the role of sociality for human beings along various dimensions, such as normativity and consciousness. Thirdly, I outline an enactive approach to the self that sheds new light on the impact of bodily impairment, such as LIS, and also offers a novel explanatory tool for the assessment of mental disorders.

Before offering an outline of this thesis and its contributions in more detail I would like to make some general remarks concerning the kind of thesis it is, and the strategy and methodology that I adopt in order to address the issues raised above. These are important for understanding my research goals and what the reader can or cannot expect from the following chapters.

I should say that this is not a typical philosophy (of mind) thesis. While it deals with concepts and questions that are clearly some of the most widely and intensively discussed within this field, it does not approach these questions in the way that most philosophers –in particular those of the analytic tradition – would probably do. I acknowledge the value of philosophical arguments *per se*, and the beauty of neat and logical reasoning. However, I am not happy to believe that this is all that philosophy is ultimately about. I do not think that philosophy is in essence a hermeneutic discipline, mainly concerned with the exegesis of previous texts and arguments. Philosophy must be freed from the dust of its own unfolding and be applied, put to practice – in life and in sciences – where and whenever conceptual groundwork, a critical eye and word, new questions or avenues are required. For this reason, it is part of my methodology to relate to empirical contexts and contribute to setting up a new conceptual framework with practical applicability. This thesis is a philosophy thesis, but it is a thesis in philosophy of cognitive science.

From my point of view, philosophy of cognitive science clearly belongs to cognitive science. Yet, it should not only be seen as one discipline among the others that constitute

cognitive science, but rather as a servant to these sciences in that it offers and develops a theoretical and conceptual background; and so it attempts to clarify confusions, provide criticism, elaborate conceptual tools and raise genuine and new questions for research. Cognitive science is interdisciplinary, but that does not mean that the concepts and models used in its various fields are interdisciplinarily shared or even coherent. Philosophy of cognitive science can be a mediator that facilitates the dialogues between the different disciplines by making suggestions towards a coherent conceptual framework in which to embed them. Philosophy of mind may be inspired by cognitive science, but it seems that the value of incorporating empirical findings of other disciplines is linked to an intrinsic purpose only. For neither is the philosophy of mind expecting cognitive science to necessarily guide its lines of reasoning nor need philosophical arguments in turn play a role for cognitive science. I take it that philosophy of cognitive science, in contrast, is necessarily affected by the other (cognitive) sciences and must consider its own theories, concepts and reasoning as ultimately connected to them. This is why philosophy of cognitive science needs to be different: on one side it should always seek to apply its concepts to concrete, empirical aspects, but on the other it should be ready to freely and reasonably choose from the rich repertoire of concepts and tools offered by cognitive science, by any of its subfields as well as of other useful approaches, such as phenomenology, social psychology, and even the fine arts. Philosophers of cognitive science should be encouraged to choose those tools and concepts that can be of use in order to advance, elaborate and develop aspects of coherent background theories for cognitive science.

Because theories in philosophy of cognitive science are not closed frameworks but intimately connected to cognitive science, they have to stay in constant development and require a readiness for exchange and open dialogue. In my view the conceptual framework that most successfully embodies such a philosophy at the moment is enactivism. The basic idea of enactivism is that cognition is grounded in life and that cognitive systems are self-organized and self-maintaining autonomous systems that actively engage with the world. In its general commitment to the mutuality between 1st and 3rd person perspectives – i.e., the subjective point of view of the cognitive agent and the agent in action scrutinized externally – enactivism reflects the will to encounter both the subjective and the organizational and mechanistic aspects of the mind. It does not shy away from integrating aspects and methodologies that were traditionally shunned or at least widely ignored, but remains vital, interested and ready to learn. It welcomes insights from various disciplines and traditions including biology, neuroscience, psychology, anthropology and sociology. Enactivism is in constant development but already provides a basic framework, with concrete suggestions as to how cognition can be defined and how cognitive science should pursue researching it. I should clearly say that I am a proponent of enactivism and the present thesis can be generally seen as an attempt for both a clarification and elaboration on it.

This thesis makes precise and original contributions to the field of embodied and enactive cognitive science. It sheds new light on the question of what cognition is at the level of human beings. The thesis illuminates human cognition by bringing together issues that have not been sufficiently acknowledged before. Its general value as a contribution to conceptual groundwork for the field of cognitive science is also that it raises novel questions and makes various suggestions for new avenues to pursue them. Throughout this thesis I use LIS as an empirical touchstone against which particular concepts are discussed. This applies for both my own considerations and those of other people working in the field. The idea behind this strategy is to get a fresh look at concepts of both body and self, and to ground them by connecting them to phenomena of the real human world. In accordance with the enactive approach I believe that cognitive systems cannot be fully understood without considering what being a cognitive system means from the perspective of the cognitive system itself. As another part of my methodology I explore this perspective by way of self-reflexion and by providing concrete phenomenological observations.

The general outline for the present thesis is as follows: in the first chapter I provide an introduction into the philosophy of cognitive science, from the orthodox perspective up to recent embodied cognitive science. This is not only a review but also invites the reader to focus on particular issues (such as the role of the body, the relation between cognition and environment as well as the issue of consciousness) that come up as important questions to be addressed throughout the thesis.

In Chapter 2 I continue this outline of the theoretical background of my thesis by offering a summary of the enactive approach to cognitive science. This introduction also provides important clarifications of the enactive approach: I discuss not only how enactivism relates to the orthodox view of cognition, but also compare it to other recent alternative views, such as the extended or sensorimotor approache, and outline general ways in which it departs from them. The second chapter has particular relevance for my argument because it provides the conceptual ground and methodology that at the same time also guides the attempt to answer the challenge posed by LIS. It is also the target of my elaborations, in particular with regards to the concept of the body and the cognitive system in social interaction.

In the third chapter I explicate the general challenge posed by LIS. I approach this in two steps. First, I provide basic (neuromedical) information about LIS and the techniques that facilitate or enable the patient's abilities, in particular so-called Brain-Computer Interfaces (BCIs). Then I introduce and discuss recent philosophical approaches to LIS and other cases of bodily paralysis. On the one hand, I show that these approaches inspire our thinking about human cognition as they connect bodily impairment with questions of identity (self) and sociality. However, their use of these concepts remains unclear and the possible linkage between these concepts at best implicit. On the other hand, I show that philosophy of cognitive science has important relevance for the empirical and medical contexts. The way in which we address particular questions in these contexts is directly affected by the kind of background epistemology we adopt in order to account for cognition and they can have immense practical relevance or ethical import. I thus show that LIS poses a two-fold basic challenge for cognitive science, with regards to the concepts used to explain human cognition but also with regards to the way in which we take them to interrelate.

Based on the different states of LIS in combination with BCI technology I formulate in Chapter 4 a particular version of this general challenge, namely a challenge to embodied cognitive science and its approach to the body. I introduce four recent approaches to embodiment and discuss how they account for the role of the body for cognition. This brings to light particular intuitions that embodied approaches to cognitive science have about the nature of the body. I introduce and discuss the sensorimotor and functionalist approach to embodiment and show that the former restricts the body to movement or biological structure and that the latter has an inflationary conception of the body. Neither of them takes the body to be relevant beyond developmental purposes. I show that these conceptions of the body cannot explain why patients who do not move are practically cognitively intact. The discussion casts doubt on the assumption that a body has mainly to do with movement and it poses again the question how tool-use figures in cognition.

Chapter 5 turns to a phenomenological account of the body. I show how embodied cognitive science can be informed by phenomenology in linking the body to consciousness. Based on Merleau-Ponty's approach to the body I shed light on the ability to appropriate tools in order to realize our goals. By introducing Henry's thought on self-affection I then bring to attention an additional dimension of bodily subjectivity, i.e. that of the body in relation to itself. The discussion of the phenomenological concept of the body in the context of LIS and BCIs offers a refinement for the concepts of body and bodily activity based on which we can account for cognition in LIS in a way unavailable to the sensorimotor or functionalist approach. However, it also leads to a further question for embodied cognitive science, viz. how the dimensions of world- and self-directed experience are integrated in the subjective body. It also

raises for the first time in the thesis a much-neglected issue in recent cognitive science: the role of the body in social interactions.

In Chapter 6 I explore the refined LIS challenges from an enactive perspective. I show that the problem of relating world- and self-directed experiences translates to a refined version of the so-called body-body problem. I offer a solution to this problem by exploring linkages between the phenomenological and enactive approach to the body and outlining how bodily subjectivity can be grounded in autonomy. I also make suggestions about how aspects of the sensorimotor and functionalist and phenomenological approach can be informed and integrated from an enactive point of view. Based on this integration I formulate a proposal for an enactive concept of the body to this day missing in an explicit formulation in the enactive literature. I defend the idea that this enactive concept can counter the challenge posed by LIS and BCI to embodied cognitive science. However, it turns out that the question of how the body matters in social interaction cannot be resolved from the perspective on the body because it implies an individualistic perspective on the cognitive system. I show that the task of spelling out what cognitive science means by embodiment also has to accommodate the fact that humans are embedded in a social world and to clarify how this affects their body. To capture this issue I introduce a refined version of the body-body problem, which I call the body-body-social problem.

In Chapter 7 I make suggestions for elaborating the enactive notion for the cognitive system (autonomy) from a social perspective. I introduce a new conceptual distinction in the dynamic quality of social processes with respect to autonomy ("distinction and participation"). These notions allow a reconsideration of the relation of identity and sense-making at a level of autonomy that is socially enacted. Based on this, I shed new light on the interrelation of the enactive concepts of operational closure, normativity and bodily experience for the case of human beings. This proposal of socially enacted autonomy makes another important contribution to enactive cognitive science. It provides a novel way to situate human beings at the continuum of life and mind in terms of social mediation.

Based on the enactive approach to social autonomy I offer a solution to the body-body-social problem and arrive at an understanding of the human cognitive system which could ultimately account for the basic challenge posited by LIS – the clarification and interrelation of the concepts of body, self and sociality. These considerations contribute to a better understanding of the impact of bodily impairment on a patient's life and the unexpected results of quality of life studies in LIS, the role of facilitation technologies like BCI and they have practical and ethical implications for improving the situation of patients with bodily restrictions.

In the last chapter I leave considerations of bodily impairment in terms of paralysis behind and test the newly developed concepts in another different domain, namely psychiatry to see whether they have some promise. What we think about the nature of human cognitive systems sets the ground for our thinking about breakdowns and what happens in cases when cognition does not work. I thus explore a possible implication of the concept of social human autonomy for mental disorders in general, and for schizophrenia in particular. I outline a new explanatory basis for understanding both the genesis and the phenomenology of disorders and contrast this with a recent interpretation of schizophrenia in terms of a disorder of the minimal self. I illustrate for bodily alienation and hyperreflexivity – two classical symptoms in schizophrenia – that the enactive perspective of the self allows for a new interpretation of schizophrenia.

Note that while this thesis pursues a clear goal – a proposal of an enactive notion of the body and indications for elaborating the enactive approach with regards to the self – these indications should by no means be conceived in all cases of as fully defended positions. This thesis does not attempt to explicitly and simultaneously give all the definitive answer to the questions it raises, and some aspects call for more extensive discussion. In particular the last two chapters are exploratory in nature. I explicitly acknowledge this – it is not only a necessary but

also an intended side effect of a work that aims to support a new direction of research in cognitive science.

I would like to illustrate the tenor of this thesis by using the analogy of the arch and the way it is constructed. In order to construct an arch we need to already have an idea of what it looks like. This idea guides our construction. We then do not start by piling up stones, hoping they will hold position and that the whole construction will not collapse. Rather, we make a sketch of the arch, and then we build a scaffolding that already reflects its overall shape (a method called "centring"). Finally, we place the stones on this scaffolding and arrange them in line with it.

In this vein, I am approaching research on cognition. We need a holistic perspective, and a big picture to understand it. The scaffolding for this is constructed in terms of the enactive framework. Based on this we can explore and interrelate the various concepts and phenomena of cognition.

Figure 1. Arch centring and a carpenter's workshop. E&E Image Library/Heritage-Images

Chapter 1.

Cognitive Science – From Cognitivism to Embodiment

> To deny the truth of our own experience in the scientific study of ourselves is not only unsatisfactory; it is to render the scientific study of ourselves without a subject matter. (Varela et al.1993, p. 13)

In the present and the following chapter I provide a historical overview over the development of (the philosophy of) cognitive science, capturing its beginning in the 1970s (cognitivism and connectionism), the shift towards a more body-centred perspective (embodied cognition) in the 1990s (Chapter 1) and four more recent strands of research in embodied cognitive science: the phenomenological, sensorimotor and functionalist approach (Chapter 1) and the enactive approach (Chapter 2). This overview constitutes the first part of the theoretical and empirical background of my thesis (the second part is provided in Chapter 3 where I will introduce LIS and BCI as well as several approaches to the role of bodily impairment for a patient's life). It consists of four major sections in which, for each of these approaches, I will attempt to highlight and compare their respective views on

1) cognitive processes,
2) the nature of the cognitive system,
3) their perspective on the relation between the cognitive system and the world (including their stance on the status of internal representation) and
4) the treatment of subjective experience.

All four approaches, i.e. the sensorimotor, functionalist, phenomenological and enactive approach are then further refined and extensively discussed in the context of LIS and BCI in the subsequent Chapters 4–6.

1.1 The Orthodox View of Cognition

The study of the mind has a very long tradition, but it was only in the beginning of the 20th century that mind became a subject of methodological exploration in the experimental psychology of Wundt and his colleagues (Stevens, 1951). Shortly after this, psychology became dominated by *behaviorism*. Proponents of behaviorism denied that the mind could be a subject of scientific study. What a person thinks or desires remains private – psychological events should be explained in terms of behavior, not in terms of internal mental states. The mind became a "black box".

Cognitive science can be seen as a response to this behaviorist disregard of the mental. Its origins fall together with the beginnings of the development of *artificial intelligence*, in the mid 1950s, the strong emphasis on formal logic and the advances in computer technology (Clark, 2001).

1.1.1 Cognitivism

The core assumption of early cognitive science was that the mind is an information-processing machine, a physical symbol system, which follows logical rules (Newell and Simon, 1963). This view of cognition is now famously known as the "good-old fashioned artificial intelligence" (GOFAI) approach (Haugeland, 1985). According to GOFAI, the best model for understanding the nature and functioning of the mind is the computer (aspect 2).

The two key hypotheses of this computer metaphor of the mind are, first, that mind consists of internal mental states which are best described as symbolic representations that (like natural and artificial languages) have syntactic and semantic properties and, second, that thinking itself should be seen as a computational process (aspect 1). While computers manipulate rigid symbols according to pre-defined algorithms, mental processes are rule-based manipulations of symbolic representations in the brain. The brain functions as the hardware on which the mind is running like a software program. Since software programs are designed to resolve a problem within a particular domain, thinking processes were, in the main, considered to consist of abstract problem solving processes governed by a specific set of rules.

The computational metaphor enabled scientists to connect behavior with inner mental states and explain intelligent behavior based on the assumption that these internal states represent facts of the external environment, effectively opening the "black box" that behaviorism took to be outside the reach of science. Given a specific problem, these representations are manipulated in a way so as to produce novel representations that finally lead to adequate, solution-directed, behavioral output. Actions of cognitive systems are causally explained in terms of mental states such as beliefs and desires. Behavior is based on particular mental states and their causal interrelations, which can be described in functional terms. Internal mental states like beliefs and desires were also seen as the appropriate background theory to account for our commonsense use of language. Language works the way it does because it refers to mental states in the brain that have a similar structure (Fodor, 1975).

An implication of the functionalism inherent in early cognitive science is that in order for a cognitive system to perform abstract reasoning, no specific material implementation was required. The mental is not unique to humans or to other animals with brains. Alien creatures could have a physical make-up entirely different from ours and they would still count as cognitive beings (Lewis, 1980). Mental states could potentially even be realized by cheese (Putnam, 1975). What mattered to cognition was that the formal organization of the system in question served a specific functional profile. This liberalism with regard to the material basis of the cognitive has become known as the multi-realizability claim. Due do this claim the role of the biological body was downplayed; it was not regarded a subject worthy of consideration for the scientific understanding of the mind. Cognition in early cognitive science was a matter of computational architectures in the head and thus "radically disembodied" (Drayson, 2009, p. 331). There is a clear-cut distinction between the body and the cognitive system.

A similarly clear-cut distinction can be observed with regards to the relation between the cognitive system and its environment (aspect 3). This distinction is another consequence of the computer analogy. Since computers are input-output devices, cognitive systems receive, via their sensorimotor sensors, subsymbolic input from their environment, which is then transformed into symbolic representations and processed so as to generate an appropriate action output. This separation between the cognitive system and the external world has been criticized as the "sandwich-model" of cognition (Hurley, 1998). This model adopts a form of realism according to which there exists an objectively given world outside of the cognitive system. The world presents problems that the cognitive system would have to solve and if the cognitive system found the right solution to a specific problem, then it could be said to function, i.e. to represent the world accurately (Varela et al., 1993).

With the nature of the cognitive defined as computational processing, the mind became a tractable subject of not only philosophical and psychological but interdisciplinary scientific

inquiry in which mind could be studied from a mechanistic or engineering perspective (Clark, 2001). It provided early cognitive science with a unified story about the nature of cognition.

But this story was limited. Dreyfus has raised some of the earliest objections to GOFAI (Dreyfus, 1972, but see also Searle, 1980, Brooks, 1991). He argued that human cognition could never be accounted for in terms of the computer metaphor. First, the human brain does not function the way the computer does. Second, human cognition is based on different sense-modalities (auditory and visual) that are processed as integrated wholes rather than as discrete bits of information. Third, in contrast to computers, humans can flexibly adapt their behavior in case of a breakdown. Fourth, human cognition does not consist of independent facts and bits of information, but is always embedded in a context. A computer, however, cannot distinguish relevant from irrelevant information. This has been also referred to as the so-called "frame problem" (McCarthy and Hayes, 1969, Wheeler, 2008).

Another major shortcoming of GOFAI was its entire neglect of the role of subjective experience (aspect 4). Cognitive processes were equated with the mechanism that manipulates symbols according to specific rules, but subjective experiences had no causal power in this mechanism. Since cognitive processes were supposed to be entirely unconscious, cognitivism was "banishing consciousness from the science of mind" (Thompson, 2007, p. 5).

1.1.2 Connectionism

In the 1980s, defenders of connectionism challenged some of the central tenets of the leading metaphor of cognitive science. Connectionism suggested that the mind should not be viewed as a computer, but – in an allusion to the workings of the brain – as an artificial neural network.

Connectionism has its roots in early work in cybernetics, in the invention of very simple neuronal models, such as the McCulloch-and-Pitts cells and Rosenblatt's perceptrons (McCulloch and Pitts, 1943, Rosenblatt, 1958). Minsky and Papert (1969) criticized early connectionist work, showing that the computational capacities of artificial neuronal networks (ANNs) were heavily restricted. The revolution of parallel distributed processing in the 1980s has however shown that their criticism is only partly applicable (Rummelhart, Smolensky, McClelland, and Hinton, 1986).

According to connectionism, to conceive of the cognitive system as a neuronal network would offer a much more biologically plausible model of the mind. Information processing in these artificial networks is not seen as digital and based on individual and rigid symbols, but as parallel and distributed. Thinking processes arise dynamically as the result of the activities of a network of interconnected neurons. In contrast to GOFAI, information is stored non-symbolically in the connection weights between different units and their activation.

The beginnings of cognitive science focused on cognitive processes involving problem solving in a particular domain based on a specified goal and according to detailed planning. Connectionism in contrast, acknowledged that the encounters of cognitive systems with the real world do not always unfold according to a pre-specified plan. Artificial neural networks took into account the context sensitivity of the cognitive system, which allowed for a greater flexibility of the artificial models in the face of novel situations.

Despite these differences, cognitive processes were still seen as problem solving processes, realized through the processing of information that is stored in internal representations. While for cognitivism these representations were symbolic and rigid, they were distributed in subsymbolic neuronal networks for connectionism.

Connectionism thus provides a somewhat different approach to the nature of cognitive system and process (aspects 1, 2), but its view of the relation between system and world is not different from the cognitivist view. Connectionism similarly assumes that there is a pre-given and objectively assessable world outside, which presents the cognitive system with problems to solve (aspect 3). Conscious experience still plays no role in connectionism either (aspect 4). Despite the connectionist's emphasis on the neural underpinnings of cognition, connectionism also neglects the role of the "rest" of the body.

Cognitivism and connectionism thus mainly differ with regard to the particularities of how problem solving is realized; yet their commonalities are much stronger. From a more recent perspective on the development of cognitive science both approaches can be treated as one unified, so-called "orthodox" view of cognition (Drayson, 2009, p. 330).

In comparison to the approaches to come, the orthodox approach to cognition appears to tell a story filled with fragmentations and separations. It implies an epistemology, which is *dichotomous*, *exclusive* and *reductive*. It is *dichotomous* because of the separation of brain and body, as well as that of cognitive system and world. It is *exclusive* with regard to what counts as a cognitive process because only very abstract reasoning processes are seen as cognitive processes. Finally it is *reductive*, for in denying conscious experience a place in the cognitive science research endeavor, it shrinks the field of phenomena to be explained to observations of material processes from "the outside".

1.2 Embodiment and Dynamics

In the 1990s the development of cognitive science shifted towards the so-called embodied and dynamical view. The new perspective on cognitive science differed greatly from the traditional one: The constitutive mechanism for cognition was no longer regarded as bound to internal (brain) processes alone, but now included bodily and environmental aspects. The primary target of research on cognition was no longer high-level "offline" reasoning processes, but rather simple and less abstract processes that were time-sensitive and situated in a realistic environment. Finally, consciousness became a proper subject matter for cognitive science.

In contrast to classical cognitive science the new "embodied cognitive science" lacks a coherent background story. The outlined differences to the orthodox view are reflections of the new approaches only in the sense that they are all generally considered as alternatives to it, but not in the sense of representing key features that are linked to each other in an overall background epistemology. There is not *one* unified embodied approach to cognitive science, but only a meshwork of partly overlapping and related, yet also partly incompatible perspectives. In other words, what appears to be characteristic for one approach need not be the case for another. For this reason, the assessment of the four basic aspects introduced above – the definition of cognitive system and cognitive process (aspects 1, 2), the relation between system and environment (aspect 3), and the role of conscious experience (aspect 4) – will reflect aspects of a fragmented collection of characteristics rather than of a conceptually interrelated network.

I begin the overview of the development of the embodied approach*es* with dynamicism and robotics. I will then consider the now classical book *The Embodied Mind* which can be seen as one basis of the development of three major and differing stances on embodied cognitive science: the sensorimotor, the extended functionalist and the phenomenological approaches. None of the approaches is able to account for all of the aspects, let alone to offer clear ideas as to how they connect. The only "embodied" approach to cognitive science that offers a unified background theory seems to be enactivism, which I introduce in the next chapter.

1.2.1 Dynamics and Robotics

With the dynamicist approach the view of the relation between the cognitive system and the environment shifted dramatically (aspect 3). It departed from the sandwich model of cognition. Cognitive systems were no longer regarded as "islands" independent of the environment, but rather as dynamical systems comprising not only neural, but also bodily and environmental features (Van Gelder, 1998). This dynamical view was also adopted in the work of roboticist and

"nouvelle AI" pioneer Rodney Brooks. In contrast to GOFAI's focus on abstract human reasoning processes, the new artificial intelligence movement was concerned with low-level behavior of simpler systems such as insects. Brooks found that in order for a robot to perform intelligent behavior, no central processing unit or explicit inner model of the environment was needed (Brooks, 1991). This was reflected in the construction of robots endowed with a so-called "subsumption architecture". The robot is able to perform intelligent behavior in a real-world environment "as the result of simple interactions between relatively self-contained behavior-producing subsystems" (Clark, 1997, p. 14). A subsumption architecture consists of different control layers that all respond differently to features of the environment; there is no central control which orchestrates the activities of the different layers:

> The idea is to first build a very simple complete autonomous system, and *test it in the real world*. Our favourite example of such a system is a Creature, actually a mobile robot, which avoids hitting things. It senses objects in its immediate vicinity and moves away from them, halting if it senses something in its path. It is still necessary to build this system by decomposing it into parts, but there need be no clear distinction between a "perception subsystem", a "central system" and an "action system". In fact, there may well be two independent channels connecting sensing to action (one for initiating motion, and one for emergency halts), so there is no single place where "perception" delivers a representation of the world in the traditional sense (Brooks, 1991, p. 147).

In contrast to the high-level human reasoning processes, much simpler actions such as finding a way around or collecting cans became the prime examples of a cognitive process. A famous example of the subsumption architecture was Brooks' robot Herbert, who would locate, collect and trash soft-drink cans in an MIT lab (see Clark, 1997, pp. 14–15 for more details). This view brought about a dramatic shift in thinking about what actually counts as a cognitive process (aspect 2). Rather than assuming that cognition is based on a pre-specified set of rules and internal representations of the external world, the characteristics of a cognitive system's embodiment and its capacities to react flexibly to its environment turned out to be of vital importance. This flexibility was considered to be biologically more parsimonious, since, rather than assuming (as in GOFAI) that cognition is general and de-contextualized, the NEW AI-movement acknowledged that animals already exhibit a high degree of intelligent behavior in their ability to exploit and to adapt to their environment. Cognition was no longer confined to the brain, but comprised the embodied whole of very simple (artificial) animals. The sensorimotor periphery of cognitive systems was not the means to provide input that would then be transformed into mental representations (as in the sandwich-view of cognition), but became directly acquainted with a given task, for instance to avoid an obstacle. The processes of perception and action were considered to be intimately connected. The body itself became the means of constantly providing information about the external world, which is always there "to be called upon just in time" (Clark, 2008a, p. 15). At least for so-called online-reasoning, i.e. low-level cognitive processes, minds would have no need for inner representations (Wheeler and Clark, 1999). The mind did not function independent of world and body, but was now seen as being based on the intimate coupling of bodily structures as well as a situated exploitation of environmental features (aspect 1).

1.2.2 "The Embodied Mind" – Consciousness Matters

The Embodied Mind, a book by Varela, Thompson and Rosch (1993) that became a classic for the embodied approach to cognition was inspired by the development in situated robotics. The orthodox view widely neglected the role of consciousness, but the authors of *The Embodied Mind* brought back the focus on the role of the lived experience of cognitive systems (aspect 4). In their eyes the study of cognition would require a complementary approach, which combined the mechanistic and engineering perspective and third-person methods with phenomenological

investigations of conscious experiences from a first-person perspective. Cognitive processes should be assessed on the one hand as behavioral processes, but on the other as experiential processes as well.

The Embodied Mind also adopted the dynamical perspective on the relation between cognitive system and the environment and transcended the "logical geography of inner versus outer" (Varela et al., 1993, p. 172) implied by the orthodox view of cognition. The idea that cognition consists of "the representation of a pregiven world by a pregiven mind" (ibid., p. 9) was entirely rejected. Instead they argued that cognition and cognitive systems' worlds are *enacted*.

"To enact" means "to bring forth" or "to bring about," expressing the conviction of enactivists that cognition requires activity. Cognition does not reside within the cognitive system, but is a relational phenomenon arising through the embodied and active engagement of a cognitive system with its environment (ibid., p. 205). Cognitive processes are also not determined by the outside world, but they depend on the system's embodied structure put into interaction with the environment. They result from "recurrent sensorimotor patterns that enable action to be perceptually guided" (ibid., p. 173). Action and perception are deeply intertwined. The body is not any longer a part of both outside layers of the sandwich, but part and parcel of cognition itself. For this reason we cannot establish a clear separation between the perception of an outer world on the one hand and the behavioral action of the individual system on the other. According to *The Embodied Mind*, the body is seen as a physiological body, as the "context or milieu of cognitive mechanisms", but it is also a subjective body, a "lived, experiential structure" (ibid., p. 238).

1.3 Varieties of Embodiment

The *Embodied Mind* marked the beginning of the development of a variety of other approaches that emphasized the role of body and environment as well as the interactive character of cognition. Some of them clearly count as a continuation and elaboration of the ideas suggested in the book. But other approaches labeled "embodied" were actually departing from these ideas because as we will see below, in fundamental aspects they remained committed to the orthodox convictions.

That there is no coherence in embodied cognitive science has been acknowledged for a while already (e.g. Ziemke, 2003, Anderson, 2003, Rohrer, 2007). A decade ago, Tom Ziemke wrote:

> Embodiment is nowadays by many researchers considered a *conditio sine qua non* for any form of natural or artificial intelligence. Moreover, it is one of the key concepts that distinguishes recent work on situated/embodied/embedded/interactive cognition from the approach of classical cognitive science which, based on functionalism, had its focus on 'disembodied', computation...while many researchers agree that cognition has to be embodied, it is less clear so far what exactly that means. (Ziemke, 2003, p. 1305)

In a similar vein, Wilson said:

> Although this general approach is enjoying increasingly broad support, there is in fact a great deal of diversity in the claims involved and the degree of controversy they attract. (Wilson, 2002, p. 625)

These authors have offered their own attempts at systematizing the field: Wilson (2002) has proposed six major *claims* characteristic of the embodied view on cognition. According to her, cognition is situated, time-pressured and involves off-loading of "cognitive work". It is, furthermore, deeply intertwined with the environment, evolved "for action" and even "off-line cognition is body based" (2002, p. 626). Ziemke (2003) disentangled the field of embodied cognition according to the question "*what kind of body*" is capable of embodied cognition (p. 1305,

my italics) and identified the following five types: (1) *structural coupling* between agent and environment, (2) *historical embodiment* as the result of a history of structural coupling, (3) *physical embodiment*, (4) *'organismoid' embodiment*, i.e. organism-like bodily form (e.g., humanoid robots), (5) *organismic embodiment* of autopoietic, living systems (2003, p. 1306, original emphasis)[1]. Rohrer (2007) identified even twelve different dimensions of embodiment, but argues that these can be usually subsumed under two senses of it, either viewing the body from an experiential perspective or "as the physical substrate" (Rohrer, 2007, p. 359).

In the following introduction I will adopt a combination of these strategies. I will roughly divide the field of embodied cognitive science into four groups where each[2] is considered with regards to both their characteristic claims and what they mean by the body.

The first group is constituted by *phenomenologically* inspired approaches that take the theories and methods of phenomenology and contemplative mindfulness traditions to be vital for research in cognitive science. The second group is represented by the proponents of *sensorimotor* views and comprises approaches emphasizing the non-negotiable role of specific bodily structures, movement and action. The third group is *functionalist* in spirit. It holds that while body and world may be of relevance, what is actually crucial for cognition is that the subconscious vehicles (of whatever kind) realize an adequate functional profile. The fourth approach, to which I am going to dedicate the entire following chapter, is the *enactive* approach.

What is important with respect to the present work is that none of these groups sufficiently explicates what the commitment to "the body" amounts to. The epistemology underlying these approaches appears fragmented, as it is not clear how different aspects of cognition relate to the body. We will come back to this issue in Chapters 4–6; the four above-mentioned approaches to embodiment and their conception of the body are discussed in the context of LIS and BCI.

1.3.1 Phenomenological Approaches to Embodiment

According to the phenomenological approach to embodiment, cognition and consciousness are both shaped by our bodies. The body is the very means by which we are in the world. It should not only be seen as a material object amongst other objects of the external world, but rather also as a *lived body*, i.e. a body with an experienced subjectivity (Thompson, 2007). The living body has two basic components. On the one hand, it is an intentional body as it is directed to the world and the means and basis for all sorts of experiences concerning the external world (Merlau-Ponty, 2002/1945). On the other hand, phenomenologists assume that the body is self-affective in that it already experiences itself where this experience is of a non-intentional nature and independent of and prior to any interaction with the world (Henry, 1988).

The distinction between the body as *Körper* (object) and *Leib* (the lived body), i.e. object and subject body (Fuchs, 2011), has inspired more recent phenomenological work. Gallagher, for example, has proposed that the lived body has a kind of pre-reflective self-awareness, comprising two tightly connected kinds of experiences: the so-called *sense of ownership* which allows us to experience our own body as the source of experiences, and a non-conceptual *sense of agency* informing us about the fact that we are the initiator of a particular bodily action (Gallagher, 2006). The distinction becomes clearer when considering involuntary actions (Gallagher, 2010, p. 27): if another person moves our arm, we certainly feel that it is our own body, which is involved in the movement (sense of ownership). However, we also feel that we are not initiating the movement (sense of agency).

---

[1] Ziemke mentions "social embodiment" as a sixth notion but does not consider it to be a type of embodiment (Ziemke, 2003, p. 1309).

[2] The first three groups may comprise several approaches, amongst which, of course, may exist more subtle differences. But for the present purpose it will suffice to consider the most salient differences and refrain from a more fine-grained analysis.

These senses are two aspects of the so-called *minimal self*, a pre-reflective bodily experience of "for-me-ness" and presence, which is primary to all our experiences. Without feeling present as an embodied subject we would not be able to have other experiences, for instance, of visual objects in the world. Dan Zahavi has recently elaborated on the basic structure of the *minimal self* (Zahavi, 2005). The minimal self as it has been referred to as the "ongoing self-affection of a living organism" (Parnass and Sass, 2010, p. 230) and thus reflects the abovementioned phenomenological approach to the body as primarily self-affected.

I come back to the phenomenological approach in Chapter 5, where I provide more details on Merleau-Ponty's and Henry's body concepts and discuss them in relation to LIS and BCI. In Chapter 6 and 7 I argue that self-affection should not be seen as independent from the interaction with the environment but as shaped by social interactions. This idea is applied in the context of psychopathology in Chapter 8 where I illustrate why self is not only based on bodily (organismic) self-affection, but also co-constituted by others.

1.3.2 Sensorimotor Approaches to Embodiment

Another group of approaches to embodied cognition puts emphasis on the constitutive role of bodily structures for cognition, in particular sensorimotor activity. The sensorimotor approach is not coherent and entails different assumptions. Contributions stem mainly from proponents of the sensorimotor contingency approach, which is mainly applied to a specific part of cognition, namely vision and visual consciousness, but is in principle considered to be extendable to other forms of perception (O'Regan and Noë, 2001, Noë, 2004). Other defenders of the sensorimotor interaction approach are concerned with cognition in general (Varela et al. 1993, Thompson and Varela, 2001). According to the sensorimotor contingency approach, perception relies on the bodily exploration of the environment and on the "structures of our biological embodiment" (Varela et al. 1993, pp. 149, 180). According to Noë and O'Regan, these claims hold only with respect to a part of cognition, not for cognition in general.

The question remains whether their theory could be extended to include not only other perceptual modalities but also to explain to what extent sensorimotor interactions shape cognition in general. In fact, for Varela et al. at least the latter seems to be the case, given that they claim that "cognitive structures emerge from the recurrent sensorimotor patterns that enable action to be perceptually guided" (Varela et al. 1993, p. 172) and Joel Krueger (forthcoming) has recently suggested an enactive account of auditory perception.

While the sensorimotor approach adopts a new perspective on cognition with regard to the intimate relation between the cognitive system and the world, the nature of the cognitive process and an explanation of visual conscious experience, it neither provides a definition of the cognitive system nor does it offer a clear take on conscious experience in general (see 2.3.3 and 4.2.3). I will return to the sensorimotor approach to cognition in the elaborated discussion of embodied approaches to cognitive science in the context of LIS and BCIs in Chapters 4 and 5.

1.3.3 The Functionalist Approach to Embodiment

A third group of embodied approaches to cognition shares the cognitivist's commitment to functionalism (for instance Wheeler, 2005, 2010, Wilson, 2004, Clark, 1997, 2008a, and Sprevak, 2009, 2010 for critical objections to this). The most prominent and recently much discussed approach is that of *extended cognition* as put forward by Clark and Chalmers (1998) and Clark (2008a).[3] According to these authors, to consider cognition to be a matter of neuronal processes

[3] I am aware of the fact that there are now several different views that could be subsumed under the label of extended cognition. The differences between these views are currently hotly debated. There exists, as it were, a discourse on extended cognition, which is quite self-contained and relatively independent of other approaches,

only is a fallacious limitation. They suggest that cognition relies on the embodied action of the cognitive system and that under certain circumstances specific aspects of the environment could become part of the underlying basis of cognition as well[4].

In order to determine which external features can be part of cognition, Clark and Chalmers suggested an "informal test" (Clark 2008a, p.97), the so-called Parity Principle, "a rule of thumb" (p. 77, Clark and Chalmers 1998, p. 8), which helps to determine whether an external process can be considered cognitive.

> If, as we confront some task, a part of the world functions as a process which, *were it done in the head*, we would have no hesitation in recognizing as part of the cognitive process, then that part of the world *is* part of the cognitive process.

According to Clark, the Parity Principle provides a "veil of ignorance" (Clark, 2008a, p. 77) to overcome the "old metabolic boundaries of skin and skull". The prominent example to illustrate this claim is still provided by the fictitious case of Otto, who suffers from "a mild version" of Alzheimer's disease and uses a notebook to back up his constricted memory capacities. The notebook is an external tool but nevertheless it enables processes that count as cognitive.

1.4 Summary in Comparison

Let me now compare the orthodox view to the embodied approach*es* to cognitive science by assessing the degree to which the embodied approaches are compatible with the orthodox perspective. Compared to the other embodied approaches, extended cognition is the approach that is most compatible with the orthodox view. Generally speaking it mainly differs from the orthodox view in that the locus of the material vehicles of cognitive processes has been expanded beyond the brain and body. But while such an elaboration may be an advance on the orthodox view of cognition, it arguably does not revolutionize it (Di Paolo et al., 2010). For that reason I first compare extended cognition and the orthodox view. This will reveal further differentiations within the embodied approaches.

1.4.1 Extended vs. Orthodox Cognition – With Respect to (Inner[5]) Representations

The extended approach does "not deny that the mind is essentially a thinking or representing thing" (Grush in Clark, 2008a, p. 149). To emphasize that on-line cognitive processes probably require no internal representations is not meant ruling out tout court that cognitive processes are representational processes. The difference to the orthodox view is that the extended approach the origin of these representations is not necessarily neurologic, i.e. within our heads: "[S]ome of the thinking, and even the representing, may supervene on activities and encodings that criss-cross brain, body, and world" (Clark, 2008a, p. 149). Extended cognition can be seen as an attempt to reconcile computationalist with dynamicist and embodied approaches to cognition (ibid., p. 55). It provides some kind of "embodied and situated story" but it still presumes that cognition is computational and relies on internal representations (Rohde, p. 14).

---

such as cognitivism. However, when I speak of extended cognition in the present context, I am thereby referring to the extended cognition hypothesis *as originally put forward by Clark and Chalmers.*

[4] To some people it may not be obvious why I discuss extended cognition as a proponent of embodied cognitive science. After all, extended cognition is famous for its emphasis on the role of the external environment and not the body. Note however that while the originality of their proposal supposedly lies in the idea that cognitive processes extend beyond the organism (i.e., beyond brain and body), proponents of extended cognition also say that cognition is embodied. This aspect of their approach however vanishes in the background. I say more on this in Chapter 4.

[5] Where "inner" means "within the cognitive system". It does not mean that they have to be neuronal.

What lies behind this adherence to internal representations is the deep commitment of the extended approach to functionalism and thus the multi-realizability claim also found in the orthodox view. If what matters for cognition is that the material implementation realizes a particular functional profile, then there is obviously no logical reason to deny that the material substrate might be found outside of the traditional boundaries of the cognitive system. Where Putnam famously defined the functional profile of mental states based on their causal relations to input, output and other mental states (Putnam, 1975), the proponent of extended cognition simply turns this definition into a version of "beyond-the-skin species of multiple realizability" (Wheeler, 2010a, p. 8). The extended functionalist thus amends Putnam's functional profile in that it conceives of a mental state as a systemic state that is "constituted by the causal relations that it bears to systemic inputs, systemic outputs, and other systemic states" (ibid., p. 6). In this context, 'systemic' refers to the fact that the cognitive functional profile is based on a system of interwoven vehicles – brain, body and environmental features[6]. Nevertheless the nature of this larger cognitive system is still presumed to be computational and the cognitive process relies on the manipulation of representations even if those transcend the neurological realm (Wheeler, 2010a, 2010b).

1.4.2 Extended vs. Orthodox Cognition – With Respect to Consciousness

Just like the orthodox view we find that extended cognition is only concerned with unconscious and subpersonal processes; it lacks any serious appeal to the conscious level of description (Kyselo and Walter, 2009). Extended cognition is concerned with the vehicles of cognitive processes and these remain unknown to the subject. Otto is not consciously aware of the role the notebook is playing in driving his cognitive processes and actions. The use of the external tool is "deeply and subpersonally integrated into his problem-solving routines" (Clark, 2008a, p. 80).

According to Clark and Chalmers, there is a reason for the denial of consciousness. "It seems far from plausible that consciousness extends outside the head" (Clark, 2008a, p. 223). It would thus be hard to accept extended cognition if it also implied an extension of consciousness. Accepting that there are cognitive processes that remain entirely unconscious to the subject makes it easier to buy into the idea that the vehicles of unconscious cognitive processes can sometimes extend to the environment (Clark, 2009). I come back to this in Chapter 5 where we will see that in contrast to these considerations tool-incorporation actually involves consciousness.

1.4.3 Extended vs. Orthodox Cognition – With Respect to Embodiment

In contrast to the orthodox view, proponents of extended cognition say that the body is relevant for cognition. Together with brain and external environment it forms part of the information processing mechanism underlying cognition. The body is seen as a negotiable component of a larger information processing system. It is "the gateway to intelligent offloading" (Clark, 2008a, p. 207) and thus helping to spread the computational load (ibid, pp. 166, 196). Yet, to say that our body is a "key player on the problem-solving stage" (Clark, 2008a, p. 42), does not mean that its functional role could not be realized by a material substrate different from specific bodily

[6] For Clark the cognitive system comprises the organismic cognitive system (brain and body) and material artifacts (such as a notebook). Rupert (2010) has objected that we should not assume the cognitive system to comprise the technology but rather conceive of it as an embodied cognitive system that happens to interact with the (technological) environment. For Rupert the cognitive system is the organism, a "persisting collection of mechanisms the integrated functioning of which causally explains, case-by-case, instances of intelligent behavior". Not everything that contributes to intelligent behavior has to be part of the cognitive process istself (Rupert, 2010, p. 432).

structures. Noë and others are called "chauvinistic" when they claim that cognition depends on the specific bodies and sensorimotor capacities of a given cognitive system (Clark and Toribio, 2001, p. 979–980, Clark 2008a, p. 177). For extended cognition the body is "nothing but the item" playing "a certain complex functional role in an information-processing economy" (Clark, 2008b, p. 25). We will discuss the role of the body in the extended approach in more detail in Chapter 4.

1.4.4 Extended Cognition in Comparison to Other Embodied Approaches

The closeness to the orthodox view and the strong commitment to functionalism makes extended cognition quite distinguishable from all other approaches to embodiment. Instead of comparing them individually to the orthodox view, I thus briefly sketch how they differ from extended cognition. This is at least clearly the case with regard to the following three issues.

1) The commitment to the representationalist nature of cognition puts extended cognition in opposition to other approaches to embodiment, specifically to the phenomenological account, the positions adopted by the authors of the *Embodied Mind,* as well as the enactivist paradigm (as we will see in the next chapter). When considering the sensorimotor approach, we do find some departure from the orthodox view with respect to the representationalist nature of cognition – at least when it comes to visual perception, no representations are required. However, the sensorimotor approach does not deny that off-line cognition involves representations. In that vein it could be seen as closer to extended cognition and therefore also to the orthodox view of cognition.[7]

2) The commitment to multi-realizability renders extended cognition incompatible with sensorimotor approaches to embodiment that emphasize that cognitive processes depend constitutively on the possession of particular bodily structures (Noë, p. 112, Thompson, 2007, p. 13).

Note, however, that O'Regan and Noë's view invites also a more functionalist interpretation. They use the example of a missile guidance system to clarify what they mean by the idea that cognition depends on bodily structures, i.e. sensorimotor contingencies: a missile guidance system can be seen as being "tuned to" the sensorimotor contingencies that govern airplane tracking. It "knows all about" or "has mastery over" the possible input/output relationships that occur during airplane tracking" (O'Regan and Noë, 2001, p. 943). To speak about a system that masters input/put relations appears to open the door for a functionalist interpretation of their approach, after all.[8]

3) Given that all of the other abovementioned proponents of embodiment[9] (even if to varying degrees) emphasize the role of subjective experience for understanding cognition, the neglect thereof in extended cognition marks another clear disparity.

The similarity of the extended and the orthodox view of cognition seems to render the former a rather special candidate of embodied cognition. It is (with regard to the abovementioned four aspects), admittedly, the most coherent one. Yet, this coherence comes at a price. Just like the orthodox view, extended cognition remains *dichotomous, reductive* and *exclusive*. It is dichotomous because it still seems to assume that there is a clear split between the cognitive system and the external environment. It is reductive because it downplays the cognitive realm to unconscious processes. Finally, it is exclusive at least with regards to the cognitive processes

[7] In fact, as Hutto has recently pointed out, since the sensorimotor approach speaks of "having the knowledge" of sensorimotor contingencies it can ultimately be conceived of as a representationalist approach (Hutto, 2005).

[8] Thompson objects to this example that such a system involves no agency, which is however a requirement for a system to have a mastery of sensorimotor contingencies (Thompson, 2007, pp. 260–261).

[9] Except for the dynamicist and nouvelle AI view. But at least it does not rule out that conscious experience figures in the explanation of cognition.

involving environmental features because these processes are rather abstract and high-level cognitive processes (I will discuss this limitation in 2.3.2).

The aspects that would make an *embodied* approach to cognitive science a clearer alternative to the orthodox view, i.e. 1) the consideration of conscious experience, 2) an account for low-level cognitive processes, 3) the role of action and interaction and 4) the dynamical interplay between the several constituting processes of cognition (in brain, body and environment) are not found within extended cognition but only within the sum of all of the remaining approaches (the dynamicist, sensorimotor, phenomenological and enactive)[10].

The above considerations are admittedly very brief. To fully spell out the differences between embodied cognition *in general* and the orthodox view of cognition *and* to account for the difference between the various approaches to cognitive science is a complex endeavor and invites a lot more work beyond the scope of this thesis. But I will make some contributions to this endeavor by discussing four recent approaches to the body throughout Chapters 4–6. With regard to the present thesis two things are important: Firstly, that there is only one common denominator of extended cognition and all of the other so far discussed embodied approaches, which at the same time marks a clear difference to the orthodox view: *the body*. But unfortunately none of the embodied approaches clarifies what it means by "the body". Secondly, while the closeness of extended cognition to the orthodox view turns it into a coherent approach, the same is not the case for the other approaches to embodiment I have mentioned so far. None of them offers a more general story of cognition and how the body figures in it.

Let me thus now come to the final proponent of embodied cognition – the enactive approach. It is the only other embodied approach that offers a holistic epistemology, i.e. a coherent and general background theory of cognition. In contrast to the classical view of cognition this approach will turn out as not *dichotomous, exclusive* and *reductive*. It denies the separation of cognition, body and environment and it takes consciousness to be a non-negotiable dimension of cognition and it provides a story of the cognitive process that is not limited to the realm of abstract reasoning processes, but begins already at the lowest level of a system's active engagement with the world. Enactivism provides, as I argue, the means to reconcile the aspects of embodied cognitive science that so far have been merely floating around, being loosely connected to one another only through the fact that they are all not entailed by the orthodox view to cognitive science. To contribute to the enactive perspective is not only one of the major targets of the following chapters, it also very clearly inspires my own view. I have thus deliberately excluded it from the previous discussion and introduce it in more detail the following Chapter 2.

[10] Extended cognition is of course an alternative view to the brainbound conception of cognition because it does not limit cognition to the brain. But all of the new views of cognitive science make this claim. The point is that among these new views extended cognition is the perspective that differs the least from the orthodox view. Moreover, we will see in Chapter 4 that the body concept of extended cognition is so lose that it appears hard to justify – on the basis that cognition is embodied, not just neurological – that extended cognition is a true alternative to the orthodox view.

Chapter 2.

Enactive Cognitive Science

> We must not underestimate the value of a new framework in allowing us to formulate questions in a different vocabulary, even if satisfactory answers are not yet forthcoming.
> (Di Paolo et al., 2010)

In the previous chapter I have sketched three recent approaches to embodied cognitive science, the phenomenological, the extended functionalist and the sensorimotor approach. In this chapter I introduce the so-called *enactive* view to cognitive science. It is probably the clearest example of a perspective that defends an embodied position while also being a genuine critic of the orthodox view, and providing an integrative alternative.

One of the main goals of this thesis is to contribute to the elaboration of the enactive framework. I would like to show that enactivism offers a unified background theory for cognition and basic conceptual instruments, in particular the notion of *autonomy,* by means of which we better understand what the human cognitive system and the role of the body is. This chapter is of particular relevance for the entire thesis because it provides the conceptual basis for this task and for the development of the main hypothesis that human beings are cognitive systems that are best described as *socially enacted autonomous* systems (Chapter 7). The reader is invited to come back to the present chapter whenever a reminder of the key concepts is needed.

The overview of enactivism consists of five parts. I begin with some general introductory remarks, in which I explicate the term "enactive", summarize the origins of that approach and introduce one of the most important background assumptions of enactivism, the so-called *hypothesis of the continuity of mind and life*. In the next part I introduce key concepts that establish the core of the enactive approach (autonomy, sense-making, adaptivity, the scale of mediacy, experience, emergence, and embodiment). In the third section these key notions are compared to the previously introduced approaches to cognition. This is done in accordance with the four aspects from the introduction of the first chapter, i.e. 1) the nature of cognitive processes and 2) cognitive systems, 3) the relation of system and environment, and 4) consciousness. Part four elaborates on the difference between enactivism and the extended cognition as well as the sensorimotor approach.

2.1 What is Enactivism?

2.1.1 What Does "Enactive" Mean – Clarification of a Label

The term "enactive" requires clarification as it is currently used ambiguously. It is classically associated with *The Embodied Mind* (Varela et al., 1993, see 1.2.3) but nowadays it is also linked with the sensorimotor approach to visual perception and consciousness (e.g. Prinz, 2009, Clark,

2008a, see Section 1.3.2)[11]. There, "enactive" denotes the active nature of visual perception and consciousness and describes only a *particular part* of cognition.

Apart from that there are at least two other groups of researchers that use the term "enactive". The main concern of the first group is the inclusion of theories and methods of phenomenology and contemplative tradition into the research of cognitive science (Petitmengin, 2006, Lutz, 2002, Lutz and Thompson, 2003). The research focus here is on the lived experience of the cognitive system. We will see that these approaches can be seen as contributing to the development of first-person methodologies for a better assessment of the subjective dimension of cognition.

The second group, more in line with Varela et al.'s approach (1993), differs fundamentally from the previous ones because it uses the label "enactive" to refer to an *integrative theoretical framework* and a new model of cognition as a whole, not only of particular aspects of cognition (Di Paolo et al., 2010). For this group, enactivism not merely advances or complements classical conceptions of cognition, but employs a critical attitude, more fundamental and encompassing than the previously discussed embodied approaches. The radical renunciation of the orthodox position outlined in the previous chapter is probably the reason why proponents of this group of enactivism consider their approach not as a reform of previous perspectives, but as a novel paradigm for cognitive science entirely (e.g. Rohde, 2011).

In order to understand the enactive perspective and the elaborations throughout this thesis it may be useful to keep in mind that the framework is still in development and that some of the concepts below may therefore not convey as strong and self-sufficient definitions. There are still many open problems and room and need for elaborations. But as Di Paolo and colleagues recently said:

> The usefulness of a research programme also lies with its capability to grow and improve itself…it is important to be engendering rather than conclusive, to indicate horizons rather than boundaries. (Di Paolo et al, 2010, p. 35)

But despite this openness for improvement one can say that this enactive approach already offers an articulated background theory for cognition and a road map towards a genuine paradigm shift.

In this vein I consider this approach a holistic, encompassing theory for a new cognitive science that other approaches subsumed under the label "embodied cognitive science" may, where compatible, be integrated with. It is this *integrated framework* of enactivism that I further refer to from now on wherever I use the term "enactive".

2.1.2 Origins of Enactivism – A Synthesis of Methods and Theories

Enactivism should be seen as a renaissance of already existing theories and their timely synthesis with frameworks and methodologies of recent and latest developments in both theory and experiment. It is inspired by and can be linked to the autopoietic theory of the biologists Maturana and Varela, the developmental psychology of Jean Piaget, the philosophy of life by Hans Jonas, cybernetics, and dynamical systems theory, as well as American Pragmatism (Rohde, 2011). It strongly relies on continental philosophy, especially existentialism and phenomenology, and is finally complemented by recent developments in Artifical Life (ALife) as well as Robotics (Harvey et al. 2005, Di Paolo et al., 2010).

The combination of these different approaches is reflected in a general methodological and epistemological attitude, according to which cognition should be approached in a naturalistic, but non-reductive way (Rohde, 2011, p. 51). This attitude is naturalistic in relying on third-person

[11] It appears that the term "enactive" has actually already been used in the 60s by social constructivist Bruner. He therewith specified a motor and action-based type of representation use involved in the cognitive development of children (Bruner and Kenney, 1966). Thanks to Sven Walter for pointing this out.

methods and approaches to understanding the mechanistic, organizational structure of the cognitive and in grounding the theory of cognition in a theory of biology. It is non-reductive in that this third-person perspective is seen as intimately tied to the first-person perspective, i.e. the experiential and subjective level of explanation. Neither perspective is self-sufficient, but requires its complement and mutually shapes the other.[12] It is also non-reductive in the sense that cognitive phenomena are considered as emergent phenomena that are inexplicable on the basis of the isolated components that constitute them (see 2.2.7). The methodological toolbox of enactivism reflects this epistemological background assumption as it ranges from phenomenological enquiries, first-person methodologies, second-person interview techniques to computer simulations of artificially evolved cognitive agents and dynamical systems theory.

2.1.3 The Life and Mind Continuity Thesis

> life comes before mind and mind is part of life (Thompson, 2007, p. 128).

The enactive framework is based on a particular perspective on the origins of cognition and its place in the world. Cognition is not regarded as a phenomenon that occurs only at abstract levels of (human) cognition, but as deeply rooted in the origins of life. This idea is essentially expressed by the so-called *life-and mind continuity* hypothesis (LMC):

> According to this thesis, life and mind share a set of basic organizational properties, and the organizational properties distinctive of mind are an enriched version of those fundamental to life. Mind is life-like and life is mind-like. (Thompson, 2007, p. 128)

If cognitive and living phenomena are "forming part of a continuum" (Di Paolo et al., 2010, p. 36), then a proper account of the cognitive needs to begin with the principles that describe the organization and the behavior of the simplest life forms[13], like for instance of single cell organisms, to use one of the prime examples in the enactive literature (e.g. Varela, 1997, Thompson, 2007, Chapter 3). The principles we find at this lower level of complexity are then applied at various other levels of description (Clark, 2001, Thompson and Varela, 2001). Note that "organizational properties" here refers to those properties that are shared by all life forms and what they imply for cognition. There are also organizational properties that are shared only amongst particular life forms, say in animals. We begin to understand human cognition by looking at the properties shared by all life forms and then continue by looking at more specific shared organization in animal and human life forms. In both views, i.e. considering shared properties of all life forms and shared properties of specific animal life forms we do not assume an evolutionary continuity but a continuity of life and mind properties.

Since cognition is grounded in life, the leading model of the enactive approach to cognitive science is *the living organism* as a whole – not the computer (as was the case with the orthodox view). Cognition is not a compared to a computational but to biological processes (Hutchins, 2011, p. 427). Note that there may not always be a direct or obvious linkage to the biological roots, but in order to understand cognitive phenomena as well as how they interconnect and at different levels of complexity, it will be helpful to get back to the basic level of the organism and the principles of its cognitive organization.

---

[12] In that sense it is possible to conceive of the second "enactive" group which focuses on phenomenological and Buddhist methodologies as an integrated part of the present framework, i.e. as a more focused elaboration of the first-person perspective.

[13] Where "life forms" refers to living systems as they are defined by the theory of autopoiesis as I will explain in Section 2.2.1.

2.2 Key Notions of Enactivism

In what follows I introduce the key notions (derived from Thompson, 2007, p. 15 and Di Paolo et al., 2010) that constitute the core of the enactive paradigm.

Recall that enactivism is still in development. These key notions are considered as the guidance for such development. Ideally, any elaboration of the enactive approach should be based on them (Di Paolo et al., 2010). The core ideas of enactivism are a deeply intertwined network of interrelated building blocks that together constitute the enactive model of the mind. Reading them separately would obscure the fact that they are all aspects of one unified background approach to cognition. In order to get a first idea of how the notions generally relate, consider again the four basic aspects based on which I compared the previous approaches to cognitive science: 1) What is a cognitive system, 2) a cognitive process, 3) the role of experience and 4) the relation between cognitive systems and the world (see Chapter 1)?

The first concept – *autonomy* – refers to the enactivist definition of the cognitive system (aspect 1). According to this definition every cognitive system is an autonomous system, generally defined as a "system composed of several processes that actively generate and sustain an identity under precarious circumstances" (De Jaegher and Di Paolo, 2008, p. 35).

There are two senses in which the cognitive system is an autonomous system. It is firstly autonomous with respect to the production and maintenance of its own *identity* as a cognitive system. Secondly, it is autonomous with respect to its interactions with the world. This part of autonomy is referred to as *sense-making*, which is the "interactional and relational side of autonomy" (Thompson and Stapleton, 2009, p. 25). *Sense-making* is the concept denoting the enactive stance on what a cognitive process is (aspect 2). *Experience* in the enactive reading is a natural and necessary consequence of both autonomy and sense-making (aspect 3). *Embodiment* is a necessary condition for a system to be autonomous. Without being embodied there can be no cognition. There is need to specify what exactly this condition amounts to. In the course of this thesis I elaborate on this question and contribute to this task. The notion of *emergence* finally explicates how autonomous systems and their sense-making activities are constituted. Understanding the constitution of autonomy will shed light on the enactive perspective on the relation of cognitive systems and the environment (aspect 4). Again, it is useful to keep in mind that if *autonomy* refers to the very nature of the cognitive system, then all of the other notions that make up the enactive framework to cognition must be seen as intrinsically linked to it.

In accordance with the life-mind continuity hypothesis, the first step for understanding autonomy, in the above-mentioned dimensions, is to consider it in its minimal form, i.e. as the kind of *biological autonomy* introduced by the theory of autopoiesis (self-production) famously put forward by Maturana and Varela (Varela, 1979, Maturana and Varela, 1980, 1987). The theory of autopoiesis describes living systems as self-organized networks that produce and sustain a systemic whole – an *identity* (Section 2.2.1).[14]

The *second* step (Section 2.2.2) is to conceive of biological autonomous systems as intrinsically purpose-driven (Jonas, 1966, Weber and Varela, 2002, Thompson, 2007). In a *third* step biological autonomy is further elaborated by the notion of adaptivity (2.2.3). Only adaptive systems are genuine sense-making, i.e. cognitive systems. They are not only following the norm of survival but generate more fine-grained norms which guide their interactions with the environment (Di Paolo, 2005). The *final* stages of autonomy discussed here are interactive regulation and the constitution of a self-image at the level of human cognition (2.2.4).

[14] Note that this perspective on cognition involves a fundamental shift in the evaluation of what counts as a cognitive system. It characterizes a much wider range of systems as cognitive. By defining the core features of the cognitive system we implicitly also define the core features of experience. Since the enactive approach broadens the perspective on cognition, it also broadens the term of experience. In the present context, experience should be understood as some form of basic sentience, or interiority but not necessarily as including a level of awareness or self-awareness as we find it in human being. These conceptions should not be interpreted anthropomorphically.

2.2.1 Autopoiesis – The Principles of Basic, Biological Autonomy

Autopoiesis is a theory about the principles according to which living systems produce their identity as a system in terms of self-organizing networks of processes of material transformation (Varela, 1997, pp. 75–76). According to this approach, a system counts as cognitive as soon as it respects the conditions of a living system. In the following I briefly explicate the principle of autopoiesis and thereby sketch the first aspect of autonomy, the construction of an identity.

According to autopoietic theory, a living system is a self-organizing "network of processes of production of components" (Varela, 1997, p. 15), where these components produce and maintain that network and thereby form "the system as a distinguishable unity in the domain in which they exist" (Weber and Varela, 2002, p. 115). Simpler formulated: Living systems consist of processes that are related to each other such that they bring about and sustain the system as a unitary whole, i.e. an *identity* within a particular domain. In order for the system to maintain its identity some process relations within the system remain constant. They sustain the invariance of the systemic organization or identity (with respect to both the process relations inside the system and the interactions with the environment) despite structural changes. Since these processes are found within the system itself, autopoietic systems are described as *operationally closed*:

> A system is operationally closed if, for any given process *P* that forms part of the system (1) we can find among its enabling conditions other processes that make up the system, and (2) we can find other processes in the system that depend on *P*. This means that at some level of description, the conditions that sustain any given process in such a network always include those conditions provided by the operation of the other processes in the network, and that the result of their global activity is a identifiable unity in the same domain or level of description (it does not, of course, mean that the system is isolated from interactions with the environment). (Di Paolo et al., 2010, p. 38)

Through operational closure the system establishes a border between itself and system-external processes that thereby become extrinsic and thus form the environment of the system. Any system whose systemic unity is established by external processes is not operationally closed. In contrast to the dichotomous view of the relation between a cognitive system and the environment in early cognitive science, the environment is not objectively given and independent from the cognitive system, but rather brought about through the system's own bodily structure and activity. The boundary between them emerges as a distinction established on the basis of the operational principle, i.e. those system-inherent processes that produce and ensure the system's identity as a whole. Processes that do not contribute to the production of identity are considered as environmental processes. On the other hand, every autopoietic system is a system only with regard to the environment it is embedded in and distinct from.

Note furthermore that operational closure takes place under so-called "precarious circumstances" (Di Paolo and Iizuka, 2008, Di Paolo, 2009). It does not suffice that the processes that constitute the autonomous network sustain themselves as that network, but that they are crucially interconnected *as* a network and thus ensure each other's maintenance. Precariousness means that without being organized as a network the processes that make up this network would "run down or extinguish" (Di Paolo, 2009, p. 16) and therefore the network as a unified whole would dissipate.

Let me illustrate in simpler terms what it means for a system to be operationally closed and to be distinguished from the environment by using the classical example of an autopoietic system in its simplest form: the single cell. Imagine at first to look only at a sea of biochemical processes. A closer look may reveal that some of the biochemical processes do not exist independently of others. An interference with a particular process *a)*, e.g., leads to a direct effect on processes *b)* and *c)*. One might say that process *a)* enables processes *b)* and *c)*. Some processes in the sea of biochemical processes interact in a way that any singular process in the interaction is

dependent on at least one other process. For that reason these processes form a network (see Figure 2.1).

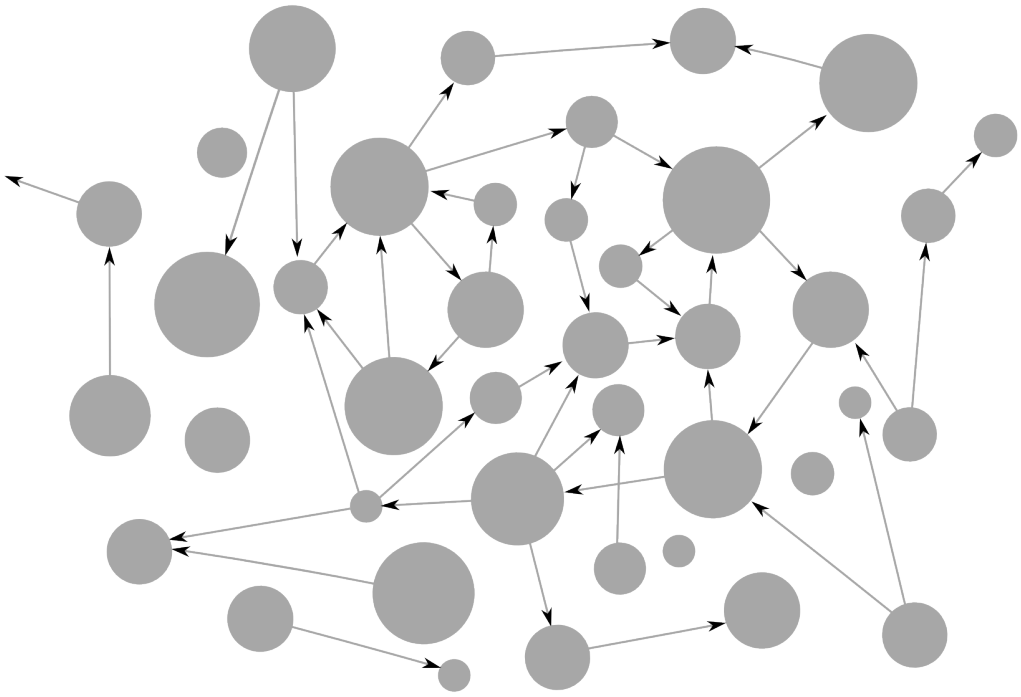


Figure 2.1: Schematic illustration of a network of biochemical orocesses. The arrows leading to and/or from each circle indicate an enabling relation between the respective processes (grey circles). Processes that are connected to at least one other process constitute a network.

Operational closure of the network arises if some of the processes are not just enabling but are also enabled by each other. If one would remove, e.g. *a), b)* or *c)* and in result of this any of these interrelations would break down then this indicates that the processes have previously not only formed a network, but also *precarious* network. The fact that they are organized as a network ensures their operation – the processes constituting this network would dissolve on their own. If we now shift the focus of observation we find in the sea of biochemical processes those processes that either only affect or that are only affected by each other. They are not part of the operationally closed network, but belong to the system's environment (see Figure 2.2).

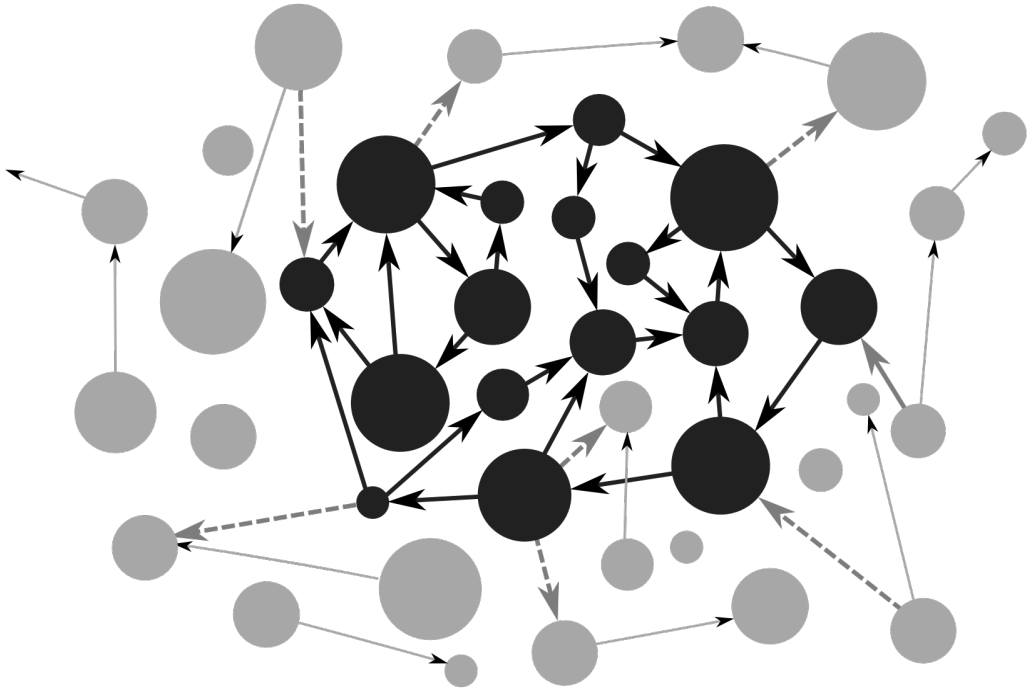


Figure 2.2: Schematic illustration of an organizationally closed network. The black circles and arrows refer to a collection of biochemical processes that are a) dependent on at least one other process of the network and at the same time b) enable themselves at least one other process within the network. These processes have at least one arrow going out *and* one arrow going in. Processes that have only a unidirectional relation (dashed arrows, going either in or out), might affect the operationally closed network but they are not part it. They belong to the network's environment.

The single cell is essentially this interconnected and self-sustained network. It is an organized collection of biochemical processes that are interacting with each other in a way that excludes other biochemical processes. To say that the cell is "operationally closed" means that the cell *as* this connected unity of processes is produced and maintained only by the interconnected processes that belong to it, i.e. the sustained network of processes. If other external processes would interfere with that network then the distinction between the network and these processes would fade and the cellular identity dissolve. In case of the cell the operations that establish the system's closure are called metabolic. The cellular metabolism creates a material membrane that establishes a physical boundary. This boundary allows the observer to then distinguish the cell as a whole against the background of several other biochemical processes.

It should be clearer now why the enactivist would not say that the environment is just objectively given and indendent from the cognitive system. We only came to consider some biochemical processes as the environment because they do not belong to the network of biochemical processes that contribute to the maintenance of the system's identity.

Here lies the basis for the definition of the cognitive in enactivism: cognitive systems need to respect the most basic condition of biological autonomous systems to generate and maintain an identity. Crucially, the systems are autonomous in the process of doing this. They are not determined from the outside, but "have a say" in their proper constitution as a cognitive

system.[15] This does not mean that the system is entirely independent from the environment (Rohde, 2011). It is dependent in the sense that the system-inherent processes, which ensure the production and maintenance of the systemic unity, require environmental resources, e.g. "matter and energy…to fuel metabolism", (Di Paolo et al., 2010, p. 38). This dependence on the environment has been referred to as *structural coupling* and it should be understood in terms of an enabling condition. However, the characterization of an autonomous system as precarious hints at a much stronger sense of dependence. The processes belonging to the system are not only enabling each other's constitution but they also *have to* enable each other's constitution in order to prevent the network from dissolving. They are determining each other. We thus make a distinction between processes that determine the network and processes that enable the network of processes to operate.

To say that the autopoietic system is structurally coupled to the domain in which it exists means with regard to the cell that the cell is produced by a network of biochemical processes that constitute a distinction from other biochemical processes. But in the same time the cell still needs these other biochemical processes to feed the metabolic network. Interactions with system-external processes thus enable the cell's operation, they can perturbe them and trigger reactions. However, as long as the operationally closed organization of the cell is preserved, the interactions with the environment do not determine it (Maturana und Varela, 1987).[16]

In order to clarify the difference between structural coupling and operational closure consider a simple example: a car's engine system. The engine has a particular mechanically defined organization that, if every one of it components functions properly, provides energy that can be used to engage the wheels so that the car moves. Of course in order to provide this energy the engine requires the appropriate fuel. For this reason one can obviously contend that the car is dependent on fuel. However, the fuel is only enabling the car to run, it does not also affect the mechanical organization of the engine system itself, i.e. if the right fuel is chosen it will not determine the car's functioning. The car's functioning as a moving vehicle is therefore operationally independent from the environment, while structurally coupled to it in that in order to realize that the operations take place it requires external resources from it such as fuel. Note that this is not to say that the car is an autonomous system because this would require that the elements realizing its functioning would be entirely determined by the relations they have with each other (precariousness). However, since a car does not repair itself but requires additional care to ensure the endurance of its mechanical components, it is determined by the external environment and not by its own operational elements.

Enactivism has been deemed to equate cognition with the original autopoietic theory (Wheeler, 2010a, in press, see Di Paolo, 2009 for a reply). But this is misleading. While the formation and maintenance of a material identity in terms of autopoiesis is one clear instantiation of biological autonomy, it can also have other realizations. We see in the next sections that autopoiesis is more recently not considered as sufficient for characterizing the cognitive.

---

[15] According to this, Herbert, Brook's can-collecting robot, does not count as an autonomous system properly because its identity as a robot is constructed externally. Its actions follow laws that are determined from outside. A robot counts as autonomous only "when it regulates its behavior according to rules or laws that it has constructed for itself" (Smithers, 1997, p. 94).

[16] To describe autonomous systems as self-organized operationally closed networks also marks a contrast with non-living systems. A system that is not self-producing and therein operationally independent from the environment does not count as an autonomous system. And since the self-production and maintenance of a systemic identity is the decisive characteristic of the living, any system that is not autonomous is also not alive. Non-autonomous systems are determined externally, and they are based on pre-designed laws. Those systems are called *heteronomous* (hetero=alien). An example of a heteronomous system would be an automatic bank machine. Its functioning is exclusively determined from outside (Thompson, 2007, p. 37).

2.2.2 Autonomous Systems are Purpose-Driven – The Basis of Sense-Making

Weber and Varela suggested elaborating the autopoietic theory by linking it to Hans Jonas' approach to the philosophy of biology. Hans Jonas was also concerned with understanding the organism as whole, but he pursued his analysis from a phenomenological-existentialistic point of view. According to him, organisms have an intrinsic purpose: once they are alive, they strive to stay alive. An organism constitutes a norm, which is not imposed by the outside, but arises naturally within itself as the need to ensure the continuation of its existence (Jonas, 1966). For Weber and Varela this provides an important source of refinement for biological autonomy in that normativity should be considered as an intrinsic dimension of the production and maintenance of biological autonomy (Weber and Varela, 2002). This elaboration would allow "to put the autonomy of the living at the center" (ibid., p. 112) in a naturalistic way, providing not only a phenomenological description of teleology but also a scientific explanation that is not based on a "derived or extrinsic normativity" (Barandiaran and Egbert, to appear). For an explanation, consider again the principle of operational closure and the notion of precariousness. The bio-chemical processes involved in the metabolism of the cell are connected to each other as a network, which only on its own ensures the maintenance of the metabolic identity. If other, external bio-chemical processes would affect the very working of metabolism, the cell as a unity would dissolve. Thus, the cell depends not only on external bio-chemical processes to feed its metabolism, but it also constantly *strives* to exclude them from the network of processes that constitute the metabolism itself to ensure that it does not disintegrate as a whole.

A precarious living system is thus permanently concerned (Weber and Varela, 2002, p. 113), it is "inherently restless" and "never fully safe" **(**Di Paolo, 2009, p. 16): once it comes to existence it is constantly facing the danger of disintegration, i.e. death, and thus needs to continuously engage in the maintenance of its existence.[17]

The basis for understanding the nature of a cognitive process in the enactive sense lies exactly here. A cognitive system is a self-organized network which sustains an identity, and for this reason it is intrinsically concerned with the need to maintain this identity. Based on this arises in every cognitive system a *perspective* on the world (Varela, 1997), from which it encounters the world by imbuing interactions with a normative status (Jonas, 1966, Thompson und Stapleton, 2009). An autonomous system does not discover meaning but is a "meaning producer" (Baerveldt and Verheggen, 1999, p. 195). According to Weber and Varela, cognitive systems are therefore not only intrinsically purpose-driven with regard to identity, i.e. the affirmation of their existence itself, but also with regard to a drive for meaning construction – they have a "sense-creation-purpose"(Weber and Varela, 2002, p. 117).

Autonomous systems do not passively respond to some pre-given, external stimuli, but evaluate their engagement with the world according to self-generated norms that are based on the need to stay alive. Cognitive processes are not reactions to a problem or an affordance posited by an external environment. Instead of representing the objective world and discovering meaning as an already given, as assumed on the orthodox view, they refer to the behavior of autonomous systems, as an active process of meaning generation or sense creation (Weber and Varela, 2002, Thompson, 2007). They result from the system's own active engagement. This process is called *sense-making* and it constitutes the second building block of the enactive model of cognition.

> Invariants are instead the outcome of the dialogue between the active principle of organisms in action and the dynamics of the environment. The "finding" of meaning…is always an activity with a *formative* trace, never merely about the innocent extraction of information as if this was already present to a fully realized (and thus inert) agent. (Di Paolo et al., 2010, pp. 39–40)

[17] In chapters 7 and 8 I apply the notion of precariousness to human level autonomy. I suggest that the environment against which a human autonomous system needs to "defend" its identity while at the same time also constructively relying on it, is a social one.

This double *intrinsic* teleology (Weber and Varela, 2002, Thompson, 2007) – in terms of identity-preservation and sense-making – explains the need to distinguish two observational perspectives: the perspective of the external observer and the perspective of the cognitive system itself. What counts as the environment of a system is not what the external observer would describe as its physical surrounding. In realizing the sense-making purpose autonomous systems bring about a distinction between the world as physical and an environment as it is seen from their own perspective (Weber and Varela, 2002). They shape their own world, i.e. specific features of the environment acquire meaning, which is meaning *for* them. The distinction between a cognitive system and the environment, which in orthodox cognitive science is considered to be ontological and pre-given, appears in enactivism to result from the organism's intrinsic drive to stay alive and the drive to ensure this norm in its (sense-making) interactions with the world (Weber and Varela, 2002, p. 120). To illustrate this, consider a bacterium swimming up a sugar gradient. How should we explain this behavior? According to the enactive view, sugar is not simply a condition of the material environment (Thompson, 2007, p. 153), but becomes meaningful for the bacterium itself because it needs to nourish itself, i.e. feed its metabolism and strives to ensure the continuation of its existence.[18] The fact that sugar has the status of food cannot be read of from the structure of sugar itself; it does not have built-in value given objectively through the environment. Its value depends on the bodily structure and the behavior of the organism guided by its intrinsic goal to stay alive (Jonas, 1966, p. 126, Barandiaran et al., 2009, p. 375, Di Paolo et al. 2010, p. 16, Barandiaran and Egbert, to appear). Values and norms should neither be seen as an internal mechanism of the individual nor as relying on environmental structures alone. They are the emergent result of the organization of a particular autonomous system and its interaction with the world.

2.2.3 Sense-Making Requires Adaptivity

The first step to understanding an autonomous system was to conceive of it as a living system that seeks to preserve its self-produced identity. At the most basic level this construction of identity is seen in comparatively simple biological systems. We have then seen that these systems further employ two complementary intrinsic purposes (Di Paolo, 2005, p. 433), the identity formation and sense-creation purpose.

However, in a second wave of replying to the worry that cognition is equated with autopoiesis, Di Paolo has recently argued that this two-step characterization of autopoiesis is still not sufficient to properly ground the notion of sense-making. The concept of autopoiesis explains the behavior of cognitive systems only in terms of an either-or norm. Environmental features acquire only a singular value, the "mother-value of all values" (Weber and Varela, 2002, p. 111). This means that the systems are always concerned with basic survival (mother value) and that the environment is thus evaluated with regards to potential perils to their existence itself (*either* it is life threatening *or* not). At the basic autopoietic level the sense-making capacities of the living systems are thus limited to an immediate reaction to the metabolic requirements, action directly concerned with the physical environment. There is no regulation on the part of the system to engage in a more flexible behavior (Rohde, 2011, p. 87). Mere structural coupling (see 2.2.1) of autopoietic systems with their environment "is a conservative, not an improving

[18] Some people may worry that there is no actual difference between a bacterium and my example of the car engine. Sugar matters for a bacterium's survival as oil matters for the car's "survival". I agree that an analogy could be drawn in that both systems require particular processes to sustain their identity as a system. The difference between car engine and bacterium is that the car depends operationally on the external environment whereas the bacterium as a precarious system ensures its survival by itself. The bacterium has an instrinsic striving to maintain its identity as a system and the operational processes that ensure this, belong to the bacterium itself. The car's maintenance (which includes to ensure that there is always enough oil) is ensured by the outside. To put it simply: the bacterium is alive and striving to stay alive, the car engine is not.

process; it admits no possible gradation" (Di Paolo, 2005, p. 437). However, a cognitive system is not only seeking to preserve its existence, it also exhibits a more sophisticated behavior.

Di Paolo thus specified the link between autopoiesis (or autonomy in general) and sense-making by adding the requirement that an autonomous system should also and crucially be an *adaptive* system (Di Paolo, 2005, see also Thompson, 2007). Adaptivity is defined as follows:

> [Adaptivity is] a system's capacity, in some circumstances, to regulate its states and its relation to the environment with the result that, if the states are sufficiently close to the boundary of viability,
> 1. tendencies are distinguished and acted upon depending on whether the states will approach or recede from the boundary and, as a consequence,
> 2. tendencies of the first kind are moved closer to or transformed into tendencies of the second and so future states are prevented from reaching the boundary with an outward velocity. (Di Paolo, 2005, p. 438)

Genuine sense-making systems are capable of monitoring and regulating their states and they can modify the conditions for the sustenance of their identity. This not only requires that they can change their inner structure as a reaction in the face of environmental factors, but also that they regulate how they interact with them. To distinguish and transform tendencies is to say that environmental features are evaluated with regards to the basic norm of survival in a graded way, adding a range of more fine-grained norms and according to which the behavior is now more flexibly regulated (ibid., p. 439). These fine-grained norms provide a measure of how good a system is in remaining and sustaining its self-production (Di Paolo et al. 2010, Di Paolo, 2005, for a recent elaboration on adaptive normative behavior see Barandiaran and Egbert, to appear). They allow to distinguish "between good, neutral, and bad ways of realizing autopoiesis" (Di Paolo, 2005, p. 440).

To illustrate this, Di Paolo also considers the example of the bacterium swimming up the sugar gradient. According to a mere autopoietic explanation the reason why the bacterium is swimming up the gradient is its intrinsic goal to stay alive. But in Di Paolo's view the additional questions to be asked are: "What makes bacteria swim up the gradient? What makes them distinguish and prefer higher sugar concentrations?" (ibid., p. 437). If the bacterium is only autopoietic, but not adaptive, it should not prefer higher concentration of sugar but remain satisfied with a concentration of sugar just enough to keep it alive. This would be a non-graded evaluation of the reaction to environmental features. However, if we consider the bacterium to be also adaptive, then it will be able to *improve* the ways in which it is keeping itself alive. The bacterium regulates its behavior flexibly. It will evaluate not only the sugar as meaningful but also the gradient itself, i.e. "the direction where the concentration grows as useful" (ibid.) and thus occasionally swim towards higher sugar concentration. This evaluation is graded because it is not following an all-or-nothing norm. It enlarges the possibilities of interaction: from an immediate reaction to the ability to recognize in environmental features additional qualities and "forces and tendencies that have to be compensated for" (Barandiaran and Egbert, to appear, ms.p. 11).

2.2.4 The Scale of Mediacy

In order to illustrate the intrinsic purposefulness of autopoietic systems and the principle of self-organized identity production, rather simple examples like the single cell and the bacterium have been used so far. This may raise the worry that enactivism is only concerned with low-level cognition. However, as Rohde reminds the critic, "the enactive approach is not obsessed with bacteria cognition" (Rohde, 2011, p. 16) but also strives to account for more complex variations of sense-making activity. The metabolic level is the basic, but the constitution of autonomy identity and sense-making can also take more sophisticated forms.

In order to illustrate how cognitive systems sustain their autonomy and generate meaning at different levels, enactivists appeal to the so-called "scale of mediacy" (in elaboration of Jonas,

1966). The scale allows to order the previously discussed aspects of autonomy in a hierarchical manner (see Figure 2.3).

The first stage on the scale of mediacy refers to the level of preservation of metabolic survival. We have seen that autopoiesis further implies goal-directedness not only with regard to the preservation of identity but also to its interaction with the environment. In order to qualify as sense-making systems, we finally added the notion of adaptivity.

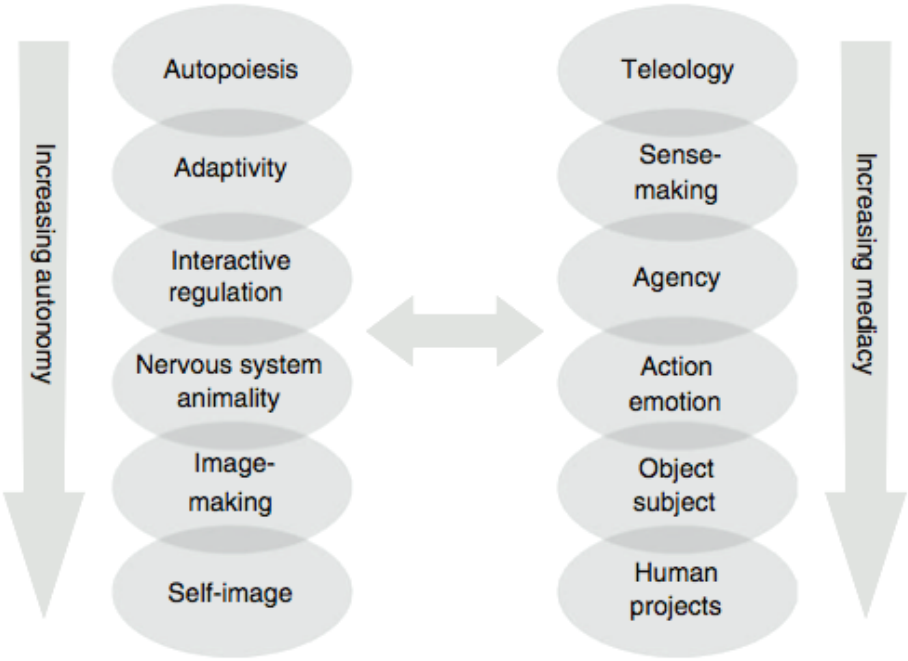


Figure 2.3: Life-cognition continuity and the scale of increasing mediacy.
(from Di Paolo et al., 2010, p. 50)

The third level at the scale is specifying autopoiesis and adaptivity even further. Cognitive systems regulate their behavior interactively, i.e. they not only regulate external states, but also themselves and thereby improve the way they maintain alive. Animals that have more complex bodily structure and use motion in order to find better prey grounds would be an example of this (Barandiaran et al. 2009, p. 380). They are not only ontologically individuated from the environment like plants, but because they can move they are also distinguished individuals in terms of space (Barbaras, 2011).

With this third step on the scale of mediacy, the "interactive regulation", cognitive systems enlarge their sense-making capacities as meaning generation can not only be decoupled from the direct physical environment, but may take a form which seems more and more distant to the basic goal of survival.

> The higher the degree of mediacy, the more complex it is for the observer to interpret a sign with respect to the process(es) of identity generation from which its value emerges. (Rohde, 2011, p. 88)

This illustrates the difference between the enactive perspective on the cognitive process as *constructive* rather than "*uncovering*" as implied by the orthodox view. To construct meaning rather than to uncover it is to evaluate aspects of the environment not simply with respect to momentary physiological concerns but also to "future consequences". The cognitive system can react to environmental circumstances in a given moment even though they do no directly affect its physiological survival, they are evaluated in an "anticipatory manner" and maybe also in

relation to other aspects of the environment: "Snowprints left by a prey are meaningful even though metabolism is not feeding on the prints themselves" (Di Paolo, 2005, p. 442). A further complexity in autonomy identity is expressed at the next stage of the scale, at which we find more complex sense-making activities of systems that are not only motile but also have a nervous system. Finally, and important with regards to the present thesis, we find the last two stages of human beings. Their interactions with the environment are mediated by sociality. At this level of complexity autonomous systems create a perspective on themselves in the sense of a self–image, they "define themselves as subjects" (Di Paolo et al. 2010, pp. 50–51). Here the organizational structure of cognitive systems comprises "both constructive and interactive aspects of the living", i.e. both sides of autonomy – identity and meaning constitution – can interact so as to bring about new forms of autonomous identity (Di Paolo, 2009, p. 18). An example of such a new form of autonomous identity is the so-called phenomenon/a of *participatory sense-making* (De Jaegher and Di Paolo, 2008) which is described as "the interactive coordination of intentional activity affecting individual sense-making, whereby new domains of sense-making may appear that were unavailable to each solitary individual" (ibid., p. 41).

Participatory sense-making is an elaborated form of sense-making, it is a socially mediated form of meaning generation based on a dynamic process of coordinated activity of several individuals which brings about a new form of autonomy that is irreducible to individual actions. It involves both the identity construction and the interaction of individual agents and generates a new, shared world of meaning, which in turn enlarges that of the individual agent on its own.[19] In Chapter 7 I elaborate on a potential implication of this claim, namely that participatory sense-making not only enlargens the individual's sense-making capacities but co-constructs its very identity.

2.2.5 Experience

According to enactivism, autonomy is intrinsically related to subjectivity and the conscious experience of the system in question (Thompson, 2007, p. 15). By being and sustaining an identity, the system generates not only a center of activity (autonomy) but also an "interiority", i.e. a specific viewpoint on the world (Thompson, 2007, p. 78). From this perspective encounters with the environment are evaluated as meaningful or not (*sense-making*). Cognition and consciousness thus belong together. They are both reflecting the embodied system's engagement with the environment (Torrance, 2005, p. 4) and allow it to "enter into a relation with the environment on the higher level of meaning" (Fuchs, 2011, p. 216). Experience is not a

[19] Some people may worry that shared meaning can only exist within a more classical account of meaning according to which it would be pregiven in the outside world so that it could be jointly discovered. But in the enactive view meaning is not in the world but only constructed by the individual's interactions with the world. If we accept this it seems hard to imagine that meaning could be shared by others.

This worry is unsubstantiated because the point about shared sense-making remains basically the same as in individual sense-making. The meaning is still not discovered but generated by the agents. In contrast to the individual sense-making, in participatory sense-making the process of meaning generation is based on the interactions that several individuals engage with and not on the activity of a single agent. Consider for an illustration a group discussion about a philosophical question, say "What is the sense of life?". Let us suppose that somebody is taking notes during that the discussion and that at the end of the meeting she is able to summarize a possible answer to this question: "The sense of life is not to be dead, and in a nice way". The point of shared sense-making is that this answer is neither produced by the person taking the notes, nor by any other individual in the group. The answer has rather emerged in a joint activity, i.e. through a conversation the content and direction of which was not determined by an individual speaker but by the dynamics of the communicative interactions of the participants in the discussion. This answer was not at once discovered together. It mirrors a shared and progressive achievement through dialogue. Shared meaning presupposes that the meaning is co-constructed in interaction. Even if such interaction, i.e. participatory sense-making may at times be asymmetric, as in a teacher-student context, there will always be a moment of mutual participation.

mysterious or epiphenomenal aspect of cognition, but arises naturally in self-preserving systems a proper part of the scientific explanation of cognition (Weber and Varela, 2002, p. 116).

For the enactivist subjectivity is therefore not only a property of more complex cognitive systems, such as human beings. Cognition is grounded in autopoiesis and autopoietic systems feature a basic form of sentience. Thus even a simple life form, such as a bacterium can therefore be said to possess basic sentience, i.e. a minimal form of subjectivity.

As a consequence, enactivists consider phenomenological inquiries to be essential for understanding the human mind. They rely on classical Husserlian phenomenology as well as on the work of Merleau-Ponty. More recently enactivists also consider insights of the phenomenologist Henry with respect to understanding the self-affection of life, as well as the social phenomenology by Schütz and Gurwitsch (Fuchs and De Jaegher, 2009, De Jaegher and Di Paolo, 2008, Krueger, 2011, Thompson et al. 2005). Rohde has recently defended the enactive view that the study of phenomenal experience is part and parcel of research on cognition, despite not being typically "scientific":

> While the aspiration to maintain scientific standards is honourable, it prohibits the study of first person non-measurable experience, a central aspect of cognition and essential for the definition of many mentalistic concepts and distinctions. Cognitive science, therefore, finds itself in denial, trying to deal with experiential phenomena whilst pretending not to be dealing with experience. (Rohde, 2011, p. 43)

But note that the study of the phenomenological dimension of cognition does not amount to mere speculative endeavor. In fact, enactivism aims to develop phenomenological inquiries into a disciplined method. One such developing method is *neurophenomenology*. It combines first- and third-person data into a natural account of access to subjective experience (e.g. Lutz, 2002, Lutz and Thompson, 2003, Thompson and Cosmelli, in press). Neurophenomenology assumes that subjects are generally not good at assessing their experiences and that they require training in phenomenological observations. The general approach of neurophenomenology is therefore to train subjects in order to acquire more fine-grained data in terms of thoroughly observed phenomenological invariances (Lutz and Thompson, 2003, p. 32). In order to increase subjects' attention and sensitivity to their experiences, so-called "first-person methods" are applied. One of them relies on Vipassana meditation (Goenka, 2000) where subjects are trained to increase their awareness by observing their (bodily) sensations and experiences with a pragmatic epoché – they are asked to only observe their occurrences yet strive not to judge or react to them. In a second step, neurophenomenology uses this refined data to acquire new third-person data about the neurological processes that underlie conscious experience. Another (also still developing) approach is the so-called 2nd-person method, an interview technique, allowing to bring the interviewee (the subject) closer to her experiences and to generate more precise descriptions of it (Petitmengin, 2006a, 2006b). Ginsburg (2005) has recently made suggestions for the design of first-person experiments involving the increase of bodily awareness in movement.

A final example is the so-called "PRISMA" method (Pieper and Clénin, 2010). It originates in sociology, Feldenkrais and arts but is currently further developed by Hanne De Jaegher for an application in cognitive science (personal communication). PRISMA investigates the phenomenology of embodiment in social interaction processes, relying on researchers' self-observation of embodied subjectivity, it acquires first-person data through notations of detailed momentary observations (of self and others).

2.2.6 Embodiment

The concept of the body in the enactive approach is still in an early developmental stage. To contribute to its elaboration is one of the major goals of the present thesis. The body concept in cognitive science is subject of Chapters 4, 5 and in particular of Chapter 6. There I suggest steps

towards an enactive notion of the body that integrates particular insights from other recent perspectives. Let me here provide only a short summary of what can be generally said about the enactive stance on the body. In enactivism a cognitive system is *necessarily* embodied. Contrary to a widely accepted view the body is not treated as a container for the mind, or a "puppet" that is controlled by the brain. The body rather is the means by which the cognitive system can engage be cognitive, i.e. engage in sense-making interaction with the environment. A system's activity "depends nontrivially on the body", there is therefore no cognition without embodiment[20] (Di Paolo et al., 2010, pp. 42–43).

Embodied cognition in an enactive sense is not only restricted to sensorimotor interactions but also at play in higher-level cognitive processes (Di Paolo et al., 2010, p. 43). In line with the complementary epistemology and resuming the tradition of *The Embodied Mind* (see 1.2.3, and phenomenology, see 1.3.2) the enactive body is not merely an "objective" body in the German sense of *Körper,* but rather a *Leib*, i.e. a *lived* body. It is not a material object in the world but the means by which we are situated in and directed to the world, it enables us to communicate with the world (Merleau-Ponty, 2002/1945, p. 106). At the metabolic level of autonomy but also at other levels of autonomy we experience the world as meaningful through our bodies. The body could be considered as a matrix upon which the significance of the world is inscribed.

2.2.7 Emergence

The fourth building block of the enactive framework is emergence. The notion has been implicitly at work in the abovementioned notions of autonomy and sense-making. It addresses the general problem in cognitive science: "how does it work"? (Di Paolo et al., 2010, p. 41). The concept of emergence reflects the enactivist's viewpoint on how cognitive phenomena are constituted.[21] Like other concepts it is empirically grounded in the principle of self-organization found at the level of autopoietic systems. An emergent property is a new process based on the dynamical interplay of several other processes. There are two conditions for emergence (ibid.):

> 1) the emergent process has its own autonomous identity, i.e. it is not reducible to the individual processes that contribute to its emergence (e.g. the interaction process as a whole) and
> 2) the maintenance of autonomy will impact on the lower level processes whose interaction constituted it (e.g. on the individual agents engaged in the interaction)

To say that an autonomous system is emergent means that it is not constituted by the sum of its components processes. It is a "relational phenomenon" resulting from the dynamical interrelations of processes that together form an organizationally closed network. As a consequence of this, enactivists reject any attempts to localize specific cognitive functions at a single level or process alone – they are against what they call "boxology" (ibid., p. 41). Cognition has no location; but needs to be assessed as a situated and dynamically evolved unity.

The prime example to illustrate this emergent relation between the constitutive processes and the new system as a whole is again the single cell. The cell as a distinguishable unified whole is explained by the principles of metabolic functioning which are observable at an entirely new level of description. Its systemic identity is not reducible to the properties of the component

[20] Although this is extensively discussed in the next chapters I want to clarify already at this point that the radical commitment of enactivism to the body does not imply that the body necessarily needs to be of biological origin. The enactive approach is fundamentally committed to computer models, artificial agents and robots as means to understand the organizational and mechanistic aspects of cognition. Rather than considering only abstract functions in their models the enactivists will seek to integrate the principles found at the level of living systems. So that any artificial agent, whether or not it is of a virtual kind will always be in some sense embodied.

[21] For more detail on the concept of emergence in enactivism see for instance Thompson, 2007, pp. 417–441.

processes that constitute the network, i.e. singular bio-chemical processes. The dynamics of the metabolism as a whole then impact on its component processes, i.e. the singular bio-chemical processes that constitute it.

A second example for the kind of emergent processes envisioned by the enactivists, is is the generation of so-called Bénard cells[22]. When heat is applied to a fluid, for instance cooking oil, a difference of temperature between lower and higher layers of the fluid occurs. The colder layer causes the fluid to rise (a process called "convection") and with increasing temperature the fluid as a whole changes its form into convection rolls. Seen as a system, the fluid thus went to a state transition (bifurcation) now exhibiting novel properties that are not based on its components, but on the interaction between (increasing) temperature and the joint behavior of two differently tempered layers of fluid. With increasing heat another transition takes place, and the rolls transform into so-called Bénard cells (see Thompson 2007, p. 61). Both phenomena, the convection rolls and the Bénard cells are entirely new properties.

The same more or less applies to the process of sense-making, i.e. meaning generation: Meaning is, as well as autonomy, a "relational phenomenon" and is not found within the world or within the organism alone. It is the emergent result of an interactive process between the two.[23] For the example of participatory sense-making this means that two interactors coordinate their behaviors in such a way that a new property of the interaction process as a whole arises, which cannot be tracked down to the individual mechanisms alone (De Jaegher and Di Paolo, 2008).

Emergent phenomena are thus not only found at the level of identity of simple cells, but also "at various levels in multicellular organisms and in particular animals and humans." (Di Paolo et al., 2010, p. 41). How these levels are interrelated has to be assessed on a case-by-case basis using for instance tools such as dynamical systems theory (ibid., p. 42).

Dynamical systems theory (DST) distinguishes between linear and non-linear dynamical systems. Cognitive systems and their constituents are described as *non-linear* dynamical systems. This holds for the nervous system, the interaction between brain and body, the system's interactions with its environment, including social interaction processes between cognitive systems. Unlike linear dynamical systems, the behavior of the specific non-linear system as a whole cannot be predicted exclusively by reference to the behavior of its individual component processes. This is because the interactions and component processes are non-linear so that the effect for a whole system cannot be obtained by adding the effects of its singular components (Rohde, 2011, p. 29, for a nice comparison between the dynamicist and traditional methodology see Thompson, 2007, pp. 40–43).

An empirical field in enactivism using DST for the study of cognitive systems as dynamical systems is evolutionary robotics (ER). It applies DST in computer simulations that model minimal cognitive dynamics. The models in ER are not used as detailed one-to-one models of human cognition (they are much simpler), explaining particular functions, but rather for generating hypotheses and proving concepts. The scientific function of generating such proofs is to bring about conceptual change and generate new hypotheses (Harvey et al., 2005, p. 84). Rohde has recently published a book that nicely illustrates the use of DST in an analysis of so-called perceptual crossing experiments. She there successfully models the assumed role of

[22] This is not to say that Bénard cells themselves are involved in cognitive processes. The example offers an analogy to illustrate the kind of emergence likely involved in cognition.

[23] Some people may worry that to claim that interactive autonomy is the emergent result of an interactive process between two individuals still allows for a localization of the cognitive, namely by pointing at these two individuals. What matters however for interactive autonomy are not the properties of two individual agents, but rather the interactive dynamics these agents are involved in. The new properties emerge only in their joint activity, they are not the sum of what individual agents do and can therefore not be localized by pointing at them.

global interaction dynamics in social interaction, thus providing empirical support for the concept of participatory sense-making (Rohde, 2011).[24]

2.3 The Enactive View of Cognition in Comparison

In the previous sections I have summarized the building blocks of the enactive perspective on the mind. Enactivism approaches cognition with three basic steps: First, it grounds the model of cognition in biology by assuming that the cognitive system bears organizational similarities with basic life forms in that it self-produces and maintains an identity (embodied center). Second, it holds that the production of autonomy necessarily implies a two-fold intrinsic purpose – with respect to identity preservation and meaning generation. This leads to the creation of a perspective on the world, which is specified as sense-making processes (center of activity). In a third step the relation between sense-making and autonomy is refined through the concept of adaptivity, i.e. the capacity to evaluate interactions with the world not only with regard to the system's survival, but also gradually with regard to various intrinsic norms.

In this section I make a further inter-theoretical clarification of the enactive framework by delineating it from the other three approaches to cognitive science, i.e. the orthodox view of cognition, the extended cognition approach and thirdly, the sensorimotor approach to perception. The differences of enactivism and the latter two are assessed even more thoroughly in the discussion of the body in the context of LIS and BCI in Chapters 4, 5 and 6.

2.3.1 Enactivism and the Orthodox View of Cognition

The enactive concept of cognition departs from the orthodox view with respect to all four dimensions introduced at the beginning of Chapter 1, i.e. *1)* the nature of cognitive systems *2)* cognitive processes, *3)* conscious experience and *4)* the relation of system and environment (see Figure 2.4 for a comparison).

1) Instead of assuming that cognition is computer-like enactivism suggests that the best model to understand cognition is the living organism. Living organisms are described as autonomous systems and an autonomous system is not an input-output device passively representing the environment but an agent. It evaluates its interactions according to its own inner norms, thus actively constructing meaning and shaping itself and the world (Di Paolo et al. 2010, p. 39). The very notion of "enaction" is exemplary for the radical departure of the enactive framework from the representationalist view of cognition.

2) Cognition does not consist in problem solving as the manipulation of symbolic representations but in constructing meaning with regards to intrinsic goals. Enactivism has a constructive epistemology and is clearly non-functionalistic. This is not to say that enactivists assume that only biological systems could be cognitive. Quite to the contrary, enactivism puts great effort in modeling and evolving artificial cognitive systems. However, they emphasize that the model should not be conflated with what it is modeling. Computer models are 'machines for

[24] Note that a lot more can be said about the concept of emergence in enactivism (see for instance Thompson, 2007, pp. 417–441). It would be useful to shed light on the relation of the enactive concept of emergence to other recent approaches to emergence, for instance, as they have been assessed by Stephan (2006). Stephan discusses three perspectives on emergence in cognitive science and philosophy of mind and argues that there does not exist a coherent notion of emergence applicable in both philosophy of mind and cognitive science. However, what counts as a coherent concept of emergence in cognitive science is dependent on a specific background paradigm and it seems that Stephan's notion is only applicable in light of a classical view of cognitive science. An elaboration of emergence against an enactive background theory might offer a concept of emergence applicable in both, philosophy of mind *and* cognitive science.

thinking' (or 'tools for thinking')" (Rohde, 2011, p. 4). This does however not mean that thinking itself is computer-like.

Functionalists do not deny that a functional profile requires a physical implementation, but they set out with a particular cognitive process and then argue that it can be realized in several ways. The main difference to the functionalistic view is that enactivists try to understand the *entire* system that is bringing about the cognitive process. Autonomy is an "organizational property of a system, not a function, a state or a mechanism" (Di Paolo, E. and Iizuka, H. (2008, p. 421). It focuses not on an abstract, single cognitive process, but on the embodied cognitive system as a whole and tries to understand how its particular embodiment will constrain and shape the actions and thereby the cognitive processes that arise with them. Enactivism thus rejects the "traditional poles of seeing cognition as responding to an environmental stimulus on the one hand, and as satisfying internal demands on the other" (ibid., p. 411). Recall from Section 2.2.2 that autonomy involves the notion of precariousness. A precarious autonomous system is dependent on particular material processes, yet it also strives to keep certain material processes outside its own organizational mechanism. It is genuinely concerned with the need to maintain its material integrity and thus "intervenes in its own substrate" (Di Paolo, 2009, p. 16). This is an aspect of cognition that cannot be captured in terms of functionalism because functionalism is essentially aiming for a "substrate-independent account" of cognition.

3) To enactivism subjective experience is a non-negotiable aspect of cognition. Inwardness is a necessary consequence of the generation of an autonomous identity (Jonas, 1966). In contrast to functionalism, the enactive approach is teleological, i.e. a cognitive system has its own intrinsic norms according to which it directs its sense-making activities and assigns meaning to particular structures in the environment. In contrast to the orthodox view, the cognitive system can therefore not be explained from an external perspective alone, but needs to be considered as being guided by the subjective experience of the cognitive system itself. For the same reason the classical explanatory gap in the philosophy mind, i.e. the question how the mental and physical as two entirely distinct processes are related to each other, is seen as a consequence of the dichotomy between mental and physical properties (or the subjective realms and the material body). Enactivism in contrast is not concerned with the relation of phenomenal properties and objective (material) properties in general, but rather with an explanation of how cognitive systems can bring about subjectivity through their sense-making activities (Thompson, 2007). The body-mind problem is transformed into the "body-body problem" and the question to be answered is how living bodies bring about an embodied experience as a kind of dynamical condition (Hanna und Thompson, 2003, Thompson, 2007, pp. 235–237). In other words, we are now asking how a living body as subjectively experienced relates to the lived body as material object. As Fuchs has recently summarized:

> The question now is about the relation between one's body as subjectively lived and one's body as an organism in the world. And the answer must be in principle that processes of *living* and processes of *living through (Leben* and *Erleben)* are both aspects of the life process seen from two complementary points of view. (Fuchs, 2011, p. 200)

From an enactive perspective the body-body problem is resolved on assuming that both, the objective and subjective body are integrated in a world-directed existence. Sinc subjective and objective levels of descriptions thus acquire a common ground there is no gap between them.

4) At the end of Chapter 1 I suggested that the orthodox view of cognitive science implies a *dichotomous, exclusive* and *reductive* epistemology. Enactivism supersedes the three kinds of dichotomies: a) the split between brain and body, b) the separation between cognitive system and environment and c), the split between conscious and cognitive processes.

CONDITION	THE ORTHODOX VIEW	THE ENACTIVE VIEW
metaphor	computer or ANN	living organism
def. cognitive system	input-output device	autonomous system
def. cognitive process	problem solving, manipulation of representations	sense-making
role of the body	contingent/negligible	necessary feature of autonomy
subjective experience	irrelevant	necessary feature of autonomy/ sense-making
epistemology	functionalist	constructivist
relation system-world	distinct, input-output, world is external, objectively given	coupled, no input-output, world is enacted by embodied interaction
explanatory gap	mind-body problem, relation experience/brain	body-body problem, how organism brings about phenomenality
norms	externally given, opposed by environment	intrinsic, basic norms: sustain existence, further norms (through adaptivity)

Figure 2.4: A comparison of the orthodox and the enactive approaches to cognitive science

For enactivism brain and body are components of a holistic network of processes that together constitute the autonomous system. What the world is depends on the structure and goals of the autonomous system. There is no ontological split between cognitive system and world, but rather a boundary as a systemic distinction enacted through the cognitive system's evaluation of particular aspects of the world as meaningful or not. Enactivism is in two ways *inclusive*. First, cognition begins not at the level of human beings, but is inherent to life and all of its forms. Second, a definition of a cognitive process is not limited to abstract thinking processes, but already begins at lower levels (e.g. the level of the cell) with directed and immediate interactions with the environment. Based on this evolve with growing complexity of autonomy increasingly more sophisticated cognitive processes. Enactivism is finally not reductive because it does not limit scientific explanation to the realm of unconscious, material processes. It incorporates the subjective and phenomenological dimension as a valid and non-negotiable aspect of cognition.

2.3.2 Enactivism and Extended Cognition

Extended cognition and enactivism are sometimes said to be similar because both posit a close interrelation of the embodied cognitive system and its environment (a key paper exploring their interrelation is Di Paolo, 2009, but see also Rowlands, 2009). This impression is superficial. Enactivism and extended cognition differ fundamentally. The most general difference has been implicit in the previous section. Extended cognition is a functionalistic framework and as such very close to the orthodox view of cognition. Many of the worries enactivists encounter with regards to functionalism will thus be equally present in their relation to extended cognition. We encounter more fine-grained differences between these approaches throughout the remainder of

the thesis. At this point, I want to clarify two possible misconceptions that can arise when trying to understand the extended and enactive cognition.

1) Boundary-Issues

Recall that extended cognition assumes that cognitive processes are not only located within body or head, but can, under certain circumstances transcend the organismic boundary and be constituted by external resources, such as a notebook or pen and paper. When coupled to such an external feature, the cognitive system is said to be an extended system, constituted by neurological, sensorimotor and environmental components. Since the enactivists quite similarly claim that cognition is not brain or body bound one might wonder where the difference between the two actually is. The answer, according to Di Paolo, lies in the notion of boundary (Di Paolo, 2009). According to him extended cognition confuses two different kinds of boundaries: the physical boundary of the system (of the organism) on the one hand and the boundary between what counts as cognitive and what not on the other hand.

In order to address the former, one needs to have an idea about the latter. Extended cognition does not offer an explication of the cognitive, but adopts a deep commitment to what we might call the principle of "location inheritance" (Kyselo, 2011, msp. 5). According to this, the nature of the cognitive (extended or not) is characterized by means of locating its realizing mechanisms. Almost all participants in the extended cognition debate – whether proponents or opponents of extended cognition – seem to make this confusing presupposition. Adams and Aizawa identify neural mechanisms as the constitutive vehicles of cognition. Cognition is therefore bound to the brain, and defined in terms of neural activity; it is "a species of causal processing involving non-derived content" (Adams and Aizawa, 2010, pp. 55, 70, 2009).[25] Rupert (2004, p. 10) in turn takes the whole system of brain and body to be part of the underlying mechanism of cognition. The locus of cognition is therefore found within the individual, i.e. the organism as a whole. Clark and other proponents of the extended view of cognition then cross the boundaries of brain and body and include certain external features. The general attitude within the debate seems to be this: take any possible cognitive process, make a claim as to whether or not you think it can be realized by an external feature and then decide that cognition is or is not extended. This polemic is to hint at a fundamental problem of the extended cognition debate. As Di Paolo aptly points out: "Before asking where it is we *must* first say what it is" (Di Paolo, 2009, p. 10). For positing that cognition extends into the world, and that any process can count as cognitive, were it functionally similar to something that is done in the head, the only guiding intuition is: cognition is what is in the head.

For enactivism cognition was never in the head in the first place. We have seen that the enactivist treats cognition as an emergent result, arising out of the adaptive interactions of an autonomous system with its environment. Cognition cannot be identified with or located within either of the component, not within the brain, body or the environment alone (but also not in combinations of them). This is not to say that we cannot localize components that contribute to the emergence of cognition. The point is that cognition cannot be localized in the components because it is not identified with the components. Cognition is identified with what their interaction brings about. A car, e.g., has a particular mechanism, consisting of engine, steering system etc., which allows the car to move along the road. However "driving on the road" cannot be equated with the mechanism that enables driving because it does not happen in these systems, but in relation to the road and other parameters. To argue that cognition is where its mechanism is analogous to arguing that driving along the road is located in a car's engine. Both assumptions are examples of the mereological fallacy, which attributes properties of the whole to one of its

[25] They do admit that extended cognition is in principle possible, but argue that "as a matter of contingent empirical fact" cognition takes place in the brain (Adams and Aizawa, 2009, pp. 55, 70).

parts.[26] The mereological fallacy is particularly problematic when the properties of the whole are relational. Since for enactivism any cognitive process is a novel and, crucially, relational property resulting from the dynamical interplay of its component processes, it has no location. The mechanism enabling a cognitive process and the cognitive process itself have a different ontological status. The mechanism may be located, but since the cognitive process acquires its ontological character as a relation[27] it would unintelligible to assess it by looking at the physical location(s) of the individual component(s).

According to enactivism a cognitive system's boundary is therefore not specified by reference to its underlying vehicles, whether or not these extend beyond the organism. It should be posited between those processes, components or phenomena that are included within the organizational network (comprising neural, embodied and interactive agent-world process) of the cognitive system, and those that are not (see Figure 2.2.2). I get back to the problem of boundaries issue in Chapter 6 (Section 6.5.4) as well as in Chapter 7.

2) On Bridging the "Cognitive Gap"

The life mind continuity hypothesis constitutes one of the biggest worries for the enactive approach. While it establishes a good alternative to classical representational views on cognition for *low-level* processes it is arguably difficult to see how it may do so for cases of higher-level human cognition. There seems to be a "cognitive gap" between action and perception, which are explained by means of sensorimotor engagements on the one hand, and the level of more complex cognitive phenomena in human beings (De Jaegher and Froese, 2009) on the other hand.

It appears that extended cognition actually deals with a very similar problem and also provides a possible route to resolve. Extended cognition suggests that cognition is partly constituted by extra-neural and extra-bodily processes. In this way, rather simple neuronal workings are augmented by exploiting bodily and environmental structures, leading to the more complex cognitive processes we find in human beings (Clark, 1998).

However, as De Jaegher and Froese point out, Clark and other proponents of the extended view of cognition were mainly concerned with abstract phenomena such as language and technology. The problem is that this focus on already complex features such as language and interaction with technologies prevents us from understanding that the cognitive mechanisms involved at this sophisticated level of cognition are already present in simpler cases of cognition (De Jaegher and Froese, 2009, p. 445).

According to De Jaegher and Froese there is no cognitive gap. It only arises as the result of a so-called "methodological solipsism" (ibid., p. 444) that is still predominant in cognitive science. Previous approaches assuming that cognition involves environmental processes ignored that complex cognitive processes are not only shaped by interactions of the individual system with the environment, but also have to be accounted in terms of inter-agent relations. This applies not only for the abstract processes discussed in the work of Clark and others, but also for more basic processes such as coordination (ibid.). In line with what we have learned about the concept of participatory sense-making, De Jaegher and Froese suggest that there is an emergent autonomy of interactions between agents which influences the individual cognitive mechanisms and enlarges their sense-making capacities. Human cognition is thus socially mediated (ibid., p. 456, see Section 2.2.4). This, the authors believe, advances our understanding of how we can get

[26] The mereological fallacy has been discussed in different places. Bennett and Hacker (2003), for instance, discuss a case of (falsely) ascribing properties of a whole to its components in the context of cognitive neuroscience; mental properties of the whole living animal are attributed to the brain, which is, however, only a part of the animal.

[27] We have seen this in sense-making where meaning is generated based on the interaction dynamics of the autonomous system interaction and the environment.

from basic cognitive processes to "higher forms of mental life" (ibid.) and is the key to transcend the presumed "cognitive gap".

Note however that this is not quite sufficient to differentiate extended from enactive cognition. De Jaegher and Froese's objection to Clark does not rule out that extended cognition could account for the coupling of biological systems with the environment as a basis for *higher-level* cognitive processes. Their point is that such intimate coupling between cognitive system and environment is not limited to these abstract processes, but begins with the simplest cognitive behavior. On this account both approaches would seem compatible: enactivism and extended cognition are the two ends of a spectrum of approaches that assume that cognition is based on coupled interaction of brain, body and environment. This is however misleading. In what follows I show why the extended approach is in fact not bridging a cognitive gap, not even at higher levels of cognition.

In order to explicate what they mean by "methodological solipsism" De Jaegher and Froese remind us of the history of cognitive science: For behaviorism environmental stimuli and the behavioral output were of crucial importance, while what was going on internally was completely irrelevant. With the rise of cognitive science this was turned upside down: the environment got completely out of focus and scientists were entirely preoccupied with accounting for what was going on inside the cognitive system, i.e. the brain.

With cognitive science's focus on the internal mechanism of cognition (the brain) it became difficult to grasp the transition from basic biological to more complex varieties of autonomy because the differences in the mechanisms of human brain and simpler animals are just tremendous. From such an internalist perspective on cognition emerges a radical version of the cognitive gap: how could we make sense of the idea that the inner workings of human cognitive system could be similar to that of much simpler systems, such as insects?

Proponents of extended cognition have similarly criticized this focus on internal processes. But their criticism is different from the internalist worry just explicated. The extended cognition objection has to be seen in the light of their commitment to functionalism, i.e. in terms of a critique of the identity theory of neuronal and mental processes. According to this, what matters for cognition is not the particular physical make-up but the implementation of a specific functional role. The extended view widens the realm of what kinds of systems could possibly count as cognitive by claiming that any creature is capable of cognition as long as it has the material set-up needed to realize it.

On this account it seems quite obvious why environmental features pose no problem for extended cognition – they are simply components of a larger supervenience base for the cognitive system. But while this liberalism constitutes a probably nice counter against species-chauvinism, it is at the same time inflationary and does not explain higher-level cognitive capacities. Why is that? It is not merely that extended approach misses the point that the same mechanisms working at the level of language and particular technology interactions may already be found at the lower level, as De Jaegher and Froese have it. The problem is rather that extended cognition does not even require any claim about the continuity between basic biological and more complex environmentally based cognitive processes. The motivation behind the extended mind hypothesis is not a proof of the continuity of life and mind, not even a limited one (with regards to language and abstract thinking processes). This is because extended cognition makes no proposal as to what the basic mechanism is that we later find in a more complex mechanism[28], nor provides an argument for a possible connection or direct transition between cognitive systems of different complexity. The complexity of higher cognitive processes in terms of extended cognition is based on the primordial acceptance of the vehicle-liberalism inherent to functionalism. It is not explained by reference to the diversity of the vehicles, which explains nothing. To repeat in contrast the enactive perspective: it explicates the mechanism and principles which are at work at lower levels of cognition and finds that the principles in rather

[28] Except, maybe, that it is computational in nature.

simple life forms such as bacteria are applicable to all cognitive systems. The relation between the generation of a metabolic identity and normativity for instance can be equally observed in the bacterium swimming up a sugar gradient. It is also found in more complex biological systems as well as socially mediated systems, as we see in my elaboration on human autonomy in Chapter 7. All of these systems similarly sustain their systemic identity and need to adapt their sense-making interaction with the environment in a way that this goal is respected. Enactivism makes suggestions as to how these principles can combine or develop in order to explicate higher-level phenomena of cognition not present in less complex systems. One example for this is the above-mentioned widening of the realm of possible sense-making engagements through participatory sense-making. I provide more support for this claim in Chapters 6 and 7 where I discuss bodily autonomy and elaborate on the several interrelations that can hold between autonomous systems. I now briefly explicate the differences between enactivism and the sensorimotor approach to cognition.

2.3.3 Enactivism and the Sensorimotor Approach

It is important to make a distinction between enactivism and the so-called enactive approach to perception, also known as the sensorimotor contingency approach. The latter uses the label "enactive", but is in fact not enactive in the sense introduced above. Generally speaking, enactive perception is concerned only with a *part* of the cognitive, while the enactive approach constitutes an entire framework for the whole of cognition. One may now ask why not subsume the enactive approach to perception into that particular overarching framework?

There are at least two reasons why this option appears rather unlikely at the moment. First, in enactivism, the label "enactive" is in inextricably linked to the notions of autonomy and sense-making and relies on the assumption of the life-mind-continuity thesis. This does not apply to Noë and O'Regan's account. Their approach to explaining visual experience relies only on the mastery of sensorimotor knowledge. But this already requires the existence of a cognitive system "a knower or agent or self" that masters this knowledge (Thompson, 2005, p. 415). Noë and O'Regan do not explicate what the cognitive system that relies on sensorimotor contingencies actually is. Enactivism in contrast has a definition of the cognitive system as an autonomous system, i.e. a self-sustaining and operationally closed network. The exercise of sensorimotor knowledge is not just some aspect of cognition, but directly linked to the preservation of the system's identity itself.

Note furthermore, that in Noë and O'Regan's approach the knowledge of the system is ascribed to the cognitive system from the outside. It is not knowledge of the system itself (ibid., p. 419). Here lies the second reason why the sensorimotor approach and enactivism are not easily compatible. The sensorimotor approach lacks a proper account of subjectivity itself (Thompson, 2007). In contrast, enactivism provides an account of basic subjectivity by introducing the notion of autonomy as the ground of cognition, i.e. normativity and integration of sense-making. The autonomous system is concerned with the maintenance of its identity and all of its goals are dependent on this. As a consequence one cannot make sense of the employment of particular cognitive capacities without linking them to system's subjective evaluation of a given situation. We should thus not only account for particular kinds of (perceptual) experience, but must also address the possibility of subjectivity as such (Thompson, 2005, p. 420).

To sum up, enactivism and the sensorimotor approach both rely on the embodied (sensorimotor) exercise of skills. In that sense they may seem compatible (but see Hutto, 2005 for a critique) and could be linked into a fruitful synthesis at some point (Thompson, 2007, p. 265, pp. 260–265). Yet, this would require to either complement Noë's approach with the very basics of enactivism or to exempt it from theoretical underpinnings that are clearly non-enactive in nature. This has not been explicitly done yet and remains a task of future research. I contribute to this task in Chapter 6 where I provide steps towards an integration of the sensorimotor approach in the enactive approach with regard to the concept of embodiment.

2.4 Concluding Remarks

One of the main questions of the present thesis is *what is a human cognitive system*? In the last two chapters I provided a brief overview of how this question has been answered throughout the history of cognitive science.

We have learned that in early cognitive science cognitive systems were seen as disembodied information processing devices. This view shifted with the beginning of embodied cognitive science. In addition to inner neurological processes, bodily and environmental processes were considered to be important for describing the nature and constitution of cognitive systems. In contrast to the orthodox view, we now also find that part of the description of cognitive systems should account for conscious experience. The phenomenological realm is considered to be complementary to the mere mechanistic and physical description of cognitive processes.

Now even though I criticized the account of cognitive systems offered by orthodox cognitive science as *reductive*, *exclusive* and *dichotomous*, it did provide a more or less unified (computational) story about the cognitive and strong tools to enable its scientific inquiry. Unfortunately, one cannot say the same about recent approaches to embodied cognition. Surely, the developments in robotics and dynamical systems theory are reflections of a radical change and they offer stunning alternative strategies of modeling cognitive processes by considering bodily and environmental capacities and without relying on a central executive organ or detailed planning procedures. The *Embodied Mind* can be seen as a first attempt to bring the alternative and new aspects of cognitive science together in a complementary perspective on cognitive systems.

However, since *The Embodied Mind* embodied cognitive science developed further and grew even more complex. There is not *one* embodied cognitive science, but rather a loose and multifaceted field, with focused elaborations on particular aspects of cognition. I have tried to highlight some of them and outline, in comparing the different approaches of embodied cognition to the orthodox view, some of their possible linkages. These linkages require more detailed elaboration, but we can already make two general observations. First, that the extended view of cognition is relatively close to the orthodox view and that this closeness puts it at distance to most other approaches of embodied cognitive science. Second, the enactive approach appears to be the only proponent of embodiment that offers a more unified picture of what a cognitive system. It grounds the various new characteristics of the cognitive of the different embodied approach in autonomy and the mind-life continuity thesis. It falls however similarly short to specify in more detail the most central tenet of embodied cognitive science, namely that cognition relies on the body. Embodiment, the sole common denominator of embodied cognitive science, appears to be generally underdetermined, a placeholder bare of any unified commitment of what "the body" amounts to.

"Embodiment" obviously happens to be the key concept towards the development of a coherent and valid alternative story to the orthodox view, as it came into some kind of focus of *every* alternative view of cognition. But if embodied cognition is supposed to provide a truly alternative description of the cognitive system the task remains to spell out in detail what the body is and to what extent it is relevant for cognition. This does not only require a clarification of the concept of the body, but also to explicate its linkages to other properties or aspects of cognition.

I suggest in this thesis that one strategy to approach this task could consist in confronting embodied cognitive science with a theory-neutral "empirical touchstone", such as LIS. I explicate in the following Chapters 3 and 4 how and why this bodily condition poses a basic challenge to cognitive science and a particular challenge to embodied cognitive science.

Chapter 3.

Locked-in Syndrome and BCI – a Basic Challenge to Cognitive Science

> I'm just a typical mind imprisoned in this body, I feel as though I'm encased in concrete that I'm constantly and painstakingly breaking through ever so slowly. (Nick Chisholm in Gillett and Chisholm, 2007)

Locked-in Syndrome (LIS) is a special condition because it appears to bracket and straightjacket an *entire human existence*. A body without movement, a person that is disconnected from the social sphere – how can we make sense of this? The radical character of the condition urges to look more closely at aspects of our existence we have always taken for granted or thought to have understood thoroughly. This is the reason why I believe LIS to be of particular relevance for research in cognitive science. I suggest that we should conceive of it as a *basic challenge* to our understanding of the nature and constitution of (human) cognition. It has been acknowledged by recent writings in philosophy that LIS is a special condition. Unfortunately however, no one so far has explicitly formulated the different challenging dimensions of LIS.

While the previous two chapters were rather theoretical and mainly concerned with the philosophy of cognitive science in general, this chapter is more empirical in nature. It considers research in (neuro)medicine and neurotechnology and discusses philosophical approaches in application to concrete cases of bodily impairment. The major purpose of this chapter is to explicate that LIS poses a *general* challenge to cognitive science and can help to reappraise our intuitions about the nature of human cognition.

This chapter constitutes the other half of the background for the present thesis. It is divided into three parts. First, I give a brief introduction to LIS and the neurotechnologies that facilitate and enable communication and specific activities the patient could not pursue otherwise. I then summarize and discuss philosophical approaches to LIS and other cases of bodily impairment in the context of sociality and identity (Cole, 2009, Dudzinski, 2001 and Gillett and Chisholm, 2007).

Finally, I discuss work by Kurthen et al. (1991) who consider LIS as a challenge to the assessment of consciousness. As a consequence of the first philosophy section I formulate the first part of the basic challenge, i.e. that before we assess the role of bodily impairment on the life of a patient we need to clarify the concepts at use in this assessment. The discussion of Kurthen et al. (1991) makes explicit a second dimension of the basic challenge to cognitive science: We should make efforts to clarify and define our concepts. But in order to determine their interrelations and relevance for cognition and consciousness, we should also make explicit our basic background epistemology.

This twofold basic challenge serves as the basis for more specific challenges to *embodied* cognitive science that I formulate and discuss in the following Chapters 4–6. It also sets the ground for the elaboration on self as socially enacted autonomy, which I introduce in Chapter 7 and further apply in the final Chapter 8.

3.1 Locked-in Syndrome and Brain Computer Interfaces

In this section I provide an introduction on the condition of LIS, as well as the use of Brain-Computer Interfaces (BCI) for facilitating communication and other capacities. Some of the empirical facts addressed below are elaborated on and considered more thoroughly in Chapter 4, but the reader is invited to return to this introduction whenever needed.[29]

3.1.1 General Information

LIS is a severe condition in which a patient's body is almost completely paralyzed. There is almost or no movement of voluntary muscles, preventing the patient from communicating either verbally or non-verbally (Plum and Posner, 1966, Bauby, 1997, Laureys et al., 2005). Despite the severe physical impairment, hearing is preserved and the body still "passively" sensible (Perrin et al., 2006). LIS patients are considered to be fully conscious and cognitively unaffected. Their minds are described as 'trapped' within their bodies (Bauer and Gerstenbrand, 1979, Inci and Özgen, 2003, Laureys et al., 2005).

Causes and States of Locked-in Syndrome

In most cases, LIS follows from a stroke or an accident leading to brain or spinal trauma with severe ventral pons lesion causing the paralysis (León-Carrión et al., 2002, p. 571). The brainstem lesions usually do not affect the vertical eye movement and the blinking capacity. But LIS might also result from a rare neurological disease, Amyotrophic Lateral Sclerosis (ALS). ALS is a neurodegenerative disease. Whereas patients suffering from a brain or spinal trauma find themselves suddenly paralyzed, in ALS the paralysis is of a progressive nature. The motor neuron (the motoric neural system) of the patient will successively degenerate. At the beginning, only the limbs are affected, but as the disease progresses the respiratory system ceases to function and eye movements become impossible.

Three states of LIS have been classified (Bauer et al., 1979). The *incomplete,* where the patient's body is completely paralyzed, except for head- and sometimes finger movements. Secondly, the so-called *classical* state (CLIS for classical locked-in syndrome), in which the patient's body is completely paralyzed, except for vertical eye movements and her ability to blink. And finally, the *complete state* (TLIS for totally locked-in syndrome) in which their body is completely paralyzed and not even eye movements are possible (see Figure 3.1). This state usually coincides with the final states of ALS. In Chapters 4–6 I refer to these cases and discuss how they challenge different conceptions of embodiment.

[29] I deliberately chose to introduce some detail on the functioning of BCI technology only later in the discussion of the body concepts. This is because the confrontation with cases of LIS raises novel questions that require particular consideration of empirical facts.

state	incomplete	classical	complete
movement	head, fingers, eyes	eyes (blinking, vertical movement)	no movement, eyes closed
BCI	yes	yes	only with *prior* training (before TLIS)
cause	ALS/stroke	ALS/stroke	ALS

Figure 3.1: States, causes and BCI-use in LIS**.** The LIS-state is defined based on the kind of movement the patient can still perform. In complete LIS no movement whatsoever is possible. This condition usually coincides with the final stadium of ALS.

Diagnosis and Cognitive Functioning

Clinical neuroscience distinguished LIS from a range of other states of severe brain damage leaving the patients non-communicative. These states are classified with respect to the preservation of two physiologically measured aspects of consciousness: arousal and awareness of self and environment (Laureys et al., 2007, see figure 3.2). Coma is situated at the lowest level, with complete absence of both arousal and awareness – coma patients are unconscious. Patients in the so-called vegetative state are awake, but usually show no sign of awareness of themselves or their environment. Cases of minimal consciousness are classified above vegetative state because the patients show limited signs of awareness, and sometimes, though inconsistently, communicate. In LIS both arousal and awareness are preserved. With respect to consciousness, LIS patients are classified close to the "normal" cases of awake, "healthy-bodied" people. We see in a moment that this classification is questioned by Kurthen et al. (1991) who argue that there is no way to assess consciousness in patients with complete LIS.

Given that most patients with LIS are not able to communicate and seem unresponsive, the diagnosis of LIS is difficult. Gillett and Chisholm (2007) report the case of a LIS-patient who "interacts with his environment…in as many ways as his impairments allow despite the fact that only a few people realized it" and he "recognized objects and people, conceptualized experiences and experienced a range of feelings" (Gillett and Chisholm, 2007, p. 2). Often the condition is not recognized for a long time and patients were misdiagnosed as being in a coma, vegetative state or as suffering from akinetic mutism (Bruno et al., 2008, p. 883). Some patients have had to wait several years before they were diagnosed as being locked-in (Laureys et al., 2005).[30] Usually it is a member of the family who notices that the patient is "still there" (Gillett and Chisholm, 2007, León-Carrión et al., 2002, p. 573, Laureys et al., 2005).

[30] A prominent example for this is the by public media controversially discussed case of Rom Houben whom doctors considered to be unconscious and comatose for over two decades until Laureys recently dismissed the diagnosis and claimed that the patient has actually been consciously locked-in all along. Laurey's diagnosis has not remained undisputed as the patient's personal reports about his reanimation turned out to stem from his assistant and not himself (e.g. Dworschak, 2009).

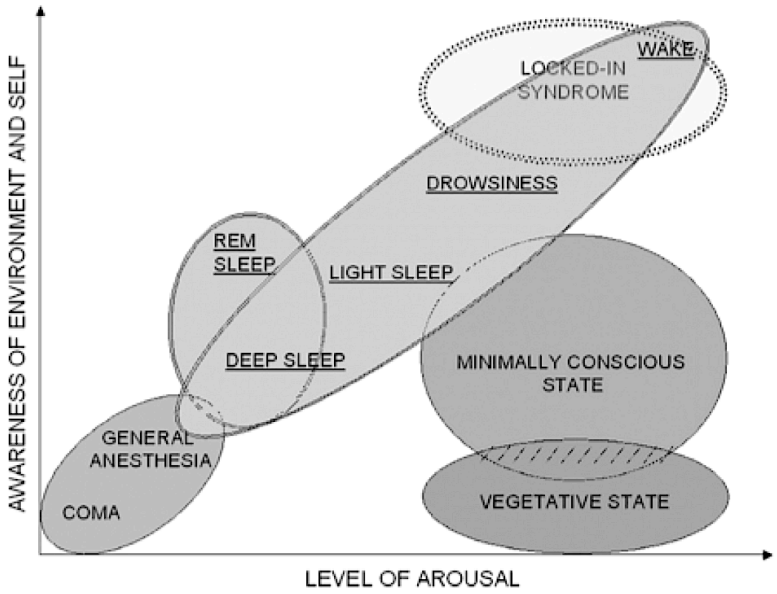


Figure 3.2: Differentiation of states of consciousness. With regard to the level of awareness and arousal LIS is located close the normal wake state, in the upper right part of the diagram (from Laureys et al., 2007).

León-Carrión and colleagues report in their 2002 survey of 44 patients with LIS that cognitive function in LIS "is excellent" (León-Carrión et al. 2002, p. 577). However, in a later article the same authors admit that they "cannot say that cognition is completely intact" – patients show deficits in attention, temporal orientation and memory (either of which occurred in less than 20% of all cases) as well as language. But León-Carrión et al. assume that these deficits result from the patients' changed emotional states and mood (León-Carrión et al., 2005). They also state that LIS patients are able to read, watch television and participate in social or family activities. Schnakers and colleagues from the COMA science group in Liège similarly reported only marginal cognitive impairments in LIS (Schnakers et al., 2008). In a series of neuropsychological tests, they assessed "major cognitive functioning", such as short- and longterm memory, attention, capacities to adapt to novel situations as well as language intelligence (ibid., pp. 2–4). The authors conclude that there is an "absence of extensive cognitive deficits" in LIS. Lower cognitive performance of subjects with LIS is due to fatigue or additional neurological causes but not due to LIS itself.

These findings suggest that even in cases of global bodily impairment, cognition is generally not affected. In the next chapter I show that this constitutes a challenge to proponents of embodied cognitive science according to whom cognition in fact depends on the body.

Quality of Life

It is often assumed that the quality of life of the severely paralyzed patient is poor and her life not worth pursuing. But reports of the quality of life with LIS are in stark contrast to this intuition. The patients' assessment of their situation is not significantly different from normal subjects once they develop strategies for coping with the disability and adapt to their new situation.

Instead well-being seems to depend on the "subjective feeling of control over one's life and the feeling of a purposeful life irrespectively of the actual physical conditions" as well as one's ability to communicate with others, i.e. engage in social interactions (Lulé et al, 2009, p. 344.). We should recognize that medical interventions affect "patients' lives rather than just […] their bodies" (Bruno et al., 2008, p. 884) and Gosseries et al. demand that medical doctors "should realize that quality of life often equates with *social rather than physical* interaction"

(Gosseries et al. 2009, p. 199, my italics). These studies emphasize the subjective perspective of the patient and that bodily impairment is not only a dysfunction of an objective and biological body, but an impairment of the subjective body, the body as *lived*. This is similarly posited in the phenomenological observations below in Section 3.2. and I discuss the relation between objective and subjective body throughout Chapters 4–6. Crucially, the studies also shed light on the role of other people and their support for the patient's self-experience. This foresees the general line of reasoning of my thesis: even from a phenomenological perspective that acknowledges the body as lived, we cannot fully account for the patient's overall existence (end of Chapter 6). In determining who we are, we need to consider the role of other subjects. Human cognitive systems are embodied, but they are also socially embedded systems (Chapters 7–8). This might explain why even extreme bodily restriction does not have to lead to an impairment of cognition and a bad quality of life. As long as patients are embedded in a social context their cognitive identity as someone amongst others is preserved. As I show later this allows for a more elaborated account of *why* the LIS patients evaluate their life situation positively (see 7.3.4).

3.1.2 Communication and Brain Computer Interfaces

LIS prevents the patient from communicating verbally (by speech) and non-verbally (by gesture), placing an uncontrolled distance[31] between herself and her social environment. Patients with intact eye-movement have limited means to express themselves based on blinking-movement. However, recent developments in BCI may help to narrow that distance (Wolpaw et al., 2002, Birbaumer et al. 2002, Birbaumer, 2006). BCIs make use of brain activation patterns voluntarily produced by the patient to control external devices such as a wheelchair, or, in the case of communication, a cursor for selecting letters on a screen. In the following, I briefly explicate how LIS patients can communicate with intact eye-movements and then summarize recent developments in BCI technology.

Eye Code

LIS patients in the classical state can use eye-movements to communicate by means of an eye code (Laureys et al., 2005). This can consist in either vertical eye movements or blinking, or a combination of both, but it always requires the help of another person. Patient and interlocutor have to agree on a shared code, coherently used in every communication. The code can e.g. consist of indicating a "yes" by looking upwards and a "no" by looking downwards or of blinking once for a "yes" and twice for a "no". This binary code allows for restricted conversation based on closed questions. The French journalist Bauby used the eye-blinking code to write his book.

BCIs

Recent technological advances in BCI provide a promising means to improve the situation of patients with LIS. BCIs are a "rehabilitation tool" (Dobkin, 2007) to facilitate and increase the speed and flexibility of communication, re-establish control, and enable a variety of activities that they could not pursue without the technology (for a comprehensive review of BCI technologies see Birbaumer and Cohen, 2007).

BCI-use neurophysiological signals recorded from the patients' brains – usually EEG oscillations or event-related brain potentials (ERP) – to operate external devices or computers (see figure 3.3). They can be invasive or non-invasive, recording either from within or outside the patient's skull. Non-invasive BCI-technologies are usually based on recording of slow cortical

[31] "Uncontrolled" because the patient has no possibility to voluntarily establish a distance. In "normal" communication we can express our wish to get away from a particular situation or just go away. The patient however is forced to stay in the situation, whether or not she desires to. Thanks to Elena Cuffari for emphasizing this aspect (which is in line with Dudzinski's observations below, see 3.2.2).

potentials (SCP), sensor-motor-rhythm (SMR) or P300 waves. The former two require extensive periods of training where the patient is instructed to produce specific target patterns of brain waves in order to achieve voluntarily control of her own brain activity. The production of "operant" SCPs enables patients e.g. to communicate by selecting letters displayed on a screen in front of them (Birbaumer and Cohen, 2007).

An alternative to SCPs and SMRs control are so-called P300-BCIs, event-related brain potentials induced by spontaneous presentation of surprising stimuli. P300-BCI has the advantage that patients do not require any prior training. A recent elaboration on P300-BCIs allows patients to learn painting with their brain-activity (Kübler et al., 2008).

Figure 3.3: Communication facilitated by EEG-based BCI technology. The patient composes words and sentences based on voluntarily produced brain-activation patterns read of by the computer (from http://www.mp.uni-tuebingen.de/mp/index.php?id=137).

ALS patients who received training were also able to maintain communication after they became fully locked-in. Unfortunately, patients who did not receive BCI training prior to being completely locked-in did not show any response to training attempts. Birbaumer and Cohen (2007, p. 626) explain this by the lack of "the contingency between a voluntary response and its feedback or subsequent reward", which is presumably required for learning to use the technology. In Chapter 5 I will discuss this claim in terms of Merleau-Ponty's approach to bodily intentionality (see 5.2.2).

A general problem with BCIs remains the great variability in user performance. They are still far from offering solutions that are "at once practical, reliable and capable of high-speed complex communication" (Wolpaw, 2007, p. 614). Despite these shortcomings, it should be clear that BCIs are of crucial importance for physiologically impaired patients. Without this technology patients would not longer be able to engage in communication (or only with heavy restrictions) and exert skills such as email correspondence, web-browsing (Bensch et al., 2007), painting (Kübler et al., 2008) or the movement of external devices (Schwartz, 2007).

With regards to the present thesis it is important to note that BCI-technology is providing the patient with a new skill (Birbaumer et al., 2003) that neither relies on the biological body nor on muscular-based movement. Throughout the following Chapters 4–6 we see that this poses a range of challenges to current concepts of embodiment, which they presently seem unable to account for.

3.2 Locked-in Syndrome from a Philosophical Perspective, I

> I acknowledged the physical condition, but I would never accept its social limits or tolerate the eclipse of my past – nor should any disabled person do so, for it is the very basis of his or her struggle for life. (Murphy, 1990, p. 82)

In the previous part of this chapter I have provided a brief overview of LIS from a neuromedical and neurotechnological perspective. Some of the researchers mentioned above also theorize about the consequences and implications of the condition for a patient's life. Laureys et al. e.g. explicitly asked the (for philosophy of mind so prototypical) "What is it like"-question (Laureys et al., 2005, Laureys and Boly, 2007) even though they are not concerned with epistemological but practical considerations. They focus particularly on ethical concerns with regards to the quality of life and the question whether a LIS patient's life is worth living. In their approaches, to account for consciousness and self-consciousness of patients with LIS is relevant for diagnostic purposes as well as for "therapy and (end-of life) management" (Laureys et al., 2007, p. 722). Importantly however, whenever topics like consciousness and self-consciousness are considered they are usually addressed in a strict 3rd-person manner, relying on neurological and behavioral parameters alone (Demertzi et al., 2008). This is reflected in a definition of consciousness as comprising *arousal* and an "*awareness* of environment and of self" (Laureys et al., 2007, p. 723, my italics). Arguably, this awareness implies a reference to the first-person perspective, i.e. "to the thoughts and feelings of an individual" (ibid.). Note, however, that awareness in this context is tested for based on

> careful and repeated examination of spontaneous motor behavior and the patient's capacity to formulate reproducible, voluntary, purposeful and sustained behavioral responses to auditory, tactile, visual, or noxious stimuli (ibid.).

The assessment of this awareness ultimately happens on the grounds of external considerations and thus from a third-person perspective alone. Crucially, it also considers the patient from an individualistic perspective, as an isolated existence. It does not consider the possible role of the social environment for the patient.

There is a (small) amount of philosophical work on LIS and (a greater) on less severe bodily impairment that accommodates the patient's subjective perspective and elaborates more concretely on the importance of social interaction or the impact of bodily impairment on the patient's capacities to engage with others. The purpose of introducing these papers in the following is twofold. I want to provide the reader with concrete ideas about the kinds of philosophical questions and topics that may arise when considering LIS. The authors offer some valid and important phenomenological observations about the role of the body in LIS or the relation between bodily impairment and identity. I do not intend to raise serious objections to any of their claims in particular. However, I will argue that in order to adequately address them, a clarification and elaboration of the concepts used in these approaches is required.

3.2.1 Locked-in Syndrome and Impaired Embodiment

In a recent paper entitled "Impaired embodiment and intersubjectivity" Jonathan Cole[32] conveys a *phenomenological* analysis of personal reports from people suffering from three different kinds of neurological impairment (Cole, 2009).[33]

First, Cole discusses the case of IW who is unable to move without consciously planning his movements. But by watching and "thinking about" the movement IW had trained and regained his bodily capacities for locomotion and instrumental actions (Cole, 2009, p. 345). More important than being able to move on his own was thereby that he also relearned to gesture, i.e. how to react bodily and express himself in social interactions. IW imitated other people's expression and recovered slowly, within a period of thirty years, a high degree of automaticity in his gesturing movements. This arguably improved his communicative skills and positively impacted on his self-confidence as well as his "social success" (ibid.).

According to Cole, the case of IW is evidence for the crucial role of embodied communication for a person's self-esteem, social recognition and success (ibid., p. 343)

Second, Cole discusses a number of cases with irreversible paralysis. Arguably restricted capacities for locomotion and instrumental action have a strong impact on both the patients' self-experience and their interactions with others. Spinal cord injury affects the person's ability to show themselves "as a whole" to others (ibid, p. 358) and thus also how a person is seen by others. Other people perceive the patient as disabled and inferior and thus alter their behavior towards the person. This leads to a decrease in the self-esteem of the person: "Damage to the body, then causes diminution of the self, which is further magnified by debasement of others" (Murphy, 1990, p. 93).

Cole's observations are in accordance with the abovementioned reports of the quality of life. The problem with paralysis is not merely the medical condition, but the *society*. "[D]isablement is nothing to do with the body. It is a consequence of social oppression" (Oliver in Cole, p. 348) or "a disease of social relations" (Murphy, 1990, p. 4). A paralysis is "not simply a physical affair…but an ontology, a condition of our being in the world" (ibid., p. 90).

These considerations suggest that bodily impairment is not merely a restriction of a biological body but also of lived bodily existence and that apparently this existence is linked to the social environment and the ability to engage with it. However, this linkage is not made explicit and requires prior elaboration.

Third, Cole discusses a group of people with Moebius syndrome who entirely lack facial muscle movements. Facial movements are a vital component of human communication (Ekman, 1993), they accompany speech and help to express our emotions or our state of mind. The ability to imitate facial expressions is present already immediately after birth (Field et al., 1982). Babies with Moebius however are not able to exhibit facial responses to show to their parents "that someone is home" (Cole, 2009, p. 351). This negatively affects early infant-parent relationships. But it is also negatively affects how the social environment perceives the person in later stages of the development. Due to the lack of facial expression and uncontrolled eye movements "people might imagine that…a child might not be that bright" (ibid., p. 351). Similarly to the above cases of paralysis Moebius affects how people perceive and react to the patient. A completely immovable face is alienating which is why the patient may receive "reduced signals from others" leading to "emotional impoverishment" (ibid., p. 354).

---

[32] Cole is a medical doctor, not a philosopher. He has worked on these topics also in close collaborations with phenomenologists like Shaun Gallagher applying phenomenological concepts such as the distinction between body image and schema to medical cases (e.g. Gallagher and Cole, 1995).

[33] LIS is notably not a condition Cole discusses. The reason why I would still like to introduce these other conditions is that they provide an idea about the impact of bodily impairment in general. They help to raise the awareness of the radical character of the impairment in LIS as they each exhibit physiological restrictions that in LIS are *all coming together*.

The role of the social environment for emotional experience seems even more evident in the case of Lydia, another Moebius patient. Lydia observed that Spanish people are much more expressive and dramatic in their communication than people in her home country (UK). She thus began to imitate their facial movements and the gestures accompanying their speech. As a consequence, she began to experience emotions she had never felt before. Cole suggests that these experiences "were not…within and about her alone" but that they "emerged within a rich, social, cultural world" (p. 356). Lydia's emotional repertoire is therefore not an individual matter but enacted through being with and learning from others.[34]

Similar to the previously mentioned paralyzed patients, people with Moebius also report a sense of alienation. Celia e.g. experiences herself as a "collection of bits" (ibid., p. 352) and a split between her physiological and mental existence. These experiences suggest that physiological impairment interrupts the immediacy and transparency of body (Fuchs, 2005). The body has turned into an observable object, distinct from but still belonging to ourselves. According to this, there seems to be a linkage between bodily alienation and the social environment. But what this linkage consists of yet needs to be made explicit.

3.2.2 Locked-in Syndrome and Identity

Dudzinski (2001) pursues a phenomenological enquiry of the relation of body and *self* in a case of classical LIS. She bases her analysis on the autobiographical reports of Dominique Bauby's *The Diving Bell and the Butterfly* (Bauby, 1997). According to Dudzinski, Bauby's case shows that and how an altered body condition leads to a change in the identity of a person. Bauby has two identities, one before the stroke and a second after awakening from the coma when he finds himself completely paralyzed. In assessing the impact of LIS on Bauby's identity, i.e. the differences in experience before and after the stroke, Dudzinski relies on phenomenological work by Gabriel Marcel and his follower Richard Zaner.

According to Marcel the body is directly tied to the experience of oneself as somebody. This is expressed in the notion of *noyau senti*, the "felt potential" and the "readiness to do and to be" (Dudzinski, 2001, p. 35). It is our body that implicitly gives us this experience of an "oriental locus" and allows us to express ourselves, our wishes and our choices. In "normal" situations – that is, when having an able body – we are not aware of the body's role as the enabler of our actions and intentions. But through the impairment imposed by LIS, the existence of the *noyau senti* becomes terribly evident.

Dudzinski suggests that Bauby's *noyau senti* is "truly crippled" as his "felt potential" and the expressive capacities of his body are no longer given. Because other people now perceive him "as an object in their world to be ignored or to be noticed" he experiences his body "uncanny" (ibid., p. 37). Dudzinski makes an observation quite similar to Cole. Bauby's impairment has an impact on how other people interact with him, which in turns affects his experience of who he is. The fact that other people feel alienated by him affects his self perception and therefore alters his identity (ibid., p. 38). Note that according to this, self-alienation would be not merely the result of an altered bodily condition, or even of a lived bodily existence, but of an altered dynamical relation to other people. We get back to this in Chapters 6 and 8.

---

[34] It is worth mentioning that this implies a perspective on the social constitution of emotions that seems even more radical than the most recent enactive approaches to emotion. Colombetti and Torrance have proposed that emotions are part of the (participatory) sense-making activities of the individual *organism* (Colombetti and Torrance, 2009). Cole's considerations would however suggest that the constitutive basis of emotions is even more widespread as it transcends the organismic boundaries of the individual person. If emotions are acknowledged as part of mind then this interpretation would be much in accordance with the leading hypothesis of this thesis that the human mind is co-constructed in social interactions.

According to Dudzinski, Bauby's new identity also lacks what Zaner describes as the ability to *effortfully possibilize*. Before his impairment Bauby was able to choose from a variety of intentions and wishes and then perform the physical actions required to realize them. But due to LIS Bauby can no longer "actualize his choices" and "free himself from the actual" (ibid., p. 40). He "no longer chooses his friends, with whom he spends time" and can now only *imagine* to doing the things he was able to do in his first life.[35] Bauby can still choose which books to read and he enjoys the letters he receives from close people; to some extent these letters give him a sense of recognition of his self as "friend, father, son", saving his "integrity" (ibid., p. 40). Yet despite of this, Bauby's self has fundamentally changed with LIS and a prevalent retreat into the world of imagination and memories has become a fundamental part of his new identity.

According to Marcel "having a body" does not amount to having it "in the sense of ownership". I do not have my body like I have an apple. The body just *is* what the person is, not an object to be owned or an instrument of her mind. The body is not *mine*, but *me*[36]. Dudzinski argues that in case of Bauby, this kind of intimate link between body and himself is split. He does not have his body in the sense of Marcel, because it is no longer useful for him and also not experienced as his own anymore (ibid., p. 42). By remembering what he was able to do, he is only evoking an experience of his old body as an instrument, a tool, but not as himself. Through the loss of bodily capacities, Bauby has thus arguably lost a part of his self.

Again we find a link of bodily impairment and the social dimension: Bauby cannot "present himself (be present to) others". His "physical availability" is deeply affected as he cannot make himself available "to any *one*" in particular (ibid., p. 44). At the same time Bauby is "available to *everyone*" – he cannot decide who is coming to see him or not, he is unable to "evade" other people (ibid., p. 43) if he wanted to. Furthermore, since he cannot respond to somebody else he cannot become available to himself as a "creative being", i.e. someone who might impact the process of becoming who he is.

Like patients with Moebius, Bauby completely lacks the facial expressions and bodily gestures that usually accompany speech, but in addition to this his reactions are much more time-consuming and effortful which seriously affects the dynamics of communication. Bauby experienced a breakdown in his ability to communicate. But even if "illness and disability cause a shift in the character of one's intentionality […] they are not conditions in which there is a *lack* of intentionality" (Diedrich, 2007, p. 85, original emphasis). Bauby is still "yearning for intersubjectivity" (Dudzinski, 2001, p. 43) and continues to be directed towards others, even though he now creates himself through his memories and imagination and the small corridor of communication via eye-blinking movements.

These considerations again seem to refer a social and not only embodied dimension of the self. They foresee what I suggest in Chapter 7, that human cognitive systems have an intrinsic strive to participate in a socially shared world and that their identity is co-constituted by individual and inter-individual actions.

3.2.3 Ethical Issues

Gillett and Chisholm (2007) consider the phenomenology of the LIS condition prevalent in the personal reports of former rugby player Nick. From his reports the authors derive the following ethical predicaments for the care and treatment of LIS patients.

---

[35] This could similarly apply to people with depression, dementia and less severe bodily restrictions.

[36] Note however that according to German philosophical anthropologist Plessner even in non-pathological cases, the human body is a centre of activity that also involves an "eccentric positionality". Other than animals human beings do not only have a body or rely on it as a means to engage with the world, but also take a distant or outside (eccentric) position to it. They become self-aware of their own bodily experiences and the fact that they encounter the world through the body (Plessner, 1981). This suggests that part of normal bodily experience sometimes involves an instrumental perspective on the body.

Despite appearances to the contrary, LIS patients are still conscious and aware of "everything that is happening to them" (Gillett and Chisholm, 2007, p. 1). The LIS patient is a "being-in-the-world-with-others", not as an objective body, but rather as an existence of "being there" in a Heideggerian sense. Being there as somebody is "an aspect of one's subjectivity as a human being" and it requires recognition from and engagement with others (ibid., p. 3). Gillett and Chisholm echo Dudzinski when they claim that we *are* our lived bodies and that for this reason LIS is not simply an "add-on" to the person the patient has been before. LIS is a "matter of altered being" affecting and hindering our being there as somebody. The LIS patient suffers from a "double injury". She is heavily restricted with respect to her physical capacities, but there is also the social injury of "paternalism, infantilization, and the loss of dignity and respect that these attitudes imply" (ibid., p. 5).

Treatment of LIS patients requires therefore special attention to the fact that the patient still is a lived subjectivity. Judgments from the outside on behalf of the patient are prone to be "conducted in abstraction and condemned to irrelevance", the patient must have a say in evaluating her situation (ibid, p. 3). More important than clinical assessment and external speculations about the outcome of the condition is, according to Gillett and Chisholm to recognize the "patient as the creator of and living being at the centre of the story of his or her own life" (ibid., p. 5) and a social participant. The patient has to be taken seriously as a person and one should ensure "his/her *being amongst us*" by enabling an ongoing dialogue.

What to some extent has implicitly guided the considerations in the previous two sections now seems explicitly turned into a fact with normative consequences. Human beings are genuinely social and they strive for recognition. And because they continue to be social – despite bodily impairment – special effort on side of care-givers and others is required to ensure that this striving for recognition is realized and responded to.

While I am clearly not opposed to these considerations, I also think they imply a particular epistemological background and combine particular interpretations of the concepts at use. In my view severe physiological impairment such as LIS precisely invite to make these explicit. In the following summary I therefore try spelling out the presumptions and conceptual interrelations most prevalent in the above approaches to bodily impairment.

3.2.4 Intermediate Summary

The most important questions all of the papers deal with concern the impact of LIS on the body, the self, the role of social interaction, and how they are related. The focus is thereby either set on the individual and the impact of LIS from an external stance on the patient's capacities, or phenomenologically, on her experiences. Body and self are assessed in relation to the social environment. Note again that I do not want to argue whether the above phenomenological considerations provide support for the presumptions and conclusions drawn from them. For the moment, I only want to make explicit *that* and *what* presumptions are implicated by them.

Dudzinski's analysis is concerned with the relation of body and self. She claims that to the extent that LIS affects the patient's body so does it affect and change her identity. This change is experienced from a first-person perspective as a sense of alienation. Bodily alienation is experienced as part of her new identity. The patient no longer just *is* her body. That patients with LIS have restricted abilities to engage socially with others implies also not only a change but also a reduction of the self. I suggest that these considerations entail two interrelated presumptions about the relation between body and self:

(1) Body and self are intimately linked, if not even identical. A self is embodied.
(2) Following this, an impairment of the body necessarily leads to an impairment (reduction) of the self (alienation).

While these premises focus mainly on the individual, Cole and Gillett consider more explicitly the role of the social in LIS. The relation of physiological impairment and the social environment is addressed in two ways.

First, with a focus on the individual we see that physiological impairment affects the person's capacities to engage with others. Gestures, facial expressions and whole body postures allow to be recognized as somebody. For patients with LIS this is almost impossible. They cannot choose to use their bodies to go away or to show with their face what they feel in the encounter. They are "exposed to their glance being "available to everyone" (Dudzinski, 2001, p. 43). With regards to the present thesis I want to make explicit the following theses that seem to underlie the authors' considerations:

(3) We are social beings, i.e. we are embedded in a social world.
(4) Humans strive to show themselves to others as subjects (teleological dimension).
(5) Bodily capacities have a social purpose.
(6) Bodily impairment thus affects social capacities

Second, the restrictions in the ability to present herself as somebody is not only alienating to the patient. As we see in LIS patients but also in people with Moebius syndrome this alienation also affects how others perceive her and the way they interact with the patients. Depending on the level of recognition or kind of interaction this has an impact on the patients themselves (feeling isolated, frustrated, low self-esteem). It appears that there is an intimate relation between bodily impairment as affecting an individual's capacities, the reaction of others and the individual's self-experience (Zaner, 2003). Note the following implicit assumptions entailed in them:

(7) Bodily impairment affects the perception of others (alienation of others).
(8) This leads to a particular behavior/reaction toward the patient.
(11) This will in turn affect the patient's experience of herself (alienation).

The fact that we are social beings and as such depend on the recognition of others brings about two important ethical predicaments. First, despite or maybe because of their inability to show to others by means of speech, bodily expression or movement that they "are there", patients need to be recognized as subjects with a perspective of their own and the will to take an active part in a social encounter. This requires, secondly, that we find means to enable the patient to express herself and to ensure that she can continue to actively engage with other people.

(12) Others need to recognize the patient as somebody.
(13) The ongoing dialogue with the patient needs to be ensured. Means need to be found in order for the patient to engage with others.[37]

All of these presuppositions concern the relation of self, body and the role of the social environment, but they entail different perspectives on how these are interrelated. According to one line of reasoning, the body constitutes the self. The individual seems to just be this *embodied self* that then engages in and is further affected by social interactions. A disturbance of the body is a disturbance of the self. A sense of alienation is explicated as constituted by an altered embodiment. The second line of reasoning seems to presuppose a more complex perspective. The self is *indirectly* affected by impaired embodiment, in that the disability impacts the individual's capacities to engage with the social environment. This leads to an altered dynamics of

[37] This entails a teleological dimension that I have already listed in 4). The authors presume that being embedded in a social environment brings about the intrinsic need to be recognized by others as being somebody. This chimes well with the enactive approach to normativity as being an intrinsic dimension of autonomy. However, as I argue later, at the level of human beings this normativity is not based within the individual system but co-constituted in interaction with others.

the social interaction process. Others are alienated by the patient and may not recognize her as a subject, leading to experience isolation, frustration etc. The issue of alienation is reoccurring throughout the previous approaches.[38]

The interesting problem with regards to this thesis is how are these two perspectives related? Are they independent or two ways of expressing the same circumstances? Both entail that bodily impairment affects the patient's self. The difference could be that in the second case the impact of the impairment on the self is understood against the background of the social. The body is endowed with a social purpose. The focus here is not on the embodied, but the social dimension of the self. This could also entail a completely different observational stance: rather than focusing on the individual and suggesting that the individual is a social body, this perspective emphasizes the role of the social environment for its co-constitution *as* an individual. The experienced alienation of the body is not the direct result of the disability, but rather an emergent result of the (altered) social interaction process.

The point of providing the above list of implicit premises and conclusions is to show that they entail much more fundamental questions. We are dealing with persons that are unable to move, unable to speak and whose connection to their social surrounding is challenged. How can one have or be a self when one is unable to move, unable to talk to others, unable to show emotions? How is movement relevant for a body and to what extent is the interaction with other people crucial for our self? How can one be a body that is entirely immovable, not temporarily, but forever?

A mere phenomenological analysis of the above kind cannot answer these questions because it describes but does not assess *why* and *how* exactly self, body and the social should be interrelated. Such analysis requires a prior clarification of

1) the relation of body and self
2) the relation of self and the social environment
3) the interrelation of all three of them

That LIS affects the body and the self goes without saying, yet that it invites us to explicate or to sharpen our look at these concepts is not explicitly addressed. But to postulate an impact of LIS on the body and the self of the patient in a specific way that somehow involves the environment, *presupposes* a particular understanding of what self and body are. What this understanding is needs to be made explicit. To contribute to this is the major goal of this thesis. In the remaining chapters I suggest that body and self should not be equated and that whether we can equate bodily impairment with an impairment of the self is at the very least an open question. I will suggest that the way to understand how bodily impairment affects a patient's life is to take the second perspective derived from the above assumption. It is to acknowledge that we are socially embedded and that this impacts who we are as cognitive beings.

[38] Alienation is not only observed in the case of bodily paralysis but also occurs in many other cases of illness (Toombs, 1993, Svenaeus, 2000), including mental illness (Fuchs, 2005).

3.3 Locked-in Syndrome from a Philosophical Perspective, II

> A distinction becomes a dualism when its components are distinguished in terms that makes their characteristic relations to one another ultimately unintelligible. (Brandom, 1994)

In the following and third part of this chapter, I consider a more subtle approach to LIS offered by Kurthen et al. (1991). These authors have focused on an epistemological challenge that a complete cessation of bodily movement (TLIS) poses for theories of the mind and their assessment of consciousness. In the following I provide an overview of Kurthen et al.'s line of reasoning. I argue subsequently that LIS is not only challenging with regards to consciousness, body, self or any other *particular* aspect or concept of mind. It is a challenge that urges to make explicit the *background epistemology* based on which these different aspects and concepts can be sensibly related.

3.3.1 Accounting for Consciousness in Totally Locked-In Syndrome (TLIS)

Kurthen and colleagues (1991) argue that the medical assessment of consciousness is actually a version of the so-called other minds problem (OMP), i.e. the philosophical question of how we can have knowledge about other peoples' mental states (Malcolm, 1962, Wittgenstein, 1953, McGinn, 1984). With regards to clinical assessment of consciousness they observe two things.

First, this assessment is often, and wrongly, assumed to imply a behaviorist or neurophysiological epistemology. Cases of TLIS with a complete cessation of bodily movement (including eye-movements) are a problem for this kind of assessment. An account of a TLIS patient's consciousness based on a behaviorist assessment would be impossible because they entirely lack overt behavioral expression, and neither can one rely on neurological data because they are no reliable indicator of consciousness. However, according to Kurthen et al. the problem is not whether or not a behaviorist or neurophysiological epistemology is appropriate because this obscures the actual problem in TLIS, which is to acquire an intrinsic assessment of consciousness and provide an account for the subjective experience of the patient with TLIS. When it comes to clinical diagnosis, we can actually not decide between behavioristic or neurophysiological assessment because both are extrinsic as they rely on 3rd person observations alone (Kurthen et al., 1991, p. 72). Even if we were to have a "perfectly reliable" neurophysiological correlate of consciousness, we would still be reluctant to solely rely on this because we could not be sure that the patient really is conscious. Both, EEG signal and verbal reports only indicate that the patient is conscious, but they do not help to solve the problem of knowing about it. As long as we are not the patient herself, we would not know whether her assessment is veridical or not. The OMP would still remain (ibid., p. 73).

Second, to approach the question of consciousness in TLIS in this way is similarly misleading. This is arguably because the strategy underlying clinical practice appears in fact to be rather something close to a Wittgensteinian argument against the privacy of language (Wittgenstein, 1953). To know that somebody is present and has mental states is "part of complex *social practice* and not an epistemic 'relation' to some 'entities' (mental states)" (Kurthen et al., 1991, p. 74). In this view, there would actually be no OMP because

> in our ordinary social communication we already *have* an unproblematic, though proto-theoretical and non-scientific, concept of 'knowledge of other minds', which is at least *practicable*…we apply

> the same *outward* (that is: behavioral) criteria as in our everyday communication… (ibid., p. 75, my underlining)

In order to diagnose whether somebody is conscious or not clinicians can actually rely on a FP adapted to clinical practice. But this accounts for the "normal cases" of assessment of consciousness. In case of TLIS this strategy must fail because somebody who does not move at all and is unable to communicate or respond in an overt way[39] is just not part of our commonsense. According to the authors, the usual folk psychological

> procedure becomes problematic as soon as essentially *new* situations arise – situations we have not yet dealt with in the context of our established mode of life. And one of these situations is the confrontation with a TLIS patient, whose state of consciousness cannot be judged in terms of outward criteria and hence cannot be evaluated within the folk-psychological framework (ibid., pp. 75–76, my underlining).

TLIS thus transcends our everyday conception of human existence. It constitutes an "epistemological *borderline case*" (ibid., p. 77). And if we were to account for consciousness in TLIS we would need a new epistemological framework offering a "*complete description of the TLIS patient (as a human being)"*. In Kurthen et al.'s opinion this new framework would have to be *neuroscientific*. Nevertheless, the authors pessimistically conclude that to achieve such a complete neuroscientific description seems "utopian" (ibid., p. 76, original emphasis). TLIS can thus neither be assessed on folk-psychological nor neuroscientific grounds. As a consequence we cannot decide whether or not TLIS patients are conscious and have to pragmatically treat them "*as if* they are conscious" (ibid., p. 77).

3.3.2 Discussion

My line of objection to Kurthen et al. unfolds in five steps. I begin by outlining what I assume to be the epistemology underlying Kurthen et al.'s consideration, i.e. a *disembodied* and internalist perspective on consciousness. Making this explicit allows me to show in the next step that on rejecting such a perspective new outward criteria become available based on which current FP could provide an account of consciousness in TLIS. In the third part, I argue that *even if* we would adopt an internalist stance on consciousness, this would not necessarily lead to the conclusion that neuroscience could not account for consciousness in LIS. In the last two sections I address a contradiction that this objection apparently has with regards to my overall criticism of the internalist perspective adopted by the authors. It appears, that in addition to an internalist and neuroscientific conception of consciousness the authors implicitly re-evoke the hard problem of consciousness – something they explicitly deny. This turns their entire line of argument into a vicious circle. I suggest that there is a deeper reason for this. It is linked to their background epistemology and can be understood as a side effect of the radical shift from a behavioristic to a cognitivist perspective on cognition.

Implicit Epistemology

The previous authors assumed LIS to play such and such a role for the patient. Kurthen et al. similarly *assume* that consciousness constitutes a challenge. But an evaluation of an epistemological framework with regards to such a challenge does not take place on "neutral

[39] The relevance of the challenge proposed by Kurthen et al. can be directly seen when considering Laureys et al.'s description of the epistemological nature of the diagnostics of LIS (2005, p.723, see Section 3.1). They adopt a behavioristic and neurological approach and do not seem to have (at least not explicitly) a folk psychological stance on the condition or a special awareness of the fact that their folk psychology does not entail cases of TLIS.

grounds". It involves a particular epistemological stance itself. What the authors did not made explicit is *what* consciousness is and *how* it is constituted.

I suggest that the authors' view can be interpreted as a form of brain-bound internalism according to which consciousness is located in the head and constituted by neural activity: if no current epistemology can account for consciousness and TLIS and the future framework which would be able to do so needs to be neuroscientific, then it seems that ultimately consciousness can only be assessed on neuroscientific grounds.

Such an approach to consciousness would have two important consequences: We will most probably not consider non-neurological aspects of the human body, such as facial expression, movement or action etc. to be relevant for consciousness. Moreover, we will not assume that consciousness could be constituted by *anything outside* of the brain.

Folk psychology has an Account of TLIS

Recall Kurthen et al.'s claim, that clinical assessment of consciousness generally relies on a "refined" version of folk psychology (FP) and that it is only in TLIS that it fails to do so.

In my view this conclusion is misleading. If we accept that FP generally entails an assessment of consciousness in other people, then we should also accept that FP is able to account for the special cases of TLIS. The problem with Kurthen et al.'s analysis is that they do not sufficiently explicate why they take TLIS to be special. I agree with the authors that TLIS is a special case, but its being special is no reason why FP could not account for it. What they propose is that TLIS lacks the "outward criteria" that we would generally apply in order to assess consciousness. But I assume there are actually further reasons that motivated Kurthen et al.'s conclusion.

Firstly, no scientist, and no philosopher can pursue her research without a particular theoretical background. Our descriptions and observation are never neutral, but informed by our own convictions, experiences and theoretical paradigms (Gadamer, 1960/2007, Kuhn, 1962/1970). If it was true that an internalist and reductive epistemology guided Kurthen et al.'s considerations then this may have also affected their perspective on the layman's folk-psychological assessment of consciousness. If I assume that consciousness arises internally and that it is only in the head, I will take (as they do) as outward criteria for its presence verbal reports, a particular behavior of the subject and/or some brain images that I take to correlate with these verbal reports. Now, if these criteria were the *only (*outward) criteria guiding my folk-psychological assessment of a TLIS patient, I might be indeed be puzzled and find the situation new and bearing reference to anything I have previously believed.

Note however that layman and medical doctor have very different knowledge. The medical doctor presupposes that TLIS is a state of complete paralysis affecting movement and the ability to communicate and that this state is the necessary result of a neurodegenerative disease.[40] Since the patient was responsive until she became completely locked-in the clinician must have implicitly assumed that she was conscious. Now that the patient is immovable, this judgment is under scrutiny. I believe things would be different in case of a layman.[41] Imagine a

[40] I assume this to be the setting that Kurthen et al. have in mind when asking whether TLIS is conscious or not. If they did not mean an assessment of consciousness *after* the patient became completely locked-in and on a patient that crucially *is diagnosed* as having LIS, then their entire considerations would be based on a paradox. The other possible scenario would be that of a previously *undiagnosed* patient who presents with symptoms of complete paralysis and for whom the doctors would want to evaluate whether or not she is conscious. This is exactly one of the greatest diagnostical problems that neuromedical staff is concerned with: to discern different medical states, such as coma, vegetative state and LIS based on similar bodily conditions. If Kurthen et al. ask whether FP can account for TLIS they cannot mean to embed the question in this alternative scenario for a patient diagnosed with TLIS is – *per definitionem* – already treated as conscious.

[41] One might object with Wittgenstein that the following considerations stand in contrast to FP. If we accept that folk psychological assessment is a social practice which can be described, but not explained, then the attempt to shed light on how it is that people assess consciousness in LIS is misleading. An evaluation based on

person who has never seen somebody who is not sleeping (and not dead!), yet also not moving and not responding. I believe such a person would most likely assume that something is wrong with the person. If she were asked to tell whether this patient is conscious, she would reply with "no". The explanation for this – i.e. *my* explanation, as an informed observer who confronts the situation while in the same time paying attention to how she is forming her judgment – is that contrary to the authors' conviction there actually are *other* indicators that can guide the layman's evaluation as e.g.:

a) absence of movement
b) no communication
c) no reaction

If we see somebody who shows none of the above criteria we will most probably take this to indicate that the person is not conscious. Either this conclusion involves an act of "testing" a set of already known criteria against a situation and realizing that none of them applies. Or it relies on perceived criteria that are directly evaluated as indicators for a particular state of mind, namely *un*consciousness. Neither would be exceptional, but a rather usual way of evaluation in "our ordinary discourse".[42] For the ordinary person the "fundamental uncertainty" (Kurthen et al., p. 77) the authors ascertain simply does not existent.

This is not to say that the commonsense conception of consciousness in TLIS is correct or sufficient for clinical purposes. However, a folk-psychological assessment based on movement may still provide us with an intuitive source of what could be at play in consciousness. Namely that it is shaped not only by neuronal, but also bodily activity. One does not have to be so radical to assume that as soon as the patient stops moving her body her consciousness would be affected. The question to be addressed is rather whether bodily impairment could impact on consciousness in the long(er) run. I discuss this possibility throughout the assessment of body's role in cognition and consciousness in the remainder of this thesis. And to be already explicit on my own background epistemology, it comes very close to the perspective I believe we should adopt.

The opposite scenario – a neuroscientific approach

In this section I am asking how a person who has background information about LIS and also an implicit neurocentric epistemology would evaluate the state of consciousness in TLIS. In contrast to Kurthen et al. I believe that such a person will have some reason for assuming that the patient is conscious.

To know that a patient has TLIS implies knowing that the patient has been conscious *before* she became totally locked-in and that TLIS is a condition of severe paralysis in its most dramatic dimension, viz. the final state of ALS.[43] The informed person could use this as an

---

commonsense is exactly not to question the basis of one's judgment, but to just perform it. However, we should remember that there is a difference between mere descriptions of what people say and the attempt to understand why they say what they say. Trying to derive knowledge about how somebody comes to assess that a person is conscious, is not to deny that the layman usually does not reflect on her judgment. These are simply two different perspectives at play. One of the layman, and the other one of a researcher who is "putting herself into the layman's shoes" and observing with more awareness how she would possibly deal with the encounter. And when Kurthen et al. claim that TLIS is not part of commonsense because there are no outward criteria they must have done just this: trying to understand commonsense.

[42] Some people may worry that this wording implies a representationalist stance on social cognition. Let me clarify that this is not the case. I do not think that the assessment of consciousness consists in an actual comparison of sentence-like bits of knowledge, I would explicate this rather in terms of a story involving an immediate and shared enactment of embodied knowledge.

[43] A related scenario could actually be construed for the layman as well. Suppose that a person who has no medical knowledge about TLIS has a close friend with ALS. The person would engage and communicate with the patient, and even find a way to have conversation with him throughout the degenerative course of the

explanation for why the TLIS patient is not talking and not moving. The patient is unable to move or speak because the paralysis prevents her from doing so. This does however not mean that the patient is unconscious. She is conscious, but unable to show it. This judgment would be supported by an implicit neurocentric background epistemology according to which consciousness arises in the brain, *not in the body*. Due to that a dysfunction of bodily movement alone would not considered a justification for assuming that somebody is *not* conscious. The absence of movement and communication are not taken as a sign of unconsciousness (say as in coma or vegetative state).

One could argue now that in this line of reasoning I only considered the *symptom* of LIS (the inability to move), but not its *cause*. We have learned earlier that LIS either results from a lesion of the brainstem or in case of TLIS, from ALS. Since brainstem and motor neuron are parts of the brain, the informed person might as well consider bodily paralysis to be the symptom of LIS, but take its cause to be lying elsewhere, namely in the brain. As a consequence the informed person could assume that consciousness in TLIS is affected not because of a lack of bodily movement but because of a neurological dysfunction. She should however *not* say that there is no account of TLIS as Kurthen et al. claim.

Curiously, now I have shown that an internalist epistemology implies the exact opposite: a negative evaluation. This seems to contradict my above assumption that an internalist epistemology leads to a positive evaluation of consciousness in TLIS. But the apparent contradiction can be resolved if one considers not only that ALS (the cause of TLIS) affects the brain, but also factors in what particular part of the brain it affects, namely areas presumably responsible for or involved in *movement*. Part of the internalist epistemology is to believe that movement plays no role in the emergence of consciousness. As a consequence a degenerated *motor* neuron should be no threat to my conviction that consciousness arises in the brain because motor areas are not where I would look for neuronal correlates of consciousness. A TLIS patient may be paralyzed, but since ALS affects only motor areas, the patient is undoubtedly conscious. Note that this line of reasoning appears to reflect the commonly accepted description of LIS as a condition in which the person is trapped in an immovable body, fully conscious and mentally intact (see Section 3.1.1).

Let me summarize the previous discussion. I made two claims that stand in contrast to Kurthen et al. Both are based on my argument that the authors misconceive the folk psychological assessment of TLIS due to an internalist and neurocentric conception of consciousness. Without such epistemology, so I suggested, the layman would most likely have a *negative* evaluation of consciousness in TLIS (involving the outward criteria of movement and the lack of communication). An internalist and neurological epistemology of consciousness, in contrast, would most likely lead to a positive evaluation of consciousness in TLIS. The difference between the outcomes of these evaluations is explained by reference to the particular stance on the role of *bodily movement* (and the lack of communication). While these seem to matter in an ordinary conception of consciousness, they play no role for an internalist and neurologically based approach to TLIS.

Note however, that my conclusion that an internalist and neurological conception of consciousness *does* in fact lead to an account of consciousness cannot explain why Kurthen et al. argue that there is *no* account of consciousness in TLIS, which I promised to do in the first place. I show now that their argumentation entails an evocation of the OMP and that this taken together with an internalist perspective on cognition may explain the pessimistic conclusion Kurthen et al. derived at.

---

paralysis, i.e. also in the classical state. When the situation worsens so that the patient is not able to communicate even with her eyes, the friend would most probably not assume that she is unconscious once she is TLIS because she has experienced her as conscious *before* and even within the less severe state of classical LIS. Thanks to Mike Beaton for reminding me of that possibility.

Implicit Evocation of the OMP

Remember that the authors do not argue that a behavioristic or neuroscientific approach would be wrong or right per se, but rather that these approaches do not suffice for diagnostic purposes because they do not tell us *whether or not* a TLIS patient is conscious. Clinical assessment of consciousness is rather based on a Wittgensteinian social practice in which, we apply – based on our holistic image of what it means to be a person – a particular attitude towards each other. Folkpsychology is applied because we do not want to have to deal with the OMP, i.e. "get […] entangled with the philosophical paradoxes that arise from the 'epistemic authority' of first-person experience and from the analogical argument", i.e. with the question of how we can "*ever* be sure about the occurrence of consciousness in other people?" (Kurthen et al., 1991, p. 77).

We have learned that according to Kurthen et al., the general knowledge of FP does not entail a condition like TLIS. Suppose we accept this claim. Why would we also have to accept that FP could not develop so that at some point the condition of TLIS would be just part of commonsense? We may not be familiar with a person who is not moving and not responding, yet presumably conscious; but could we not become familiar with it? If present commonsense has a way to evaluate consciousness in ordinary cases then at least a different, *future* commonsense, which also entails knowledge about the human condition of TLIS should in principle very well account for consciousness in TLIS. Why would Kurthen et al. ignore this possibility? One reason could be that it would be exactly *not* commonsense as it implies the need to "create new language games" (ibid., p. 76). However, this cannot be meant in a general, but only temporary sense, for denying such development and thus the creation of new language games would imply that commonsense knowledge is static, which is rather unlikely.

I believe the actual reason for neglecting the possibility that a future FP can account for consciousness in TLIS lies in their claim that in order to account for consciousness in TLIS we need to understand "what it means to be a human being in *all* respects" (ibid., p. 77). But if this was just for the reason that we lack knowledge about this special condition then why is it that FP already had an account for consciousness in normal cases? Why do we *suddenly* need a complete image whereas apparently the previous one was completely sufficient for the ordinary cases?

I believe this is because the authors expect actually much more from a complete image of man than just the acquisition of additional knowledge about patients with TLIS. And it seems that this extra bit of knowledge would desirably allow us to tell whether a person is conscious or not (even if we are not the person herself). But this would mean exactly that the complete knowledge is expected to also entail a solution to the OMP.

If it was not for sheer intuitive reasons that this urge to account for the "hard problem of consciousness" still resides, then it will be a consequence of the author's own argumentation as I show now. Consider again their claims:

> 1) Any account of consciousness should avoid the OMP, and apply commonsense, i.e. social practice (FP).
> 2) Current commonsense social practice (FP) is not applicable to TLIS.

Note that the second claim does not imply that the first would not be valid anymore. Just because current commonsense is not applicable to TLIS does not mean that consciousness generally, i.e. in ordinary case, should *never* be assessed based on commonsense. The problem is that Kurthen et al. seem to simultaneously deny and accept this point. They explicitly accept it in that they apply (1) to the case of TLIS. They then see that there occurs a problem, yet they still keep assuming that (1) is necessary as they now argue that we need a complete image of man. This "complete image" would obviously refer to the "incomplete" image of current FP. It is needed for ascribing consciousness based on *more* knowledge about what it means to be human. However, if a complete image of man is (in principle) to be provided by neuroscience, and *not* by FP, then one should not also assume (1) that FP should guide our assessment. In other words,

since the authors argue that FP is unable to provide a complete image of man, yet that a complete image of man is required to assess consciousness, the conclusion should be that we cannot apply (1) in order to assess consciousness. But this would mean that we have to give up on (1) which re-invites the classical OMP that the authors wanted to avoid. Paradoxically however, by arguing that (neuro)science, and *not* FP could provide the necessary complete image of man the authors actually *are* giving up on FP. They leave the "safe grounds" which ensured that the OMP is not threatening. When the authors assume that only a future rich neuroscientific image of man can account for consciousness in TLIS they cannot logically rely on (1) anymore and yet they *explicitly* do because they reject that an assessment of consciousness involves the OMP. The argumentation seems to be lost in a *vicious circle*.

Now seems clear why the authors conclude that only neuroscience could provide the complete image but that it this is an *utopian* endeavor. This conclusion is actually an attempt to escape the circle. Since the authors want to *avoid* the OMP, the only possible resolution to their problem must be to *ultimately* deny that neuroscience is able to account for consciousness in TLIS. If a neuroscientific approach does not entail a resolution to the OMP *and* I consider the assessment of consciousness to exclude the OMP, I should not assume that neuroscience is able to give an answer. Unfortunately, even this move would have solved nothing. Due to the implicit argumentative circularity it would bring us back at (1), which the authors assume to be unable to account for TLIS. We see that despite arguments to the contrary, it could be the very presumption of OMP that has motivated the author's line of reasoning. If the implication of their argument is that a complete image of man must also entail an answer to the hard problem of consciousness, then, of course, neither folk psychology, nor neuroscience, nor any other approach for that matter could *ever* account for consciousness in TLIS.[44]

Why evoke the OMP?

Before concluding this chapter let me explicate why I think that the authors necessarily have to evoke the OMP. As already suggested it seems that the authors' implicit epistemology is well in accordance with a cognitivist view of cognition. If we consider the context in which the authors' internalist epistemology emerges, i.e. the shift from a behavioristic to a cognitivist epistemology the apparent problem to theorize about another person's conscious state becomes quite evident.[45]

Recall that behaviorism is interested in outward criteria, whether they concern the input, i.e. the stimuli that the person receives or the, output, i.e. the behavior that the person performs. It completely neglects what is going on *inside* of the person. Thereby it fails not only to account for the subjective dimension, but at the same time also fails to provide an explanation of the functioning of the *mechanism* underlying stimuli input and behavioral output.

With the beginning of the cognitivist view a complete shift into the opposite direction took place. This shift implied two things, a turn *towards the inside* and the objective to also *explicate* the inner *mechanism* driving cognition. And here lies the problem: To reject behaviorism *and* to wholeheartedly embrace cognitivism amounts to an *epistemological black-and-white* maneuver where one extreme position is taking over for the previous. Crucially, in some sense cognitivism and behaviorism are equally concerned by "the inside". It is ignored by behaviorism while being

[44] If it were true that a reference to the OMP has rendered their argumentation conclusive, there would still be a fundamental problem. To say there is no account of consciousness in TLIS should be based on the assumption that an account of consciousness in TLIS requires a solution of the OMP. But this would then similarly apply to all other conditions and thus render the claim that TLIS is a special case obsolete.

[45] Interestingly even researchers who are willing to accept that body and action not only shape cognition but are involved in its very constitution, remain reluctant to consider the condition of LIS to have an affect on consciousness. There is a split between consciousness and cognition, which explains why it currently seems to be quite à la mode to postulate a kind of vehicle externalism while holding *in the same time* that the external vehicles have nothing to do with consciousness. It is a phenomenon that rests within the head (Clark, 2008b, see also chapter 4).

focused on by cognitivism. Cognitivism aims to explain the mechanism of the cognitive *and* considers this mechanism to be internal (where internal refers to the brain). However, to account for what is going on *within* a cognitive system and to explain the mechanism of the cognitive are two different things, they do not necessarily entail each other (see Section 2.3.2). That the mechanism of the cognitive may as well be relying on processes located outside of the brain is ruled out by cognitivism. Seemingly the behaviorist's radicalness in ignoring the inside has "triggered" the exact opposite response from cognitivism: *if* we were to explain the mechanism then it must be located exactly where the behaviorist *did not* want to look, namely inside the head.

By trying to overcome behaviorism the cognitivist has thus thrown the baby out with the bathwater. Anything that does not concern the brain such as the rest of the body, movement or even the external environment plays no role. It will not assume that bodily action and movement could represent more than the output of something that is going on inside.

To conclude, if the option for an epistemology of consciousness in TLIS consists in choosing between either behaviorism *or* cognitivism, we are indeed unable to account for consciousness. The first has no explanation of the mechanism of cognition as it *only* looks at outward criteria and does not connect them to a story about *how* cognition works. The second is looking for a strictly *internal* account of the mechanism of cognition. It misses that while the inner, i.e. neurological processes may be of importance for cognition, they maybe not the only components relevant for it. Such epistemological black-and-white maneuver could then lead to the assumption that consciousness resides only within the head.

Upon believing that consciousness is in the brain and that everything external can only be treated as a result or expression of the inner cognitive mechanism, accounting for another person's consciousness must indeed amount to a Sisyphean task. We would try explicating something that is – due to our background commitment – *in principle* inaccessible from the outside. Bearing any other indicators of a person's conscious state, a solution to resolve this problem must then appear to adopt the Wittgensteinian strategy. We will then arrive at the authors' central claim that since clinical diagnosis of consciousness is based on a Wittgensteinian strategy, it has no account for consciousness in TLIS because TLIS is not part of our commonsense.

This line of reasoning implies that it would be possible to come to another conclusion if one adopted a different background epistemology. One would not have to presume that consciousness resides in the head alone by assuming a non-cognitivist attitude about the mechanism of cognition. For a non-cognitivist approach to cognitive science the hard problem of consciousness would have not necessarily arisen as we have see for example in the enactive approach (see 2.3.1).

Before coming to the final conclusion of this chapter, let me now conclude the second part of philosophical considerations of bodily impairment and LIS. I agree with Kurthen and colleagues that LIS poses a challenge. But I also think that what they take this challenge to consist of is far too limited. The authors have picked the most radical condition of LIS – TLIS – in combination with what seems to be one of the probably most difficult phenomena of human existence: consciousness. This combination runs the risk of obscuring much more general and multifaceted questions arising in the context of LIS – it therefore does not reflect the actual scope of the challenge. *What* is challenging in LIS is not only to assess *whether* or not the patient is conscious. This question woud require to first explicate *what* being conscious actually amounts to. Definitions of consciousness are always based on a particular background theory and to respond to the challenge demands to lay open our own epistemological background. Nevertheless, this applies also more generally to all aspects of cognition. Consciousness may be treated as *the* mystery of human existence, however, I am inclined to say: equally mysterious are also other concepts in cognitive science.

3.4 Conclusion

In this chapter I have introduced and discussed work that approaches LIS from two different angles. On the one hand from a *phenomenological* perspective, based on which the impact of bodily impairment on self, body and social interaction of the LIS patient was described. The phenomenological observations raise important issues to be considered in the context of bodily impairment. We learn that in human cognition, the body seems closely linked with questions of identity (self) and sociality. However, these concepts remain undetermined and the linkages between them only implicit. I have argued that in order to establish a perspective on the interrelation of body, self and sociality we are required to make explicit what we mean by them. I have therefore suggested that cases of bodily impairment and especially of LIS should be considered a challenge to the concepts by which we explicate human cognitive systems.

On the other hand I have discussed work that acknowledges the special epistemological status of LIS in the complete state and holds that it is a challenge to our assessment of consciousness. However I have argued that LIS poses not only a challenge to consciousness but to our background theory of cognition in general. The outcome of an assessment of consciousness will depend on what background theory of cognition we apply.

Cases of bodily impairment and in particular of LIS thus pose a *general* and two-fold challenge to cognitive science, i.e. to provide definitions of the concepts of body, self and social interaction and to make explicit what kind of epistemology is driving our perspective on cognition. Both are intimately tied. A definition of individual concepts remains limited without relating them to other concepts, yet in order to relate them to other concepts, a perspective is needed from which we can tell how they are interrelated.

I agree with Kurthen et al. that this perspective should be holistic. But I believe that it could neither be offered by neuroscience nor by a phenomenological approach. While the first is reductive, the second is merely descriptive. Phenomenology does not offer a mechanistic or organizational explanation of the phenomena it describes and neuroscience fails not because it is currently insufficiently developed, but because even a future neuroscience will always remain a limited explanation of cognitive phenomena at the *neurobiological level.* A framework is required in which phenomenology and neuroscience assume their proper role as complementary perspectives in an encompassing interdisciplinary endeavor, not as superior or inferior accounts of the cognitive. The interdisciplinary endeavor is cognitive science and the epistemological framework, which I believe is most promising for an integration of the various dimensions of cognition is enactivism.

The basic idea guiding the following chapters is therefore that adopting an enactive background theory helps not only providing definitions of body, self, or even consciousness, but also understanding in the context of LIS how they are related.

While I have discussed the previous two approaches to cognition and consciousness to show *that* and *how* LIS poses a general challenge to cognitive science I will now address and refine the challenge, by putting it to practice. In Chapters 4–6 I discuss LIS as a challenge for recent approaches to *embodied* cognitive science and their conception of the body. However, I also show that understanding the impact of LIS requires transcending considerations of embodiment and adopt a social perspective. In Chapter 7 I provide steps towards a refined notion of autonomy (as socially constituted) and illustrate how the enactive epistemology can offer a way to intelligibly connect these concepts. This allows understanding how the patients are affected by cases of severe paralysis, such as LIS. Throughout the chapters I foster my basic hypothesis that human cognitive systems are best described and explicated from a social perspective. Adopting this perspective will allow contrary to Kurthen et al. for a theoretical assessment of consciousness in TLIS. Such an assessment implies not only to consider the role of active bodily engagement but crucially, that of an ongoing social interaction. While the emphasis on the social embeddedness of LIS patients in particular, and human being in general chimes well with the phenomenological

observations provided by Dudzinski and others, the enactive alternative can additionally explicate them (see Section 7.5).

Chapter 4.

LIS and BCI – a Challenge to Embodiment, I:

The Sensorimotor and Extended Functionalist Approach

> We are just at the beginning phase of understanding the myriad ways in which the body is in the mind. (Rohrer, 2007)

4.1 General Introduction

In the previous chapter I have suggested that LIS poses a general challenge to cognitive science. It invites us to reappraise the concepts and epistemology underlying our perspective on cognition. In the present chapter I introduce a more specific variant of this challenge.

Recall the shift in cognitive science away from the orthodox view of cognition toward a dynamical and embodied perspective on cognition (Chapter 1). The new approaches clearly emphasize that body and environment play a crucial role for cognition, though they reject the classical view to a varying degree of radicalness. Importantly, they do not constitute a coherent framework. "Embodied Cognition" appears to be a label for a range of different, partly overlapping, partly even incompatible approaches. It is not clear what the body in these so-called embodied approaches actually is and to what extent it supposed to matter for cognition.

LIS-patients are most deeply affected with respect to their bodily condition and their ability to interact with the environment. In BCI-use there is even no bodily movement involved at all. This poses a multisided challenge to every theory of cognition that takes the body and the interaction with the environment to be important. How does a theory that assumes a crucial role of the body for cognition account for the fact that a patient is presumably cognitively intact and engaging with the environment despite severe bodily restrictions? My suggestion is that LIS and in particular the cases of LIS-patients that rely on BCI technology to enable movement or facilitate communication are of crucial relevance for these new embodied approaches for cognitive science as they invite us to make explicit what we mean by the body and to elaborate on the kinds of relation that presumably hold between cognitive system, body and environment.[46] It

[46] To be sure, LIS is not the only condition leading to this kind of questions. In so far as the role of the body is put into question by LIS, the challenge to embodied accounts of cognition would seem to be an extreme case of a series of similar challenges where a restriction to the body's capabilities for action is used to question the role of those capabilities for sustaining some form of cognitive performance. LIS is located at the far end of a spectrum – intermediate places are occupied by cases affecting bodily motility as for instance in tetra- and paraplegia, the capabilities for bodily expressivity (deafferentiation, Moebius syndrome, see chapter 3), distortions of the body image (Lotze and Moseley, 2007) and others. In all these cases it is possible to question whether the affected capability has any effect on the cognitive skills that it supposedly supports according to embodied views. At the nearest, non-pathological extreme of this spectrum we find simply the case where a voluntary decrease in overt bodily engagements with the world seem to have very little effect on some mental capabilities, such as counting, reasoning, imagining, remembering, planning, etc. Another example is the so-called deep sleep paralysis (Goode, 1962). There are meditation techniques requiring the student to sit still, a case of voluntary and temporary immobility (Goenka, 2000).

A motivation for working out in some detail how this kind of challenge is to be formulated in a useful manner for the case of LIS and for exploring the possibility of embodied responses is that such re-formulations and such responses might be applicable to large parts of this spectrum of problems. Cole (2009) has recently formulated a variant of the challenge with respect to patients with quadra- and paraplegia as well as Möbius

can act as a kind of empirical touchstone, which can be referred to in assessing their conceptual, empirical and phenomenological plausibility.

The challenge of LIS to embodied cognitive science is discussed within three chapters, the present and the subsequent Chapters 5 and 6. I begin my considerations by providing a first intuitive formulation of the problem, the so-called crude version of the challenge as well as a similarly intuitive and crude solution, the developmental response. I argue that both are missing or avoiding the actual challenge. I then provide a more refined version of the challenge by way of confronting four major approaches to embodied cognitive science with different cases of LIS and BCI. These approaches are, the sensorimotor approach and a version of functionalism – the extended cognition approach (in Chapter 4), the phenomenological (in Chapter 5) and the enactive approach to the body (in Chapter 6).

4.1.1 The Crude Challenge

Explicating the challenge does require caution, because at a first and hasty glance it presents itself in a radical but also shallow version. I call this the *crude challenge*: LIS patients cannot move their bodies. But if the body is constitutive for cognition and consciousness (as the embodied approach claims), then there could not be cognition or consciousness in LIS. This version of the challenge seems to be already present in discussions of embodied cognitive science, not explicated but only in the form of an "informal criticism" as Anderson (2003) observes:

> One informal criticism—informal because although the objection has come up in nearly every discussion of EC-related research I have been involved in, I have not seen it worked out in detail in print—is that EC cannot be true because the physically disabled are obviously able to learn, acquire concepts, and communicate. If concepts were, indeed, related to, or their contents depended upon the specific physical capacities of embodied agents, then there ought to be detectable differences in the conceptual repertoire of differently abled individuals; yet there are no such differences. (Anderson, 2003, p. 113, my italics)

This could bring critics of the embodied view of cognition to the scene. In line with a cognitivist and internalist view of cognition they could use this version of the challenge to deny the plausibility of the embodied perspective altogether: Because LIS patients obviously *are* conscious and cognitive function is preserved, the body or bodily action cannot be a necessary enabler of cognition. The abovementioned challenge would only illustrate that the crucial claim of proponents of embodied cognition – that experience and cognition require the body – must be wrong. LIS simply shows that no bodily activity is required for cognition. An explanation of how cognition is possible in LIS could not or should not be given from an embodied perspective.[47]

4.1.2 An Easy Way out – The Developmental Response[48]

Let me consider another way to respond to the crude version of the challenge as a defender of embodied cognitive science and to formulate a criticism by its opponents. It may be called the developmental response. This response emphasizes the role of embodied engagements with the

---

syndrome (chapter 3). His analysis may be seen a kind of foregoer of the present challenge, however, in terms of argumentative structure, the LIS case presents embodied views with the strongest version of this family of challenges.

[47] We have already encountered a version of this reply in the previous Chapter 3. Kurthen et al., (1991) acknowledged the challenging dimension of LIS for cognitive science but their assessment remained limited due to an underlying cognitivist epistemology.

[48] This section has been formulated in the main by Ezequiel Di Paolo as part of a joint paper on LIS as a challenge to embodied cognitive science (Kyselo and Di Paolo, in preparation).

world in bringing forth a certain cognitive skill. It seems likely that learning to move in the world, bumping into things, feeling effort, etc. are embodied experiences that influence our capabilities to reason about, say, causality, intentions, plans, goals, objectives, and so on.

Several bodily schemas underlie our culturally mediated use of linguistic structures (e.g. Lakoff and Johnson, 1999) and our mathematical skills (Nuñez, R., 2004). A history of bodily interpersonal engagements since infancy seems to underlie our social skills, our abilities to take the other's perspective and to think in objective terms. So a rough and ready reply to the above challenges could simply be: The body plays a crucial role in the *development* of cognitive skills. Even in cases when the body is impaired and cognitive skills unaffected, we cannot get rid of the body in understanding how such skills have *come about*, even if they so happen not to depend on bodily factors for their re-enactment.

There are at least two dangers with this strategy. While it seems undeniable that a compelling case could be made for the *enabling* role played by bodily factors in the development of some cognitive capabilities, it is far from clear that the case can be made convincingly *in general* (save for the trivial, but uninformative fact that all instances of genuine human level cognition are to be found in humans who have living bodies). Adopting this kind of response forces us to empirically determine case by case whether the body plays or does not play a role in a developmental history (as well as determining what counts as a crucial role or not – a far from trivial question).

Thus, in so far as there is (as many embodied theorists would like to claim) an essential link between embodiment and mind, this link may be obscured if the body is investigated only by examining its role as an enabling, i.e. causal factor in a historical account. This, of course, in no way is equivalent to suggesting that such investigations would be uninteresting or irrelevant for an embodied science of cognition.

Methodologically, attempts at a general statement concerning the role of the body in cognition would run the risk of becoming a collection of "just-so" stories – a criticism already raised against the adaptationist program in evolution (Gould and Lewontin, 1979) whereby the mere formulation of a plausible story runs the risk of replacing the actual empirically supported examination. This is especially likely since the role of any factor in complex, multi-layered developmental histories is quite hard to determine. The undesirable result of adopting such a strategy would be an embodied cognitive science that dogmatically assumes a developmental role for the body as a general rule without a theoretical justification.

The other danger is even more worrying. This strategy might in fact be conceding too much in accepting that the body plays at least, but conceivably sometimes also *only*, a historical role. If this were the case, embodiment-skeptics could easily question whether such a role, while historically necessary along a given developmental path, is merely contingent, in other words, whether other developmental (or design) routes might not arrive at the same results with bodily factors playing no role, or playing a different role (thus revealing the contingent connection between a given bodily factor and a given cognitive skill). If some of our mathematical abilities seem to be supported by an embodied history of bodily coping with objects and people but can, at the same time, (for the sake of argument) be described in purely formal, disembodied terms, why could these capabilities not be constructed in an artificial cognitive system using the resulting formalism, obviating the embodied history that led to them in the human case?

Even if it could somehow be proven that a bodily factor plays a role, in fact the same role, in all possible historical paths leading to the development of a cognitive capability (something that approaches a theoretical impossibility), the door would remain open for the embodiment-skeptic to claim that a physical necessity in a historical account does not amount to a logical necessity at the level of essence. The cognitive skill under question would just happen to necessitate a bodily factor as a historical enabler without this meaning that the body plays an essential role in such a skill, for instance, every time it is enacted.

It would seem that claiming a developmental role for the body as the answer to the challenge is in fact a way of avoiding the challenge in a similar way as we have already seen in the

crude response of the embodiment-critics. As I shall argue later, the very separation between developmental history and current cognitive skills may not be so clear-cut; a possibility that would also be obscured by the developmental response (see Section 4.2.2.). In contrast, to claim that the body is essential for cognition, we would have to show that fundamental aspects of all cognitive skills could not be understood without recourse to some fundamental aspects of embodiment, not only in terms of a historical progression, but also in terms of the disposition and enactments of such skills. Herein lies the demanding challenge posed by LIS for embodied accounts of cognition.

The crude version of the challenge and the developmental response are a reflection of the original confusion in the field of embodied cognition. They may come rather naturally when it is not clearly specified what exactly the role of the body for cognition is and what "the body" actually means. Yet for the exact reason that there is not "the" embodied approach, and "the" possible answer to what the body is, a crude version of the challenge would be too superficial and lacks an actual addressee. A fair analysis should acknowledge the variety of embodied approaches and not reject embodied cognition altogether. For that reason in this thesis I pose the question what embodiment is and to what extent is it required for cognition with respect to the different approaches subsumed under the label of embodied cognition. Their assessment will rely on the following cases formulated based on the different states (incomplete, classical and complete) and causes of LIS (stroke or ALS) on the one hand and patients relying on BCI technologies on the other:

1) Classical state (global paralysis, except for eye-movements)
2) Complete state (complete paralysis)
3) Complete state (complete paralysis) + communication with BCI
4) Classical state + ALS + BCI
5) Complete state + ALS + BCI

I discuss the extent to which the abovementioned approaches are able to address these cases and shed light on their particular conception of the body and the degree to which they assume it is necessary for cognition. The strategy for each approach is as follows. Firstly, I select and present particular claims about the nature and role of the body. Secondly, I confront them with cases of LIS and BCI and assess whether they can sufficiently explain them. Thirdly, I will hopefully arrive at a refinement of the challenge by working out related concepts and key issues that need to be addressed. At each instance of refinement, the challenge thus undergoes a transformation. And the embodied approach, which can be said to counter the challenge, must preferably not only account for some of its aspects, but for the challenge in its growing complexity. In other words, if a specific claim is not or only insufficiently accounting for a particular instance, then this is considered as an invitation for further elaboration on the concept of embodiment in general.

4.2. Refining the Challenge I: The Sensorimotor Approach to Cognition

As is the case with the overall field of embodied cognition, the sensorimotor approach is not coherent and makes different, partly contradicting assumptions. Contributions stem mainly from the sensorimotor contingency hypothesis, which is chiefly concerned with a specific part of cognition, namely vision and visual consciousness (O'Regan and Noë, 2001, Noë, 2004). Other proponents of the sensorimotor interaction approach are concerned with cognition in general (Varela et al. 1993, Thompson and Varela, 2001). Generally speaking the sensorimotor approach to embodied cognition puts emphasis on the constitutive role of specific bodily structures for cognition. In the discussion I mainly focus on the approach proposed by Noë and O'Regan.

In accordance with the strategy outlined in the introduction I am going to assess two major claims of their sensorimotor approach. Firstly, that cognition relies on *bodily movement* and secondly, that it requires sensorimotor *skills*. Confronting these claims with cases of LIS provides the first refinement of the crude challenge to embodied cognitive science. The general question "how can embodied cognitive science account for LIS?" is now reformulated: how does an embodied approach to cognition according to which embodiment amounts to bodily movement and the possession of sensorimotor skills, account for cases of LIS?

4.2.1 Movement and Action

According to the sensorimotor approach, visual perception relies on the bodily exploration of the environment and on the "structures of our biological embodiment" and "capacities for action" (Varela et al., 1993, pp. 149, 180). A prominent example in support of this claim has been taken from research on sensory substitution. By using a so-called tactile-visual-substitution-system (TVSS), blind subjects are enabled to develop quasi-visual perception. An array of vibrators is connected to the subjects' skin and they are connected to a camera mounted on the subjects' head. The activation of the vibrating impulses is depending on the subjects' active control of the camera when exploring the environment (Bach-y-Rita, 1972). Subjects have to be "behaviourally active…using head, hand or body movements" (Varela et al., 1993, p. 177). The same holds for touch: you are able to feel the shape of a specific object's surface "by your active touching (probing, prodding, stroking, rubbing, squeezing) with your hands". The "tactile impression" does not only rely on "sensations in your hands and feet" but on their "movement through space" (Noë, 2004, p. 15).

According to this, perception depends on sensorimotor interactions that consist in at least minimal *movements of the body* (ibid., p. 13). The agent has to not only engage with its environment at some point during her development, but she has to do it currently. In order to see, for instance, she has to move the camera "now". Without instantaneous movement no perceptual experiences could arise.

Noë and O'Regan argue that these claims hold only with respect to a part of cognition, namely vision, but not for cognition in general. However it is worthwhile to ask whether this theory could be extended so as not only to include the other perceptual modalities, but also to explain to what extent sensorimotor interactions shape cognition in general. In fact for Varela et al. at least the latter seems to be the case, as they claim that "cognitive structures emerge from the recurrent sensorimotor patterns that enable action to be perceptually guided" (Varela et al. 1993, p. 172).

The concept put into question by the context of LIS and BCI is 'sensorimotor interaction'. If sensorimotor interaction is required for perception then the first and narrow variant of the challenge would consist in asking what role *active bodily movement* has for perceptual experience of patients with global physiological impairment. The second variant would ask what role sensorimotor interaction associated with active movement has more *generally* for cognition in patients with global paralysis. In what follows I focus mainly on the second route.

(Whole) Bodily Movement and Classical LIS

Consider the case of classical state of LIS. The patient's body is immovable as a whole and the bodily exploration of the environment would be restricted to the small field of vision of the patient. However, there is presumably no doubt that these patients are fully cognitive.

If sensorimotor interaction consists in bodily movement in terms of the *whole body moving* and requires active exploration then perception in cases of classical and complete LIS must be affected since such global bodily movement is not possible for patients with LIS. The same goes for other cognitive abilities. If those were to emerge "from recurrent sensorimotor patterns" as

Varela and colleagues have it (Varela et al., 1993, p. 172) then a LIS patient's cognitive abilities must be limited given the fact that the LIS patient has no capacities for global bodily movement.

(Partial) Bodily Movement in LIS and BCI

Instead of demanding movement of the body as a whole, proponents of the sensorimotor approach could argue less restrictively that *some* kind of bodily movement might be sufficient for cognition. CLIS patients are still able to blink and vertically move their eyes. A weaker version of the claim that bodily movement is required for cognition can be thus reformulated: *minimal* movement may be sufficient for realizing the sensorimotor interactions (Noë, 2004, p. 13). From this perspective CLIS patients can still be seen as relying on sensorimotor interactions as blinking and saccadic movements of the eye(s) are preserved.

However, even this weaker version of the movement constraint can be put into question by LIS. Consider the case of patients in the classical (and complete) state who use BCI technologies thereby relying on their own brain activation patterns to communicate and control external devices. These BCI-facilitated interactions with the environment are achieved without movement of the body *at all* (see 3.1). A concept of embodiment defined in terms of sensorimotor interactions that involve (minimal) movement cannot sufficiently account for these cases. It also does not seem to account for cases of sensorimotor interaction that are realized in part without relying on the *biological* structures of bodily movement.[49] As a consequence one can say that both versions of the bodily movement constraint, i.e. the radical version (movement of the body as a whole) and the weaker version (some kind of movement, as say, in movements of the head) appear as *too restrictive* to properly account for general cognition in LIS and BCI.

This is clearly not to deny that cognition relies on sensorimotor loops (closed in the interaction with the environment). However it appears that such a claim needs further refinement. If sensorimotor action was to play an *active* role in driving cognition and consciousness then it has to be made explicit what this claim amounts to and this involves an elaboration on what a sensorimotor interaction actually is. The "body" concept in the crude challenge could thus be refined by the first two basic questions:

1) Do sensorimotor interactions necessarily involve movement?
2) Do sensorimotor interactions rely on biological structures alone?

4.2.2 Cognition Requires Mastery of Sensorimotor Skills

Noë has an implicit reply to the above considerations. According to him, visual cognition requires the possession or mastery of *sensorimotor bodily skills*, i.e. the practical knowledge of how to exercise these. This means that if you want to see what is around the corner, then you *know how* to move your body and head, and where to glance in order to perceive what you want (Noë, 2004, pp. 11–13). Noë and O'Regan deny that such mastery has to necessarily include action or movement to be *instantaneously* exerted (as seen in the TVSS example). Important is rather that the person possesses a knowing-how in the sense of *having* cognitive capacities and that she previously had the ability to actively exert "control over her visual input" (O'Regan and Noë, 2001, p. 1015).

This appears to provide a direct response to the discussion above. If what matters for cognition are not actual movements, but only the knowledge about what would happen *were* I to

[49] Some people may wonder whether I am not missing the point. TVSS-use is also a case of a sensorimotor interaction that is in part non-biologically realized. However there is a crucial difference between TVSS and BCI-use: while TVSS still requires active movement to replace vision with tactile stimulation, BCI relies on neuronal activity alone. When I claim that the sensorimotor approach cannot account for non-physiological components of the sensorimotor interaction cycle then I refer to those cases of facilitation or substitution devices that do not rely on bodily muscular activity.

move my body, then such an account would not be threatened by LIS. The consequence would not have to be that "a paralyzed person would be unable to see" (ibid.) as the previous discussion seems to suggest.

Coming back to LIS this would mean firstly, that even a LIS patient in the complete state (with no eye-movement) should have visual experiences (in imagination or dreams), despite of a lack of active use of her visual apparatus. I come back to this in the subsequent Sections 4.2.3 and 4.2.4). Secondly, if we were to extend the claim that what matters is just the mastery of bodily skills to cognition in general then the sensorimotor proponent could offer at an explanation of why cognition (not only visual experiences or perception) in LIS is preserved. This explanation could still contain a reference to the role of bodily movement. One could argue that bodily movement plays a role in the realization of cognitive processes in LIS but that this role is only *developmentally* relevant. If the LIS patient had engaged in sensorimotor interaction at some point before becoming paralyzed the paralysis itself will not provide reasons to doubt that she will continue to be cognitive.[50] This answer to the challenge would be a clear example of the developmental escape (see 4.1.2); it provides a rather weak embodied account of cognition in LIS. While momentary enactment of movement is necessary in situated perception as illustrated for instance by cases of TVSS, it seems no longer required for other cognitive processes or those occurring at a later moment in time. Importantly such a reply should be at odds with the invocation of the notion of skills, as I argue in the next section.

Skills require Training

If cognition were to rely on mastery of bodily skills, for which movement is only developmentally relevant, then this invites an interpretation of skill that does not involve any sort of training or effort to *maintain* it.[51] This would be a highly counterintuitive implication for it seems a rather fundamental aspect of skill that it is not preserved once for all time. You will find myriads of examples for this. Consider the following two. It can take up to ten years of constant practice to become a fairly good piano player. But even after the piano player masters the instrument she needs to continue with the practice. If she stopped practicing for a significant period of time, this would directly affect her virtuosity. Another example: I have spent more than ten years developing a decent proficiency in Russian. Yet if I stop talking to people or reading books in the Russian language, this linguistic skill will be affected, no matter how well it was developed. It will most probably not fade entirely, but after a longer break, fluency is unlikely, even a simple conversation is clearly slower and appears bumpy. The point is that in order to *become* and develop into a skill a particular activity not only requires intensive training but also an ongoing practice, or some form of update, to really *persist* as a skill. It would appear prima facie likely that this applies not only to musical or linguistic skills but also to other forms of cognitive abilities. As a consequence I suggest that bodily skills involve a double temporal dimension. They require not only an initial developmental phase but also a continuous phase of sustaining them (see Figure 4.1).

[50] Note that I do not argue against a developmental perspective per se. It is in fact part and parcel of an account of bodily skills. The objection is targeted at those proponents of an embodied approach to cognition who *merely* adopt a developmental stance: If it turns out that a developmental stance on the role of the body trivializes the body, then this can run against their own view, which in alternative to other perspectives, initially emphasizes the role of the body. Simply put, why should we stress that the body plays a role for cognition if that role turns out to be trivial?

[51] To insist that skills do not require a timely update and that cognitive processes remain unaffected once developed amounts to an implicit cognitivist and thus representationalist view of cognition.

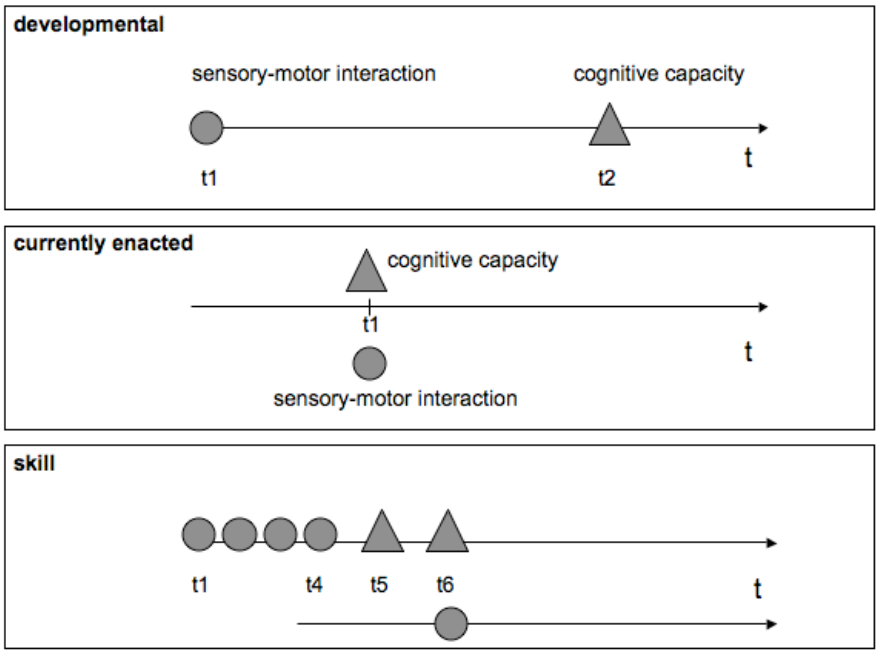


Figure 4.1: Three perspectives on the role of sensorimotor interactions for cognition. The figure illustrates three approaches to the role of sensorimotor interactions (circles). From the *developmental* perspective, they are only historically relevant. Once the cognitive capacity (triangles) has developed (at t1) they are no longer required. From a second perspective sensorimotor interactions are *currently enacted* to bring about the cognitive capacity (in this view both the skill and the enactment of sensorimotor interactions temporarily coincide). According to the third perspective (that I would like to promote), after an initial phase of developmental the skill is trained and sustained. Both phases rely on the contemporaneous enactment of sensorimotor interactions.

On acknowledging that the notion of bodily skills entails a double temporality, there would be an alternative, refined route to accounting for their role in cognition. The proponent of sensorimotor skills would not have to assume that bodily movement is *either* only developmentally relevant, in the sense that once a person has performed bodily movement she will have the practical knowledge, no matter whether she is ever going to enact it, *or* to make the radical claim that bodily movement is required for cognition at *any moment*. In an intermediate account of the role of bodily movement in skills she might argue that sensorimotor interaction loops need to be enacted for the development and every now and then thereafter to ensure that the skill (or practical knowledge) is not fading.

Sensorimotor skills and LIS – Can BCI sustain or update previous cognitive skills?

Can a concept of sensorimotor skills that involves a regular enactment of sensorimotor loops account for LIS? It seems obvious that in the case of LIS an update could not rely on bodily movement. The restriction of the sensorimotor contingencies theory to biological and movement-based aspects of embodiment should preclude a further account of LIS in terms of sensorimotor skills. However, there are LIS-patients who rely on BCI. They are obviously cognitively engaged with the environment. But they do so without any movement and with the help of non-biological processes. One could adopt a more liberal concept of sensorimotor skills (a concept that does not imply movement in terms of the whole body and need not necessarily be biologically based) and see whether such a liberal notion of sensorimotor skill may shed light on LIS and BCI communication. The question would be whether BCIs could be treated as a means for training and sustaining sensorimotor skills that originally developed through biological and movement-based bodily engagement.

Consider therefore the mechanism underlying BCI-use in a little more detail. LIS patients who use the technology show an adaptive change in the function of the brain areas involved in BCI control. In normal motor-actions signals of the motor neurons act on muscles and thus enable movement. In BCI however cortical areas that usually contribute to exert control on the motor neurons now additionally assume the role of the motor neurons and thus become themselves the "final product" of the central nervous system (Wolpaw, 2007, p. 616). BCI-use generates "new brain output pathways" (Wolpaw, 2007) and "changes electrophysiological signals from mere reflections of central nervous system (CNS) activity into the intended products of that activity: messages and commands that *act on the world*" (Birbaumer et al., 2003, p. 121, my italics). By learning how cortical areas can directly enable interaction with the environment LIS patients acquire a capacity not present prior to the training. This capacity could count as a novel non-movement based *skill*.[52]

As a consequence of this BCIs are unfortunately not easily conceived of as a means to replace or substitute pre-existing cognitive abilities. They neither seem to be only a tool of expression of already given capacities. BCI-use constitutes rather an entirely new non-biological skill. This leads to another issue that needs be addressed, namely how a new BCI-based sensorimotor interaction loop is related to pre-existing cognitive processes that were acquired through bodily movements.

In this section I have suggested that the skill-based sensorimotor approach presumes that cognitive skills require no update (via actual bodily movement). This is highly counterintuitive and leads to the developmental escape. On avoiding a developmental escape and accepting that skills require an update, one would have either have to predict that cognitive processes are affected in patients with LIS. This is because those skills cannot be preserved without further performance of bodily movement. Alternatively one may argue that cognition in classical LIS relies on timely updated skills while these are not maintained by bodily movement, but with the aid of non-bodily and non-movement based faculties (provided, e.g. by BCI). This alternative does not come without problems because BCIs appear not to replace or express previously existing skills, but require another form of interaction cycle leading to a new kind of skill. In order to claim that non-biological and non-movement based sensorimotor loops can update cognitive skills, an explication of the kind of relations that exist between previous, movement-based and novel, non-movement based sensorimotor interactions is required.

The challenge to embodied cognitive science could be refined by three further questions. These need to be addressed in order to clarify the concept of embodiment as suggested by the sensorimotor approach to cognition:

3) To what extent are sensorimotor interactions constitutive of cognition?
4) Does cognition require a concurrent enactment of sensorimotor interactions?
5) How can we relate biological and non-biological sensorimotor interactions?

4.2.3 Abbreviated Sensorimotor Interactions?

There is a way for the sensorimotor proponent to argue that bodily movement is involved in cognition in LIS patients, without adopting a developmental perspective or having to consider non-biological realizations of sensorimotor interactions.

For Hanna and Maiese (2009) the concept of bodily movement actually involves two kinds of intertwined processes. They refer to the kinds of movement discussed in the previous section such as movement of head, hands, eyes etc., or of the body as a whole, that are all visible externally, as *overt* bodily movements and "*behavioral* processes" (Hanna and Maiese 2009, p. 102).

[52] Even in BCI technologies that require no training (see 3.1.2.2), the use of the technology enables the patient to perform actions in a way she was not able to before.

There is, however, a second variant of body movements: so-called *covert* body movements, intentional neurobiological processes that precede overt movements and are not visible externally.

If this second concept of covert bodily movement is applied to the context of LIS a plausible response to the above challenges by LIS as well as several of the sister challenges could present itself (see above footnote, p. 108). A person who displays no *overt* bodily movement may very well engage in *covert* bodily movements. Consider for example a person who is sitting still on a chair with her eyes closed and merely imagining visually scanning a house. The imagination of scanning a house can be realized by the same sensorimotor circuits that are active during an actual case of visually scanning a house; with the exception that no muscular activity, i.e. no *overt* movement is involved.[53]

Covert Movements in LIS

If the concept of sensorimotor interactions was widened to imply covert activities then we could account for LIS in terms of sensorimotor interactions. But recall from the introduction to LIS that there are two general causes of LIS, a trauma to the brain stem or ALS and depending on these there are two ways of formulating another version of the challenge.

In case of trauma one could say that the patient still has the capacity of imagination and that this relies on covert bodily movement. However, this invites the question whether and/or how long these internal circuits can work entirely independent of actualization through "real" sensorimotor interactions (Clark and Grush, 1999).

For a LIS patient who lost mobility due to ALS things may be different. Here the concept of covert bodily movement seems not applicable. Covert bodily movement relies on activity of the motor neuron and ALS destroys the motor neuron. Thus the capacities of the patient to enact covert bodily movement and thus perform imagined vision should be increasingly affected as the disease progresses.

BCI-movement – a non-biological form of overt movement?

Let us come back to LIS patients using BCI and discuss a further way to refine sensorimotor interactions in terms of the distinction between covert and overt movement. BCI is not only used for communication but also for controlling a wheelchair or an artificial limb. Such cases could involve a kind of *overt*, behavioral expressive movement, which only differs from "ordinary" overt behavior in that is not biologically realized. The examples could be seen as a sort of *hybrid* sensorimotor interaction, realized in a distributed manner in that BCI software is taking the job of the motor neurons. The software is thus receiving signals from the cortex, which are transformed so as to act on the external device, in the analogous way that the motor neurons would do in affecting the muscles.

This hybrid form of sensorimotor interaction is not unproblematic. As just seen we find in BCI-use an entirely new sensori-"motor" loop. It consists of neurobiological processes directly governing external devices. But covert intentional body movements require the distributed activity of different CNS areas. In normal cases the function of these areas consists in controlling the motor neurons that are activating the muscles. In BCI control the neurons of the CNS have to change their function. They are no longer controlling the spinal motor neurons, but substitute them, and control the BCI device directly. BCIs therefore partly bypass the areas supposedly

[53] A defender of Noë and O'Regan's approach may now suggest that Hanna and Maiese are exactly expressing what Noë and O'Regan have in mind: a sensorimotor skill does not require *muscular* enactment. It could simply rely on inner movement processes. But to use Hanna and Maiese's concept to elaborate the sensorimotor approach in this way is not so easy. It is actually paradoxical because the voluntary activation even of *covert* movement still counts as a case of momentary enactment. Yet the idea of merging both approaches was precisely to support Noë 's claim that *no* momentary enactment is required.

involved in "normal" internal sensorimotor circuits. Crucially, this applies for both, LIS cases with intact and impaired motor neuron.

A notion of embodiment that allows "inner" (covert) and "outer" (overt) bodily movement seemed promising at a first glance because it offered an account of BCI based interactions with the environment. But even on a second look, the notion appears too restrictive because the BCI circumvents parts of the neurologically based sensorimotor circuits involved in *covert* bodily movements. As a consequence it seems that biological sensorimotor interaction cannot be the sole basis for cognition in LIS.

Yet, I would suggest keeping not only the baby, but also some of the water in the bathtub. Even if BCI-use is not based on "natural" sensorimotor circuits it still appears to involve some kind of *covert* activity. That embodied cognition has set off as an alternative to the classical cognitivist view does not have to mean that the inner ongoings are now completely irrelevant for cognition. If we include the range of bodily activities that are not behaviorally visible, but *inside* the body then this could lead to a further refinement of the concept of embodiment. One could refer to either the natural or the hybrid *covert* sensorimotor circuits not as kind of bodily *movement*, but rather as a type of bodily *action.* In this way we could keep the intuitive understanding of bodily movement (in humans) as *overt,* visible, expressive behaviors of the body (and sometimes artificial complements, such as BCI).

From the third part of the discussion of the sensorimotor approach arises a further question that helps refining the challenge to embodied cognitive science:

> 6) Do sensorimotor interactions always involve a loop into the environment? Can there be bodily action, which does not involve any movement nor transcendence of the cognitive system into the world?

4.2.4 Intermediate Conclusion

The basic upshot of the discussion in the first part of this chapter is that the sensorimotor approach in its present state is too restrictive to account for cognition (and consciousness) in the context of LIS and BCI. Note that I do not say that the sensorimotor approach is wrong per se. The point is rather that it needs important elaborations in order to provide a non-trivializing story about the role of the body in cognition. The above discussion entails four dimensions in which this can be achieved.

Firstly, the present concept of sensorimotor interaction is restrictive in terms of *kind.* I have shown that global paralysis in LIS resists an explanation of cognitive capacities in terms of *bodily movement*. I have suggested that cases of LIS and BCI invite us to reconsider what we take to be the ontological nature of sensorimotor interactions and possibly transcend this limitation by including *non-biological* and *non-movement* based sensorimotor interaction processes in the explanation of cognitive capacities.

Secondly, we can assess the concept of sensorimotor interaction with regards to a particular *form* (sensorimotor interaction as a *skill*) and consider therefore a *temporal aspect* of the challenge posed by LIS. The challenge is not only related to the question of *what* embodiment is, but *to what extent* it is required for cognition. There are at least three ways in which this can be formulated. The first variant is to argue that embodiment, i.e. sensorimotor interaction, needs to be contemporaneously enacted (as illustrated by TVSS). Here it appears to be momentarily and constitutively necessary for cognition. A second way is to conceive of embodiment, i.e. sensorimotor interaction, as *merely* developmentally relevant. Once a person has engaged in sensorimotor interactions she continues to be cognitive, without any further bodily engagement. We have seen that on accepting this claim, the challenge posed by LIS would be avoided and the sensorimotor approach runs the risk of providing only a weak embodied position. A third variant is that repeated sensorimotor interactions can, after an initial period of time, become a habit or a skill which is inherently prone to fade without further maintenance. This involves a *double*

temporal dimension: embodiment in this sensorimotor approach would not only be developmentally but also continuously relevant for cognition.

Thirdly, we have seen that the concept of bodily movement might also include the neurological processes preceding overt movement. This could provide further refinement of embodiment in terms of a division between *inner* and *outer* forms of sensorimotor interactions. While it is not clear whether BCI control counts as a case of sensorimotor interaction proper, it definitely relies on some kind of inner bodily activity. This activity is maybe best described as hybrid (brain and BCI). I have suggested that it is useful to widen the concept of sensorimotor interaction to also include processes *within* the body. But we may require a similar freeing from the idea that only biological structures can underlie sensorimotor interactions.

In the second part of this Chapter I assess the extended cognition approach, which essentially adopts a functionalistic stance on the role of the body.

4.3. Refining the Challenge II: The Extended Functionalist Approach to Embodiment

A second variant of the embodied approach to cognition is based on the extended cognition approach introduced in the first chapter (1.3.3). There I have mainly focused on the controversial claim that cognition is extended beyond brain and organism in that it comprises parts of the external environment. Note however that this claim is only one of two assumptions made by the extended cognition approach. In addition to an extension of cognition beyond the boundaries of the body there is also already an extension of cognition from neuronal processes to *embodied* processes. In other words, the extended approach is also committed to embodiment (Clark, 2008b).

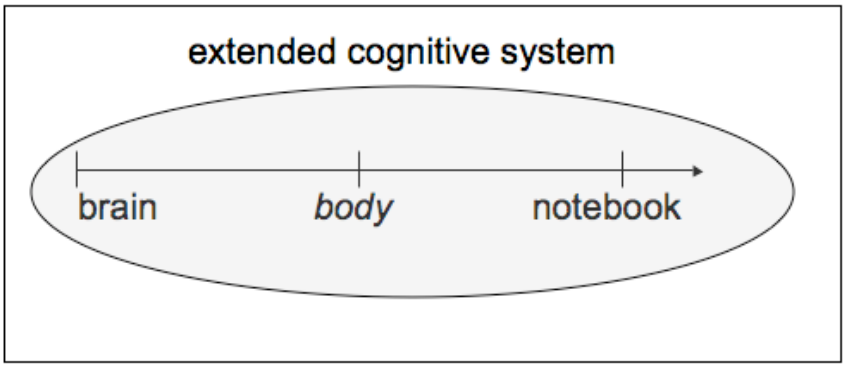


Figure 4.2: The extended cognitive system. The extended cognitive system is classically seen as a distributed computational whole of which the body is one component[54].

What this commitment amounts to is unfortunately not quite clear. What seems clear is that because extended cognition is a version of functionalism it inherits the functionalist liberalism concerning the material basis of the cognitive system (see also 1.4.2). Recall that according to functionalism mind can be defined in terms of abstract internal mental states that are characterized by the input processes causing them, their relations to each other, and the behavioral output they cause. What matters thereby is not how exactly these processes are materially implemented, but that their implementation realizes the overall functional profile of an abstract mental state. This is the multi-realizability claim that the functionalist considers essential for every theory of mind (Braddon-Mitchell and Jackson, 1996, p. 44).

[54] Though there are now recent approaches to extended cognition that depart from the claim that this system is computational (e.g. Silberstein and Chemero, 2012).

According to "extended functionalism" this liberalism holds analogously for the role of the body and the status of particular environmental aspects for cognition. Even though Clark admits that in the actual human world cognition happens to be realized by brain, body and particular parts of the environment. The body thus usually plays a role in constituting cognition. However, this role is by no means special and the body needs not necessarily to assume it. The body matters for extended cognition only as part of what Clark called a "larger mechanism story" (2008b, p. 39). As Clark says:

> [I]nsofar as the bodily platform matters, it matters by virtue of the suite of abstract opportunities (encodings, operations) that it makes available, and contrariwise, the suite of encodings and operations that it makes unnecessary […]. (Clark, 2008a, p. 201)

This liberalism places the extended cognition approach at far distance from the previously discussed sensorimotor perspective that assumes "*full-and-principled* sensitivity to all the details of a being's embodiment and/or sensory apparatus" (Clark, 2008b, p. 53, original emphasis). According to him, this difference between the functionalistic and the sensorimotor approach creates a non-reconcilable tension in "embodied, embedded" cognition (Clark, 2008b). While it is certainly right to consider that the body plays a role for cognition, Clark strictly denies that this means that cognition is tied to a *specific* structure of the body. The sensorimotor approach to the role of the body should be rejected for its implicit assumption that "sameness of mental state requires sameness of bodily structure" (Clark 2008b, p. 51). Such a claim would amount to the same "unacceptable meat or species chauvinism" as the attempts of reducing mind to neural activities in the brain (ibid., p. 43). The tension in "embodied, embedded" cognition can thus be resolved if the body is seen as functionally replaceable part of the "mechanistic supervenience base" of cognition (Clark, 2008b, p. 40).

Interestingly, my discussion of the sensorimotor approach led to a similar conclusion. I have argued that the emphasis on biological bodily structures and movement rendered the sensorimotor approach too restrictive and thus unable to account for a variety of challenges posed by LIS and BCI. One might thus be inclined to think that these limitations are exactly expressing Clark's worry and that extended functionalism would provide just the right alternative solution to the challenge posed by LIS.

Unfortunately, this impression is only superficial. I share Clark's worry to the degree that the role of the body in these approaches is too restrictive. But I also contend that this offers no reason to assume that the extended functionalistic view entails a better approach to the body's role in cognition. With regards to the relevance of the body for cognition conceiving of the body as a contingent part in a greater information processing system is a retreat into the exact opposite extreme. While the sensorimotor approach strongly relies on *particular* bodily structures (defined in terms of muscular movement) extended cognition rejects not only that these structures have to be of a particular kind but apparently also the idea that we need *bodily* structures at all.[55]. I do not mean that the functionalist approach has nothing to offer in order to account for the challenge posed by cases of LIS and BCI. But we will see in what follows that the role of the body in extended cognition is widely underestimated. It is so loose that extended cognition appears not embodied at all, despite declarations to the contrary.

From the perspective of extended cognition BCI could serve as a practical example in support of extended cognition (Fenton and Alpert, 2008). This would require to assesses whether BCI qualify as a part of the cognitive system and thus pass the parity principle (Clark and Chalmers, 1998). Recall from Section 1.3.3 that the parity principle applies if the external resource allows resolving a problem in the same way as e.g. the brain does. Indeed, Fenton and Alpert hold that BCIs are "functionally integrated devices [that] will extend a patient's cognitive as well as physical capacities" (Fenton and Alpert, 2008, p. 127). Accordingly, the LIS patient and the BCI coupled to her brain activity would constitute a kind of extended cognitive system.

[55] Which is of course not to say that they deny that some sort of implementation is necessary.

In the following four sections I discuss whether BCIs extend a patient's mind. In the beginning I defend Fenton and Alpert's claim against a recent objection by Sven Walter. But I will then develop an alternative line of reasoning against the idea that BCI extend the mind. This is based on the assumption that extended cognition entails a very weak concept of the body.

4.3.1 Why BCIs are not a Case of Extended Cognition[56]

Sven Walter has rejected Fenton and Alpert's attempt to apply the theory of extended cognition to cases of BCI-use. In order to count as a vehicle of cognition it "would have to realize the functional role characteristic of a cognitive faculty or process" (Walter, 2010, p. 67). In the case of BCI, Walter suggests the following three candidates for a role of BCI in cognitive extension: a) the voluntary control over an external tool, b) allowing the patient to act and c) to communicate. According to Walter, a) and b) would be bad examples, because the movement of an arm or cursor is not a cognitive, but a bodily process. At a first glance, c) seems to be thus more suitable: BCIs allow patients with LIS to communicate and express their thoughts. The problem with this, Walter says, is that the cognitive part in communication is still done in the head because the EEG does not "contribute to formation of thought or intention to communicate" (p. 68). The extension Fenton and Alpert refer to is thus only an extension of bodily, but not of cognitive processes; BCIs do not substitute or extend communication or any process that we would call cognitive. This is why, according to Walter, the extended cognition framework is not a suitable framework to account for how BCI affects patients with LIS.

I believe this raises a series of worries. Firstly, remember that the proponent of extended cognition accepts that certain bodily processes count as cognitive; he does not argue that the body does not matter for cognition. Examples in support include cases of mental rotation (Clark and Chalmers, 1998) and gestures (Clark, 2008a, pp. 123–127). Now imagine that the LIS patient uses a BCI to control an artificial arm, for performing "artificial gestures". If, according to Clark, gestures are cognitive, then following the principle of parity, BCIs used for moving an artificial limb can be seen as a part of constitutive mechanism realizing the cognitive process of gesturing. They could count as a case of extended cognition after all. An analogical example would be the game Jigsaw[57]. Pieces of a puzzle are moved, probed and turned around as to compose a bigger image. In all of these cases bodily movements realize a process that is cognitive.

Secondly, consider the BCI control of an external cursor. The patient might use a cursor to draw arrows between squares and bubbles of a virtual mind map on a screen. He might thereby order them in a specific logical sequence or connect them in a way so as to display an abstract property of the relations that hold between different parts of a more complex whole. Bringing items into a logical order, or conceptualizing a dynamic relationship clearly counts as cognitive. The proponent of extended cognition could hold that this is a case of extended cognition, where in order to perform a complex mental task movement of an external tool allows outsourcing parts of the memory process as well as of our abstraction capacities.

Thirdly, there is a recent, highly innovative elaboration on BCI technology used for letter selection on a screen allowing patients to learn painting with their brain-activity. Instead of letters or numbers and links for communication or manoeuvring the Internet, patients can select different commands for painting, varying for instances colors and brush sizes (Kübler et al., 2008). Painting does not solely happen in the head, in the sense that the painter thinks of an image and then just puts it out there. It involves bodily movement and checking back constantly with the result on canvas. If painting counts as a cognitive process, involving bodily movement and the emerging product on canvas, then BCIs that enable one to move a virtual brush on a screen might be considered a case of extended cognition.

---

[56] This section is for the most part based on a recently published article of mine (see Kyselo, 2011).

[57] I owe this example to Ezequiel Di Paolo.

Consider finally Walter's claim against BCIs as constituting communication. He says that EEGs do not "contribute to formation of thought or intention to communicate" (Walter, 2010, p. 68) which is why BCI cannot be considered a case of extended cognition. This line of argument, it seems, implies a classical Fodorian view of the relation between language and thought (Fodor, 1987). Fodor considers language and thought to be two disparate things. Thinking might have structures similar to those of natural language; the latter, however, is just a means to communicate the first. BCIs only help a LIS patient to express thoughts in a linguistic and visually tacit fashion, but they are not involved in any thinking process. However, such a perspective on the relation of language and thought stands in contrast to the one advocated by proponents of extended cognition who reject the Fodorian distinction between language and thought, and consider language to be in fact a prime example of extended cognition (Clark, 2008a, pp. 50–58). If language and thought are intimately intertwined and words "augment human computation", as Clark has put it, then BCIs realizing the formation of words based on letter selection, could in principle be seen as a part of cognition.[58]

In my view, Walter's objections reflect an implicit cognitivist epistemology of a kind that the proponents of extended cognition precisely seek to overcome. It appears that he believes only those processes to count as cognitive that are in fact happening in the head, or in the brain. This might be because he overlooks the fact that extended cognition is not only a theory about the brain in interaction with external tools (cognition as transcending the brain), but supposedly also about cognition as embodied, which entails an extension from brain to body (see introduction of this Section 4.3). To be fair, I also believe that the reason *that* Walter would neglect the body as playing a cognitive role precisely lies in the shortcoming of extended cognition itself to sufficiently explicate what the body is and what this role amounts to.

4.3.2 Can BCIs Functionally Substitute for the Body?

Recall again the original issue in this discussion, namely to examine what embodiment in functionalism actually is. The above considerations show that the extended functionalist has a story ("larger mechanism story") to tell about the role of non-biological environmental tools in LIS. Yet even though we could understand on a case by case how BCIs in the examples above are substitutes for the body, what a body is and what it role generally amounts to remains unclear.

What does the idea that BCI substitutes the body tell us about the body and its role in cognition to begin with? According to functionalism, there is a "cognitive role of the body, which is to act as a bridge enabling biological intelligence and the wider world to intermingle in the service of adaptive success" (Clark, 2008b, ibid., p. 57). Based on that we would arrive at the following explanation of why LIS patients who use BCI are cognitive even though they are globally paralyzed: If the body plays a functional role for enabling adaptive success and this functional role is not dependent on a particular implementation then, in principle, the BCI technology could take over the functional role that the paralyzed body now fails to fulfill. BCIs would act as a substitute for those parts of the overall "intelligent organization" that have so far

[58] An anonymous reviewer has objected that both Walter's and my claim are misguided, because communication based on BCI differs so greatly from ordinary communication that a discussion of whether or not BCI communication is extended is futile to begin with. While ordinary communication may be extended, BCI communication is not. This objection seems to be a spin-off of the classic argument by Adams and Aizawa who claim (vs. Clark and Chalmers) that the differences between processes that are supposedly extended and original cognitive processes are so great that the former could not possibly count as cognitive (Adams and Aizawa, 2010). Now, Clark has conceded that the symbols in Otto's notebook vary greatly from the encodings of our biological memory. However, in order to qualify as a physical vehicle of mental content, sameness of kind is not a necessary condition (Clark, 2008a, p. 90). Importantly, by accepting that "ordinary communication" is extended, all of this seems to be warranted. The problem with the objection is that it reintroduces Adams and Aizawa's worry while at the same time rejecting it. If one allows for Clark's vehicle-liberalism for the case of "ordinary communication", why would one resist doing the same for the case of BCI communication?

been implemented by the body (ibid., p. 56). As a consequence, that LIS patients cannot move their bodies would not pose a problem to extended cognition because cognition is realized by the interplay of brain and BCI.

This story is short and neat. But it is not sufficient to account for cognition in LIS and BCI. To claim that an artificial component can substitute for a bodily component is only telling if one also knows what the external tool actually replaces. In other words in order for BCI to count as a substitute, one would need to first spell out exactly what it means to say that the biological is "in service of adaptive success" and to clarify the extent to which BCI would replace this service. To say for instance that BCI replaces the role of the body in gesturing says nothing about the role of gesturing for cognition.

There are two further and related options for the extended functionalist to account for LIS and BCI. The first considers the technology to substitute not only for the body, but also for the functionalistic role that the body occupies. The second approach argues the direct opposite, i.e. that in LIS the patient can still rely on neuronal processes. I show that both options inherit a similar problem.

4.3.3 Can BCIs Substitute the Body and the Functional Role Underlying it?

A response to the above worry could be the following: While cognition is normally realized by brain, body and features of the environment, the assignment of functional roles to specific material substrates does not necessarily require such a threefold contribution. The functionalist might claim that BCI-use in LIS illustrates not only that the biological body's role is replaceable, but moreover that the role occupied by the body needs not to be played at all. Brain and BCI technology alone can form a new computational whole (assuming now two, not three possible functional roles). According to this, LIS patients may be said to be cognitively intact despite bodily impairment because their cognitive processes are realized by their brains and the BCI.

Unfortunately, this reply does not resolve, but only replace the problem with another one. To grant that the body and its role are not a matter of necessity would still require a clarification of what the cognitive process or capacity is that a distributed supervenience basis consisting of brain and BCI (without the body) is solving and how it relates to the cognitive system?

4.3.4 Emulation as Internal Substitution of the Body

Recall that the original issue was not only to account for BCI technology, but rather, what (organismic) *embodiment* in the extended approach amounts to. One might thus expect that the extended approach should also attempt to account for patients with LIS who do not use BCI – and are thus not considered a case of cognitive extension beyond the organism.

One possibility for extended cognition to account for LIS patients in the classical state is to argue that their cognitive processes are realized by an internal "emulator circuitry" in the brain (Clark and Grush, 1999, p. 13). That the body sometimes helps minimizing the cognitive load does not rule out the possibility of going back to default and that cognition is realized by internal states of the brain alone. In the globally paralyzed patient neural circuitries can have the functional role to "act as *de-coupleable surrogates* for specifiable (usually extra-neural) states of affairs" (ibid., p. 8) and thus to emulate bodily processes.

This response adopts a strategy similar to the previous one. Here the body's contribution is substituted not by BCI, but by neurons emulating motor activities. The realization of the cognitive mechanism is neither distributed in a threefold manner (brain, body, environmental aspect) nor in a twofold way (brain and BCI, as environmental aspect). It is not distributed at all and provided solely by the brain's contribution. As Wilson has put it:

> Many centralized, allegedly abstract cognitive activities may in fact make use of sensorimotor functions in exactly this kind of covert way. Mental structures that originally evolved for perception or action appear to be co-opted and run "off-line," decoupled from the physical inputs and outputs that were their original purpose, to assist in thinking and knowing. (Wilson, 2002, p. 633)

However, this strategy inherits the same problem as the attempt to substitute the body's role by the BCI. Saying that the brain takes over the job of body and/or environmental components immediately does still not answer what exactly this job has consisted of.

In addition, the emulation-approach encounters a problem already present in the sensorimotor discussion. I have suggested that a sensorimotor skill requires a continuous update to persist as a skill. How plausible is it then to assume that the inner circuits can remain entirely independent from actualization through overt movement? If the answer to 'how plausible is it to assume' cannot be yes, then an account of LIS in terms of extended functionalism seems to be essentially disembodied. It appears to escape into a developmental response.

Let me now conclude the discussion of the extended approach to the body. The extended functionalist has basically two ways to account for LIS. In both the body plays only a trivial role. It is either replaced by environmental features, such as BCI – or, in the other case, by the brain. According to the latter option, the brain can substitute for bodily function and constitute cognition by relying on inner representations. If we were to designate the body's role for cognition this would again constitute a case of a mere developmental response – bodily activity has been relevant in the past, but now that the patient is immovable, it is no longer needed. We thus encounter the same problem as in the sensorimotor approach of Noë and O'Regan (see Section 4.2.2).

When it comes to cases of BCI interaction, the extended approach appears to offer an alternative to the restrictions of the sensorimotor approach. BCIs are driving cognition as external constituents of a distributed cognitive mechanism. But this story circumvented the original aim, i.e. to provide an explanation of LIS and BCI from an *embodied* perspective. To argue that BCI replaces the body's role in constituting cognition tells nothing about the body's contribution and more crucially, it can do entirely without the body.

At least for the problem at stake, extended functionalism cannot seem to get away by basing arguments on the multirealizability claim. LIS and BCI-use present real world cases of human beings whose cognitive functioning is supposedly preserved despite heavy bodily impairment. Even if we grant that the body is one implementation of a functional role realizing cognition in LIS patients, the question to be answered is what exactly is that functional role?

I believe that an answer to this question is impossible without answering what cognition is to begin with and how it is related to its realizing material substrates. Even if we accept that the body is replaceable by particular surrogates then we need to know what it is, that these alternative surrogates are constituents of. If the body is supposed to be a vehicle of cognition, yet we do not know what the thing is it act as a vehicle for, then to assess the role of the body in extended cognition seems an impossible endeavor.

The confrontation of the extended functionalist account of embodiment with LIS and BCI thus reveals two worries. Its account of embodiment is too loose – where this looseness reflects a deeply rooted problem, namely the lack of a clear account of what cognition is in the first place. Or it offers a representationalist solution, by escaping into the developmental response, which unfortunately offers a trivial story of the body's role.

According to Clark it would be a mistake to infer from the fact that the body does play a role that this role depends particularly on the body. As the above considerations suggest, it seems however similarly misleading to criticize a restrictive notion of embodiment (the sensorimotor approach) by replacing it with another notion whose inherent liberalism turns it so loose that it might account for Martian cognition, but not for that of human beings. If we were to look for an embodied approach as an alternative to the classical cognitivist epistemology, the functionalistic approach would be a clear non-candidate.

4.4 Conclusion

Let me give a final conclusion for the entire chapter. In the first part I have discussed the extent to which the sensorimotor approach is able to account for cases of LIS and BCI-use. I have argued that equating embodiment with sensorimotor interactions that rely on physiological bodily movement and require no update fails to provide a satisfying answer to the challenge posed by LIS and BCI-use. The sensorimotor approach in the spirit of O'Regan and Noë is too restrictive. This is of course not to deny that sensorimotor interactions are of importance to cognition. Quite to the contrary, the confrontation has revealed important hints and questions that can guide and refine the further discussion and research on the role of sensorimotor interactions.
The main questions to be addressed by future research of embodiment with regards to sensorimotor interactions are:

1) What is a sensorimotor interaction? Does it require physiological structures? Does it require movement?
2) What is a (sensorimotor) skill?
3) Is bodily action necessarily overt?

The second candidate approach was extended cognition. I have argued that despite its advantage over the sensorimotor approach to account for the role of external and artificial add-ons in driving cognitive mechanisms, it offers no better alternative. To conceive of the body as a contingent part in a greater information processing device which can be replaced by brain or BCI actually dissolves the body's role altogether. The advantage turned out to be explanatorily void.

Nevertheless, extended cognition offered some refinement for future elaborations on the concept of embodiment. On the one hand it emphasized the idea that embodiment might not be limited to organismic components. On the other it revealed that a proper account of embodiment and its role for cognition requires explicating what cognitive processes and cognitive systems respectively are. Both issues are inextricably linked and cannot be addressed in an isolated manner.

Interestingly it turns out that extended cognition and sensorimotor approach have something in common. By adopting a mere developmental perspective both positions face the danger of trivializing the body's role for cognition such that it becomes questionable whether they can count as an alternative to the orthodox view whose outstanding characteristic ironically is to bring the body back into the explanation of cognition.

Another shortcoming that applies to both positions is their lack of treatment of conscious experience. Despite the stated emphasis on the essential role of the body for cognition, a split remains between the body and consciousness itself. The body's role is conceived of as a means to realizing the representation of the external world. This seems to hold independently from arguments that support or discount the claim that in doing so the body's role is merely contingent. Extended cognition classically denies that consciousness can be realized by something outside the brain (Clark, 2009) and Noë's approach seemingly allows for the existence of phenomenal content without bodily engagement. In both views consciousness is separated from the world, and seen as "enclosed in the body" (Merleau-Ponty, 2002/1945, p. 69). The body is an external and material object that can be perceived just as any other object in the outside world. This shortcoming is also observed in Thompson's objection to the sensorimotor perspective on subjectivity. In his view, the sensorimotor approach does not accommodate subjectivity *itself*, but only the subjective experience of perceived *objects* (Thompson, 2005, 2007, Sections 2.3.1, 2.3.3 this thesis).

The confrontation of both approaches with cases of LIS and BCI revealed that the tension in embodied cognitive science is, contrary to Clark (2008b), not resolved at all. It rather shows that there are many aspects and dimensions in embodiment that are in urgent need of elaboration. This elaboration implies to answer how bodily activity is related to cognition without

assuming that the body's role is contingent or merely developmentally relevant. It should explicate various forms of bodily action, which must not be restricted to physiology, movement or inner processes alone. Finally it should relate embodiment and bodily action to consciousness and conscious experience.

In the following chapter I discuss how a phenomenological approach to the body fares in the assessment of the body's role in LIS and BCI.

Chapter 5.

LIS and BCI – a Challenge to Embodiment, II:

The Phenomenological Approach

> We cannot remain in this dilemma of having to fail to understand either the subject or the object (Merleau-Ponty, p. 82, 2002/1945)

In this chapter I consider a phenomenologically inspired embodied approach to cognition. The chapter consists of three parts. In the first two sections I introduce Merleau-Ponty's concept of the *habit body* and confront it with different cases of LIS. In Sections 3 and 4 I introduce another aspect of the phenomenological body, namely *self-affection*, as put forward by Henry and also discuss to what extent it can account for LIS.

The phenomenological concepts of embodiment defended in the following sections are able to elaborate on the concept of embodiment by linking it to subjectivity and reassigning the body's role for cognition as relevant beyond developmental purposes. They thus refine the notion of bodily action and account for aspects of the challenge posed by LIS that the sensorimotor approach and extended cognition were not able to. At the end of this chapter I provide an overview capturing the refined concept of bodily action. However, it turns out that the phenomenological approach still has limitations with regards to understanding the constitution of bodily experience from an organizational viewpoint. It does also not properly account for the role of the body in social interaction. I will thus conclude that the phenomenological approach to embodiment is enriching, but not sufficient to properly account for cognition in LIS and BCI.

5.1 The Phenomenological Concept of Embodiment I – Merleau-Ponty

Husserl has been positively referred to by some of the defenders of enactivism, but despite his emphasis on the constitutive role of conscious experience, there remains within his position a split of the world and conscious experience. One might say that Husserl has argued for the direct opposite of cognitivism. While cognitivism holds that the world is absolute and pre-given, Husserl contended that the appropriate departing point of understanding is an absolute and prior conscious subjectivity (Husserl, 1992/1930). Merleau-Ponty has criticized Husserl specifically for this assumption. According to him, phenomenology should not be considered as an analysis of what occurs to an absolute consciousness, but rather as the "study of the *advent* of being to consciousness" (Merleau-Ponty, 2002/1945, p. 71, original emphasis). In line with Heidegger, Merleau-Ponty held that prior to any analytical or scientific reflection, phenomenology has to acknowledge that we are already situated and embedded in a human world and thus account for this pre-reflective (or pre-objective) existence, our being-in-the-world (Heidegger, 2001/1927). Crucially, this being-in-the-world, Merleau-Ponty argues, is revealed through our embodied existence, i.e. through the perceptual capacities of our bodies. In the following I introduce Merleau-Ponty's approach to the body.

5.1.1 Being-in-the-World and the Habit Body

For Merleau-Ponty, the body is the vehicle of our being-in-the-world. It thus is the very means by which we can perceive it. The body is not only an object in an externally and absolutely given world that can be perceived as any other material thing from a third-person perspective. A body is also primarily a living and thus subjectively experienced body and its "function" can only be fully grasped "by enacting it" (Merleau-Ponty, 2002/1945, p. 87).

Merleau-Ponty distinguishes two different aspects of our embodied existence: the *habit body* and the *body at that moment*. The body at that moment refers to our immediate personal experience in a specific moment, say, a sadness arising through the loss of something deeply valued. However, personal experiences of the body at the moment are already bound to other experiences. Even though, in this particular moment, one might be entirely immersed in grief, this experience does not concern the totality of existence – the body continues to strive towards further and different experiences. This striving hints at another more general body, the habit body. The habit body is the organism (ibid., p. 97) which is permanently there in the background of our personal existence. When enacting our body we come to understand that part of the bodily structure of the habit body is a sense of directedness towards the world as well as an experience of an "I can", i.e. the felt "potentiality of a certain world" (ibid., p. 122).[59] These bodily structures shape our experiences. But that they do is however not immediately evident because most of the time our "personal existence represses the organism" (Merleau-Ponty, 2002/1945, p. 95). However, even if we are not usually aware of it, our habit body constantly perceives a world, and this world serves as the very basis for all other experiences at the personal level (ibid., p. 105).

5.1.2 The Habit Body and Bodily Impairment

Reconciliation of Body and World through Existence

To clarify the bodily structures of *ongoing directedness* towards the world and the *felt potential* to act, Merleau-Ponty considers cases of bodily impairment in which the "flow" in our every-day encounters with the world is disrupted and we experience bodily alienation, e.g. due to disease or immobility (see also Chapter 4). Such experiences can be part of comparatively normal situations, such as having a migraine or a toothache. But they become much more evident in the case of the phantom limb phenomenon. Patients with a phantom limb suffered the loss of a limb yet still continue to have the experience that they possess the lost body part. In this "denial", we can see, according to Merleau-Ponty, that and how our striving towards the world still structures our experience:

> What it is in us which refuses mutilation and disablement is an *I* committed to a certain physical and inter-human world, who continues to tend towards his world despite handicaps and amputations and who, to this extent, does not recognize them *de jure*.
>
> In the self-evidence of this complete world in which manipulatable objects still figure, in the force of their movement which still flows towards him, and in which is still present the project of writing or playing the piano, the cripple still finds the guarantee of his wholeness (Merleau-Ponty, 2002/1945, p. 94).

[59] At least in men. Iris M. Young (1980) observed that the phenomenology of the feminine body differs fundamentally from the one introduced by Merleau-Ponty. Women in patriarchic societies experience their body not with the same transparency and potentiality as men. For example, rather than having experiences of a bodily "I can", their bodily existence is inhibited and "withholds bodily commitment to the end in an self-imposed "I cannot" (Young, 1980, p. 146).

While the "manipulatory movements" given through the habit body (the limb) may be missing in the experience of the (personal) body at this moment, the patient can still perceive objects in the world as "manipulatable" (ibid., p. 95). This cannot be explicated from a third-person perspective, say as a misrepresentation of the brain or material dysfunction. It can neither be explained psychologically, e.g., as an act of ignorance or mere imagination. We do not merely think that we do not want to give up on the lost limb. The reason why the lost limb is experienced as "quasi-present" is that (other) bodily structures continue to provide the patient with a perception of an impersonal and external world (ibid., p. 98). A phantom limb patient is still embedded in a world filled with objects and other people that "invite" interactions that she is no longer able to perform. If, e.g., an insect loses a leg and substitutes this loss with a healthy leg, then not because the leg was "automatically put into operation" or because the insect "is aware of an aim to be achieved" (ibid., p. 90). The insect accommodates the restriction because it is still embedded in a world and directed towards it.

As a consequence of this there is no split between an objective world and the body. What we take to be an objectively and constantly given world is the result of our embodied existence in the world, which through an ongoing directedness – whether or not we are personally aware of it – provides us with a relatively stable impression of our surroundings (habit body).

No Split between Consciousness and Body

For the same reason it is also not possible to "relate certain movements to bodily mechanism and others to consciousness". The habit body and the personal realm are linked and integrated in our existence so that how we perceive the world depends on the experienced structural aspects of our bodily existence. Bodily capacities and consciousness are inextricably linked as being "both directed towards an intentional pole or towards the world" (Merleau-Ponty, 2002/1945, p. 101). We can therefore not choose between *either* an empirical *or* a psychological perspective:

> It is because it is a pre-objective view that being-in-the-world can be distinguished from every third person process, from every modality of the *res extensa,* as from every *cogitatio,* from every first person form of knowledge – and that I can effect the union of the 'psychic' and the 'physiological'. (ibid., p. 92)

The task is thus not to determine the relation between consciousness and the body as an material object, as suggested by defenders of the classical explanatory gap-argument, but to account for the very possibility of consciousness in living beings. This cannot be done from a third-person perspective alone because living beings are in principle "possessed of an interiority" (Thompson, 2007, p. 236). The issue is to spell out how exactly the living and the lived body are integrated in a world-directed active existence (see Section 2.3.1).

Consider again the case of the phantom limb. The amputation of an arm is a dramatic example of a "rearrangement" of the habit body as it changes basic capacities in mediating the subject's intentions with the world (Merleau-Ponty, 2002/1945, p. 164). However, the patient's interactive space is not enlarged but diminished. However, what the phantom limb phenomenon has in common with the other, non-severe cases, such as dancing, is that the world does not cease to appear as a place to be engaged with. The amputation only perturbes but does not change the habit body's overall intentional structure of being directed and the experience of the "*I can*". The question is whether the same can be said for cases of global paralysis such as LIS, in which not only parts of the body are affected but also its overall capacity to move and, arguably, to relate to the world.

5.2. Merleau-Ponty's Concept of Embodiment and LIS

At a general level the phenomenological approach to embodiment would deny that the body could be fully accounted for from a third-person perspective. This would reduce the patient's body to a physiological object and the impairment to an observable malfunctioning in the mechanism of the biological body. Subjective experience is not just optionally added to considerations of the physiological body, but a basis for describing the nature of embodiment as it is experienced by the person *being* that body.

Approaching the challenge (to understand in the context of LIS the phenomenological approach to the body) from this perspective can therefore not consist in trying to understand LIS as a physiological impairment, in which experience is the "condition of an object" (Merleau-Ponty, 2002/1945, p. 141). But neither can we simply explicate LIS in terms of a disembodied first-person perspective.

The appropriate perspective requires treating the patient, as a "human subject as an irresolvable consciousness…wholly present in every one of its manifestations" (ibid., p. 138). This means for the present context that LIS should be assessed in relation to the *whole existence* of the patient – in her lived corporeality and with respect to the above-mentioned structural aspects of that lived corporeality: the *directedness* towards the world and the *"I can"*, i.e. the felt potential of being able to realize one's intentions and projects by providing oneself with new "personal" worlds.

The case of the phantom limb illustrates the nature of our habit body as being constantly directed at the world. The loss of a limb does not affect the body's general ability to provide us with an outside world. But should we expect the intentional structure of the *habit body* to remain equally unaffected when it comes to cases of *global* bodily impairment? In my view the answer should be negative. This is because in contrast to the previously discussed alterations of the habit body, the restrictions in LIS are of the most radical kind. In CLIS where eye movements are still preserved, the patient can provide herself with a world given to her in the small field of vision; yet, despite of this, the patient is not able to move her body at all. Even more terrifying is the situation for the complete LIS-patient (TLIS) who cannot even move the eyes. If our embodied existence as being in the world crucially depends on having a relatively stable organismic incarnation that is *put to practice*, the question for the phenomenological concept of embodiment is: What sense of directedness or ability of relating to the world can patients with CLIS or TLIS be said to have? What kind of *being-in-the-world* conditions the existence of the LIS patient when her embodied being is so radically restricted?

5.2.1 LIS and "I can"

I begin to address these questions by first considering the impact of LIS on the "I can", the felt potential of the patient. The perceptual realm of the patient is diminished to an extreme extent. In contrast to the phantom limb, LIS affects the *body as a whole*, the restrictions are not temporary, but presumably irreversible. In less restrictive cases of bodily impairment objects remain manipulable as such, even though they might be temporarily not manipulable *for* the patient. This is because the patient still has other means to continue providing herself with an impersonal world of manipulable objects.

However, to what extent can the habit body retain the "*I can*", the potentiality to manipulate objects, when this potentiality depends on (some sort) of continuous bodily engagement with the world? Remember that the LIS patient has not only lost the ability to manipulate objects, but that she can also no longer (or hardly) confront herself with any objects that could be manipulated. Because the paralysis in LIS is global this basic capacity of the body to relate to the world is lost. The phenomenological approach to the body seems to imply that a condition like LIS would not only affect the patient in that she loses her ability to move. If consciousness, according to Merleau-Ponty, is intrinsically tied in with the capacity of relating to the world, we should expect that the loss leads not only to perturbations of the body at that moment but also that it has an impact at the structural level. One might hypothesize that even the potentiality of having experiences might cease to persist because there is no world available against which a loss of ability to engage with could be experienced as a deficit. The patient should ultimately lose her personal conscious experience of being potentially able to engage with the world and of objects in the world as to be potentially engaged with. It is questionable whether LIS leads only to a "loss in differentiation in the function […]" (ibid., p. 85). What seems to be at stake in this case is rather that the patient's very ability to provide herself with a world.

5.2.2 LIS and Directedness

Let me now consider a second structural aspect of our embodiment, i.e. the sense of directedness towards the world. Can the patient continue to be directed even though she lacks any self-transcending form of enacting this directedness? Or is it conceivable that while she is unable to realize her intentions, she could still have them, and may remain committed to them, not ceasing to strive towards an interaction with the world?

In response to the latter question one might argue based on the consideration in the previous chapter, that bodily action should not only include *overt* movements but that the 'intentional arc' (Merleau-Ponty, 2002/1945, p. 157) begins "below the surface" of the body (see 4.2.3). Bodily action could then be understood in a broader sense as a directed process originating in the individual. The commitment towards the world itself could be considered as an act, namely the *act of relating*. As a consequence, being directed to the world could involve two aspects: 1) an action referring to the envisaged kind or nature of relation to the world, viz. an intention, which is not movement based and 2) an action that transcends the individual and into the world. In this way one could account for cognition in LIS by saying that while the second kind of action is restricted or prevented by LIS, the first remains unaffected. Unfortunately this would imply that inner directedness is preserved even though there is no possibility to enact it beyond one's inner realms. It would mean to consider the two actions as independent of one another. To maintain this position would amount to a kind of developmental response, according to which bodily movement is not relevant for cognition after an initial period of development.

I would suggest in contrast that the directedness towards the world is not independent from one's ability to (bodily) enact one's intention. In everyday life the failure to achieve a goal, the suffering that accompanies the failure and the inability to just let go, not only temporarily, but

sometimes even for longer periods of our lives, just show how deeply we are committed, and thus incessantly are directed to the world. Likewise it would seem that there are cases of bodily restrictions similar to LIS (e.g. a person being tied on a chair) for which it seems counterintuitive to expect a loss of basic bodily capacities such as the directedness towards the world. But there is a crucial difference between such cases and the case of TLIS. It is not with respect to a particular wish or desire that patients have to accept limitations, but rather with respect to the general possibility of ever being able to realize a wish or desire. In TLIS actions can neither be realized nor even justifiably be imagined to being carried out. To what extent will that allow the TLIS patient to remain directed towards the world? Can we still say that the patient is embedded in the world?

From an external, third-person perspective one might quickly respond that since we can see the patient, say, in a room of a hospital, containing objects as well as other people, the patient clearly is in the world. Furthermore we should not forget that even though the patient might be unable to move, she is still sensitive to the world, for instance with regard to particular auditory and visual experiences. Yet this is most likely not what Merleau-Ponty had in mind. To be embedded in the world is not to be physically located *in* the world, as a physical objective space to which we are passively perceptive. Embeddedness is not a given, it is a relation that is established based on an active engagement with the world. We are thus embedded in the world *qua* our active embodied existence.

Now recall from Chapter 3 that ALS patients with TLIS show no response to attempts to provide them with BCI training (see Section 3.1.2). Birbaumer and Cohen explained this by the patient's inability to produce a distinct brain activation pattern with successive feedback (Birbaumer and Cohen, 2007, p. 626). In terms of present phenomenological perspective, one might say that the subject cannot learn a new habit of (BCI-use) because learning a habit is "to experience the harmony between what we aim at and what is given, between the intention and the performance" (Merleau-Ponty, 2002/1945, p. 167). In order for the BCI to become a part of the TLIS patient's habit body she would have needed to experience how the activation of a particular brain pattern (e.g. imagining a particular movement or place) would have translated into the successful manipulation of an external object.

However, this considers only the patient's inability to learn a new skill. It does not explicate *why* the patient does not respond anymore. *Understanding* the patient's unresponsiveness would require to presume a subjective perspective from the patient's overall existence, not just with regards to the particular action of learning a new skill. We have to approach the "human subject as an irresolvable consciousness which is wholly present in every one of its manifestation" (Merleau/Ponty, 2002/1945, p.138). On adopting this perspective, we might begin to account for the lack of response as follows: the patient experiences that she is not able to manipulate *any* objects in the world. They cease to exist as objects she can relate to. Now if the sense of directedness involved a continuous update through actual engagement with the world but such engagement is impossible then due to the lack of available objects there is nothing that this directedness could be related it. To be embedded in the world implies that our relation to the world usually takes a determining form of an ongoing urge. We are in the world with a certain need to "feed" our existence through enactment. This is as an active process, *usually* realized through bodily movements. Yet, if the possibility to realize an action is not given anymore, this need cannot be fulfilled. TLIS might thus fundamentally affect the patient's existence as embedded in the world, not only with respect to the "I can"– the felt potential of being able to realize her actions – but also with respect to her very striving to realize those actions, itself. It may be possible that if the patient never experiences the harmony between her goal and the achieved action that she ceases to be directed at the world and that the world gradually ceases to

be a world she could relate to. The world ceases to be world for her[60]. This might explain why a patient in TLIS shows no response to training attempts.[61]

Consider in support of this an experience reported by anthropologist Robert Murphy (1990). Even though Murphy suffered from a less severe state of paralysis (he became quadriplegic in result of a spinal cord tumor) his situation mirrors the difficulty in maintaining bodily directedness towards the world when every effort to move, i.e. to realize this directedness, is unsuccessful:

> For a while, I occasionally tried to will the legs to move, but each futile attempt was psychologically devastating, leaving me feeling broken and helpless. *I soon stopped trying.* The average nondisabled person could be driven to the edge of breakdown if his legs were pinioned and rendered totally immobile for long periods, and an accident victim in a body cast finds his only comfort in the fact that his situation is temporary. I was saved from this, however, because the slow progress of paralysis of my limbs was paralleled by a *progressive atrophy of the need and impulse for physical activity. I was losing the will to move.* (Murphy, 1990, p. 88, my emphasis)

Compare this to cases of depression. Common symptoms of depression are a lack of motivation, feelings of emptiness and indifference. Maybe there is an analogy to be drawn between the depressed healthy-bodied person, Murphy's case and the TLIS patient's unresponsiveness. In all of these cases we see that directedness and engagement towards engagement the world is affected. They might all suffer from some form of existential hardship that is disembbed*ing* and that might therefore lead to loss of will to engage with the world.

Let me now consider how the phenomenological approach accounts for LIS patients who use BCI.

5.2.3 BCIs and LIS

I have suggested that relating to the world in having a particular wish or desire may be seen as a kind of inner, non-movement action. Now I want to discuss how the concept of the habit body can account for the use of BCI for facilitating interaction and communication with the world. We are therefore not concerned with the "inner" part of the "intentional arc", but with those parts of the action that transcend the person's organismic boundary. This can be approached from two different angles:

1) Can the notion of bodily movement be expanded so as to account for movements that are not realized by our natural physiology, but by instruments of the external world, i.e. it is conceivable to treat BCIs as enabling a non-physiological bodily movement?

---

[60] A note of clarification: By saying that the world ceases to be a world, I really suggest that in the case of an unresponsive TLIS patient the experienced world of the patient as such might cease to exist, not in the sense that it is only deficient or turns into a disengaged world. Heidegger argued that there is a difference between the relations that material objects, animals and humans have with the world. A stone, he says, has no world (*weltlos*), because in being a stone the stone does not relate to anything. Animals have a world, but this world is poor (*weltarm*) because the interaction with the world is driven by instinct; the animal is immediately affected by the world, but it lacks a reference to itself as the being that is interacting with the world. Humans in contrast form their world (*weltbildend*) because they can detache themselves from an immediate engagement in it. They take a reflexive perspective on the world as a whole and thus on themselves as beings that are in the world (Heidegger, 1929–30/1992, pp. 284, 288, 414). From this perspective, to say that the world ceases to exist refers to human existence as world-forming. It is this property of being that could be at stake in TLIS.

[61] Again, this is not say that every action requires being immediately acted out in the world. But "inner action" is independent from the world to a limited degree only. If it is not regularly updated it should be affected at some point.

2) Is bodily intentionality restricted to movement, i.e. can there be a form of action reaching out into the world that does not involve movement, whether or not movement is physiologically grounded?

Are BCIs a Form of Non-Physiological, yet Bodily Movement?

In order to address the first question, let me return to Merleau-Ponty and consider a further aspect of the *habit body*. We have seen that the relative stability of our embodied existence is presumably given through a consistent biological structure provided by organs and reliable sensorimotor circuits. In the elaboration of the concept of habit, Merleau-Ponty however also says that "sometimes, finally, the meaning aimed at cannot be achieved by the body's natural means; it must then build itself an instrument" (Merleau-Ponty, 2000/1945, ibid., p. 169). In order to relate to something we usually rely on biological bodily structures, such as our eyes or hands for example. But in the same way that we do not relate to our eyes or hands explicitly but just use them in order to establish the relation to the world, some external tools can become part of the means by which we are in the world. Originally non-bodily objects become appropriated by our habit body and thus cease to exist for us as independent objects. Merleau-Ponty illustrates this famously with the stick of the blind person that allows substituting visual feedback of the person's position with a tactile one. The stick is no longer an external object to the blind person but "has become an area of sensitivity…providing a parallel to sight" (ibid., p. 165).

This will probably immediately ring a bell. For it was one of the major claims of extended cognition that, granted specific conditions are met, an external object can become part of the machinery underlying cognition (see Section 4.3). In the following section I show that there is however a fundamental difference between phenomenological talk of the incorporation of an external tool and the view held by extended cognition.

The Incorporation of BCI – Extended Cognition vs. Phenomenology

Remember that extended cognition classically denies that the case of extension holds for consciousness. This implies not only *a)* that consciousness is regarded as separated from our bodily existence, but also *b)* that even if consciousness was considered as independent from our embodiment, then it would be constituted exclusively by neurological processes in the brain (see Section 1.4.3). In contrast, for the phenomenological approach the body is the means by which we are conscious of "the world". "[T]he body has understood and habit has been cultivated when it has absorbed a new meaning, and assimilated a fresh core of significance" (ibid., p. 169). Learning a new habit, i.e. the very process of incorporating an external object into our habit body, thus *already* and necessarily involves consciousness, viz. an "*experience* [of] the harmony between the intention and the performance" (ibid., p. 167, my italics). Already two decades ago, Dreyfus and Dreyfus explored Merleau-Ponty in relation to embodied cognitive science and explain this idea as follows:

> According to Merleau-Ponty, in everyday, absorbed, skillful coping, *acting is experienced as a steady flow* of skillful activity in response to one's sense of the situation. Part of that *experience* is a sense that when one's situation deviates from some optimal body-environment relationship, one's motion takes one closer to that optimum and thereby relieves the "'tension" ' of the deviation. One does not need a goal or intention to act. One's body is simply solicited by the situation to get into equilibrium with it (Dreyfus and Dreyfus, 1996, p. 111; my italics).[62]

Therefore, in order to acquire a new domain of significance we need past and present embodied conscious acquaintance with the world. Only *for* and based on that bodily conscious acquaintance

[62] This is in line with recent work by Rietveld who elaborates the Wittgensteinian concept of "directed discontent" into a notion of *situated normativity* according to which in the process of applying bodily skills on objects we unreflectively reduce an experience of dissatisfaction (Rietveld, 2008, p. 980).

a new domain can at all acquire the status of being significant. But this claim would be at odds with *a)* the claim of extended cognition that consciousness and body can be separated.

Now, if according to this phenomenological approach the body is (what makes us) conscious, then the incorporation of an external tool altering the habit body's constitutive structure that the external tool would also co-constitute consciousness. This claim would reject *b)*.

Consider the four conditions that according to Clark and Chalmers have to be met in order for an external feature to count as part of an extended cognitive system (Clark, 2008, p. 79):

(1) The device is a constant in the life of a cognitive system. It is "reliably *available*" (my italics) and "typically invoked".
(2) The further processing of information stored in the device is executed *automatically*.
(3) The retrieval of information is achieved *easily* and "when required".
(4) The use of the device requires that the cognitive system has (at some time in the past) *consciously endorsed* the information stored in it.

It is ironic that the authors were sure that the fourth requirement had only marginal relevance (for a criticism see Kyselo and Walter, 2009), for it seems to be the most important constraint of all: what we experience in acquiring a new habit, say, in learning to use a stick to substitute impaired vision, affects us directly. As a consequence, it would be impossible to judge that a device is 1) "reliably available", 2) executed automatically or 3) information stored on it easily and timely achieved without having thereby already relied on the previous conscious experience of the habit body and an experienced harmony or flow given through the fit of our intention and the resultant in its enactment. In that vein, constraints 1–3 can be seen as referring to different aspects of the process of learning and the nature of a habit, or skill, i.e. they are observations only *resulting from the conscious experience of learning or having learned a skill*. With growing "proficiency" dealing with the external object becomes flow-like, i.e. feels *automatic* (constraint 2). It is through my repetitive engagement with a particular object and the accompanying experience of a flow between the intention and what could be realized with the instrument that I perceive it now as *reliable* (constraint 1) and which I would invoke, for the exact reason that it enabled me to experience this successful contingency between what I intended and how I was able to realize this intention. To say furthermore, that the information is *"easily" acquired* (constraint 3) is just another way of expressing the *experience* of a flow that comes with having developed a habit.

This shows that in fact only by reference to conscious experience can we at all arrive at the four constraints. This is not only with regards to the process of tool incorporation but also with regards to actual use of the incorporated tool. While proponents of extended cognition deny that conscious experience has any relevance for extended cognition it seems to constitute the very basis of their argument.

Let me come back to the question of how to account for the body's role in BCI. I suggest that if our embodied existence can change by "appropriating fresh instruments", then BCI should count as an example of such a non-biological instrument (Merleau-Ponty, 2000/1945, p. 166). Having learned how to reliably produce distinct activation patterns to control an artificial limb or a cursor on a screen could be seen as having acquired a new habit. Crucially, BCIs do not extend a pre-existing mind. Just like the stick for the blind person they are "incorporate[d] into the bulk of our own body". They hence provide an additional structuring element that allows the patient to create a new domain of significance and thus continuously enact her conscious and directed existence in the world (ibid., p. 166).

A problem for this could be that Merleau-Ponty's perspective on tool-incorporation relies on movement. The stick of the blind person is an example of how an external device can become incorporated through bodily effort in terms of neuro-muscular movements. One might argue that there are cases of incorporation that do not require direct movements of the muscles, as for example in wearing glasses. Clearly, there is a difference between using a stick and wearing glasses

in that the latter does not involve displacing entire body parts (e.g. limbs). Nevertheless, there is still some sort of minimal muscular activity involved in the act of visually "aiming at" objects in the world. For this reason the incorporation does require – even if to lesser extent – muscular movement of the physiological body. The situation is however fundamentally different in cases of BCI-use. BCI-use does not comprise *any* muscular activity, nor does it involve the motor neurons. The following Figure 5.1 illustrates the realizing pathways in both variants of movements.

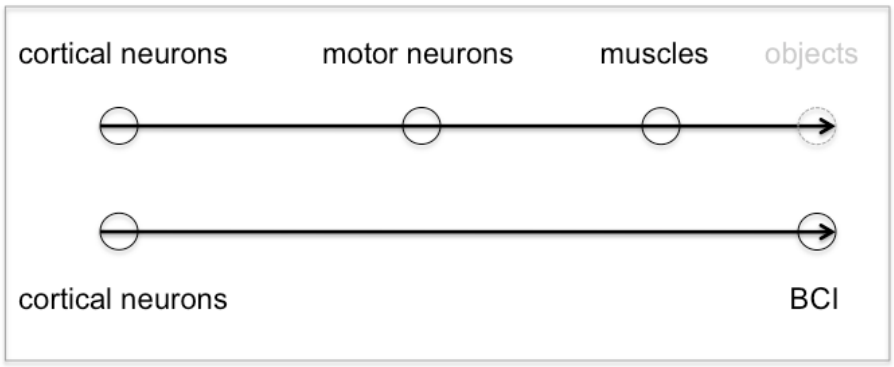


Figure 5.1: Comparing the pathways underlying "normal" and BCI-based movements. Whereas in the first case signals from the cortical neurons activate the motor neurons that then control muscular activity, in cases of BCI-use, the cortical neurons are directly connected to the BCI interface, additionally assuming the role of the motor neurons.

In neuro-muscular movement cortical signals excite the motor neurons, which then activate the muscles. But in case of BCI-use the cortical neurons are connected to the external device – they directly control the BCI.

If the appropriation of external tools was limited to neuro-muscular movement then we could not account for the incorporation of BCI. Recall that I suggested that directedness might involve action that is non-movement based, i.e. an act of relating originating "in" the individual (5.2.2.) If bodily action was not limited to movement then this might help to account for BCI incorporation: Even though BCI-use is not entirely realized by biological means and does not involve movement it can still realize an intentional bodily act of "aiming at things". In this way it can help create a new domain of significance.

If these considerations are plausible, they have consequences for our general conception of embodiment. It now seems that the body is not exhaustively described by reference to its biological underpinnings. This assumption is not based on a multirealizability constraint of functionalism (as in extended cognition), but grounded in the concept of being-in-the-world. The BCI becomes a part of our body in so far it has been incorporated into the habitual structure of our existence. Accordingly we can now say that BCI may under certain conditions (sufficient training, familiarity, matching of intentions and outcome etc.) realize *bodily movement*, where "bodily" also includes external objects (granted that they have been transformed into a structural component of enacting our existence) and non-movement based action. Furthermore, if bodily structures, such as the sense of directedness and the felt potential to be able to perceive a world depend on a continuous engagement with the world, then we can further specify the role of BCIs for the LIS patient. By ensuring the patient's embeddedness they can help to preserve these basic bodily structures. It is an important task to spell out in detail how the structure of the habitual body has changed with the incorporation of BCI. As the discussion of Clark and Chalmers' constraints suggests, the subjective perspective is thereby going to be of fundamental relevance. An assessment of whether and to what extent the BCI technology has become incorporated into the habit body cannot possibly be given from an external perspective. One would have to also

ask the patient herself to describe her experiences with using the technology in order to determine whether or not BCI has become a part of her embodiment. This could be done by e.g. using some sort of questionnaire, as recently done by Nizzi et al. (2012).

Contra the Developmental Response – The Body of Habit

The world-directed structure of the lived body is seen in the acquisition of a habit. To learn a habit is not a matter of intellectual understanding. When I learn, e.g. how to dance I do not organize "by analysis, the formula of the movement" (ibid., p. 165) but I understand the new movements in a prereflective, bodily manner, in a "grasping of a motor significance". This requires that I must already have had a prior bodily grasp on the world. To learn a habit is then basically to modify this prior grasp (ibid., p. 164). As Merleau-Ponty writes:

> If habit is neither a form of knowledge nor an involuntary action, what then is it? It is knowledge in the hands, which is forthcoming *only when bodily effort* is made, and cannot be formulated in detachment from that effort. (ibid., p. 166, my italics)

The habit body has a physiological grounding through "stable organs and pre-established [sensori-motor] circuits" (ibid., p. 100). But crucially, in order to preserve the *habitual* nature of our embodiment this physiology needs to be put to practice.

We have seen that according to the sensorimotor approach bodily engagement is only developmentally relevant (see Section 4.2.2). On the phenomenological account in contrast mind can never be decoupled from an actual engagement with the world (ibid., p. 167). We cannot separate in our analysis the role of sensorimotor circuits (i.e. the body) from that of consciousness (or cognition). The sensorimotor circuits are the very *means* by which we are directed to the world and have experiences of it. In so far as the bodily functions are affected, conscious experience and thus sensorimotor knowledge should be affected, too.

For that reason Merleau-Ponty's notion of embodiment cannot be interpreted in terms of a developmental response of the kind discussed in the previous chapter (see Section 4.2.4). Such response would presuppose that body, cognition and consciousness could be seen as separate things. But since for Merleau-Ponty being embodied implies in principle being a subjective body that engages with the world, the body cannot be seen as a historical cause of cognition or consciousness. The habit body implies that in encountering the world we constantly change our way of being in the world. To learn how to dance, to play an instrument or to write on a typewriter changes our habitual abilities. These habits augment previous capacities reduce or open up new spaces in which the subject can relate via her bodily experience to the world. Nevertheless they all rely on the fact that we already have a habit body, which was and is intentionally related to the world, where being intentionally related involves engaging with the world.

BCI and Communication – a Bodily, but not Movement-Based Action?

In the previous sections we have seen that movement of the body might not be limited to neuro-muscular, or to physiological structures in general, but can implicate non-biological objects. However, to say that incorporation does not require neuro-muscular movement can imply two different emphases. The first one, referred to in the previous section, focuses on the *kind* of realization of movement, i.e. whether or not bodily movement needs to be physiological. However, even non-physiologically realized movement can still share some crucial similarity with neuro-muscular movement. Using, for example, an artificial hand to manipulate external objects, to navigate a remote control or a wheel chair with a BCI also involves either a kind of displacement or the interaction with an external object.

There is, however, a case of BCI-use that does not seem to fall under the same category. In facilitated *communication*, the technology realizes an action without relying on muscles or (or not

only) involving a displacement or the manipulation of an external device. For these cases even a more liberal concept of bodily incorporation through movement seems unsatisfactory, for what kind of movement is involved in an act of communication? Recall, therefore, the second refined challenge: can a notion of embodiment refined in terms of phenomenology account for an action that is not based on movement of physiological or artificial body parts or objects? In what follows I attempt to provide a positive answer.

To begin with, let us compare the concrete structure of the action cycle in cases of (1) navigating external devices with (2) cases of communication (see Figure 5.2). Each action cycle comprises four basic structural components: *a)* the source of intention, i.e. the particular goal or wish that the patient aims to realize, *b)* the direction of the intention (towards herself, towards the world), *c)* the type of intentional object, and *d)*, derived from *(a-c),* a more general realization type of intentional action cycle.

Let us begin with (1) *BCI and the navigation of external devices*. Imagine, e.g., that the patient uses an artificial hand to adjust the cover sheets of her bed. The goal of moving the sheets could be to cover up a part of her body to keep her warmer (a). Her goal is directed at something *external* to her, something not belonging to her body (b). The intentional object of this bodily directedness is the cover sheet, i.e. a *(material) object* (c). Finally, the type of action that realizes her goal can be described as a kind of (non-biologically realized) *movement* of the artificial arm moving the sheet (manipulating an external object) (d). The same structure could apply to cases of controlling a web browser or a TV.

Let us similarly decompose (2) *communication with BCI*. Suppose the patient uses the BCI technology to compose the following phrase on the screen: "How are you today?" Let us further assume that the patient's sister is sitting next to her, and that she is able to read the display. The motivation behind the patient's question is to know whether her sister is fine (a). The patient is addressing her sister, and thus is, similar to the first case, directed at something *external* to her (b). Crucially however, and in contrast to (1) she is not directed towards some *thing* external to her, i.e. a material object, such as the sheet or the remote control, but to *another person*, who is a *subject* of her own (c). This leads to another major difference between these two cases. In (1) the dynamics of the action cycle can be described as *uni-directional*, the patient's action, involving the external material object has come "to an end", as soon as the manipulation of the object has been performed. Phenomenologically speaking, the action is successful if there is an experienced harmony, i.e. a match between the intended goal (become warmer) and the new position of the sheet (now covering the body part). The entire action complex can be seen as a linear, uni-directional process.

In cases of communication however, the action complex is not terminated by the patient's formulation of the question on the screen. In order to satisfy her goal of action (a), i.e. coming to know how her sister is doing, she requires an answer from her. The intentional "object" of the patient is not merely manipulatable but reacts based on her own. In order to close the intentional action cycle a contribution on part of the sister will be required.

BCI usage for	Object manipulation	Communication
goal of action	improve state of being	share meaning
direction	external	external/internal
intentional object	material **object**	another **subject**
dynamics	**uni**-directional	**bi**-directional
movement	yes	?

Figure 5.2: A comparison of BCI-use in object manipulation and communication. The case of communication resists a straightforward reference to movement, as this usually implied uni-directional interaction with objects and not a dynamical engagement with another subject.

Furthermore, in the communicative action the patient is not only directed towards the external world (her sister). By composing the sentence "How are you?" and expecting the sister to reply the patient is implicitly already directed towards herself. The case of BCI communication draws attention to the fact that in speaking of embeddedness in and directedness towards a world, we cannot solely consider the world of external objects. We also need to accommodate the fact that human beings are embedded in a world of other subjects. Other subjects do not remain independent of the actions involving them. They engage in joint interaction processes and co-construct a dynamics that is irreducible to their individual capacities. This participatory aspect of bodily action fundamentally differs from the interaction dynamics in "mere" object manipulation.

The phenomenological concept of embodiment based on Merleau-Ponty's notion of *habit body* is able to counter the challenge posed by LIS and BCI in that it:

> 1) helps to specify the link between bodily impairment and its relation to conscious experience: consciousness and body are intimately tied through existence, our being embedded in and thus directed towards the world;
> 2) can explain BCI-use for manipulating external objects as a case of incorporation of an external tool into the habit body, thus allowing that embodiment involves movements that are not necessarily biologically grounded.

Merleau-Ponty did explicitly acknowledge that we are "committed to a certain physical and inter-human world" (Merleau-Ponty, 2002/1945, p. 94). However the differences between actions that are directed towards objects and those that deal with other human beings invite further elaborations of embodiment in terms of the relation between the body and sociality (see Chapters 6 and 7). The confrontation of LIS and the habit body thus led to a further refinement of the challenge:

> 3) What is the role of the body in social interaction (communication)?

In the following I introduce another phenomenological approach to the living body. This is the concept of self-affection as put forward by phenomenologist Michel Henry.

5.3 The Phenomenological Concept of Embodiment II – Michel Henry

Like Merleau-Ponty, Henry criticized a reduced conception of the body as a material object i.e. the body to be seen, heard and touched like any other object in the world (*corps materiel*). Without first being a *sensing* body we cannot come to understand the body as *sensed*. The sensing body is not a given in the world but is the very source of giving the world. (Henry, 1995, mp. 5). This is in accordance with Merleau-Ponty's conception of the body as intentional in the sense that the experiencing body is directed at providing experiences from the outside world (including the body as a material object). However, Henry's objection to the Cartesian view of the body goes beyond Merleau-Ponty's approach.[63]

He questions precisely the claim that the living body is primarily intentional as being directed towards (having) experiences from the outside world. Henry rejects that whenever I experience something, I hear what can be heard, see what can be seen, that this is only situated in a space *outside of me*. According to him, this description misses an essential aspect, namely the relation that the sensing and understanding body has to itself (ibid.). There is a more fundamental, original sense of corporeality, namely *self-affection*, which is the immediate self-experience of the body itself (Henry, 1988, 1995, 2004).[64] Henry describes as it as radical subjectivity standing in relation with itself (Henry, 1988). Crucially the source of this affection lays within the body itself and not the world. As Zahavi explains:

> In [Henry's] view, subjectivity is absolute in the sense of being irrelative, and completely self-sufficient in its radical interiority. It is immanent in the sense that it manifests itself without ever leaving itself, without producing or presupposing any kind of fracture of alterity. (Zahavi, 2007, p. 136).

This self-affective relation to myself is not to be compared to the relation I can have to external things. It is the experience of my inner body that is not visible from the outside but that is the actual source of my ability to walk, grasp and let myself be with others (Henry, 2004, p. 206). I am the source of self-affection but not in that I notice myself as being that source. If a hand touches a body part then there is an intentional relation because the hand senses the body like it can sense an object in the world. The body part is that which can be seen by the eyes, heard by the ears, etc. However, the relation of the hand to itself differs from this. The hand as it affects itself is not directed at anything, it is non-intentional providing an immediate evidence of its power for movement or touch. Auto-affection reveals an "I can" that is an ultimate power (*pouvoir ultime*) because it possesses every potential that the body has (as for instance grasping, moving and touching). This power is thus much more fundamental then what Merleau-Ponty refers to with the "I can" of the habit body.

For Henry bodily self-affection is grounded in the phenomenology of life itself, which we understand not by looking at experience from the external world, but by "being in it". He says:

> Life is phenomenological in the sense that it constitutes phenomenality itself, i.e. it constitutes the givenness itself and also the original mode in which this phenomenalisation is phenomenalised. Life is not givenness but precisely the givenness of the givenness, it is auto-donation. Auto-donation of life means that what life is giving is itself, what life is sensing, is just itself. Life is not first sensing the world, its resistance, or pressure…it does not sense first what is sensed, the quality of things, or the things. Life is not affected by another thing than itself, it is not affected by a heterogeneousness of any kind, it is affected only by itself. In this sense I say that life is auto-affection. (Henry, 1995, mps. 10-11, my translation)

---

[63] I should say already that I disagree with Henry's radicalness in this point as we will see in the discussion of self-affection in the context of LIS in the following Section 5.4 as well as in the next chapter, Section 6.1.2.

[64] This is similar to what Damasio has described as core consciousness (Damasio, 2006).

It is a property of life itself to be phenomenological and for this reason it is the basis of (any other kind of) phenomenology. Life does initially not sense the outside world but is affected by nothing other than itself. Which is why, as Henry puts it, "life is self-affection" (Henry, 1978, p. 143). The living body is a life that affects itself. For that reason "experience of the body is that of a reality that I do not have, but *am"* (Azevedo, 2006, p. 324, original emphasis). Experiences like being tired, exhausted or even being depressed as well as more positive experiences such as the feeling of lightness, or being motivated or maybe irrationally happy, are examples of such basic auto-affective bodily experiences. We do not have these experiences "through the intervention of a (inner) senseorgan or an intentional act, but are immediately aware of it." (ibid., p. 227). They immediately relate to ourselves and they not only originate in us, but also remain within us, letting us *be* a particular mode of being in the world. Self-affective bodily experience shows how "subjectivity can be aware of *itself*" (Zahavi, 1999, ibid., p. 225). Self-affection thus gives us a direct sense of our own state. And this unmediated self-awareness of subjectivity itself conditions all other experience of the world.[65]

In the following section I discuss self-affection as a dimension of the living body in the context of LIS.

5.4 Self-Affection and Locked-in Syndrome

We have seen that a concept of embodiment that necessarily implies movements (whether or not they are physiologically realized) is too restrictive to account for cases of LIS patients who communicate using a BCI. Communication in BCI seems to involve a form of directed action originating in the body that does not require any movement at all. I have also suggested that if one conceives of the directed interaction with the world as comprising an "outer" action (realizing the intended goal) and a kind of "inner" action (the very act of relating to the world) then the latter seems to be a form of embodied action that is directed towards the world without involving movements. While LIS patients may not be able to realize parts of the action cycle, in communication with BCI they are still able to actively relate to the world and enact the relatedness in the form of voluntarily produced brain activation patterns selecting letters from a grid on a screen. We thus assumed that "interiority is understood as an act" (Barbaras, 2011, p. 92). With Henry's notion of self-affection in mind there now seems to be an additional type of non-movement based action available, which refines interior bodily activity. It is an action, which is (in contrast to the habit body) *not intentional* but referring to the subject's pre-personal existence as a living being that affects itself. If the immediate and non-intentional self-affection precedes arguably every other (directed) type of action then cases of BCI communication in LIS can be approached based on a tripartite structure of bodily action:

> a) a form of self-affective internal bodily action that is directed at herself. It does not require movement and it does not transcend the patient's organismic boundary
> b) a form of intentional bodily action that is directed towards the world. It is also non-movement based and internal.
> c) a form of intentional bodily action that is directed at the world. It is external, mediated by BCI and involves a bi-directional dynamics (the participation of another subject).

With this more liberal concept of bodily action we might also account for the body's role in cases of TLIS patients who do *not* use BCI technology. Even though TLIS patients may not perform

[65] Lavigne worries that even though Henry seems to acknowledge that there is a difference between self-affective (immanence) and hetero-affective experience (transcendence), this difference seems to dissolve because he says that world-directed experience is only revealed through self-affection. If everything is based on the experience of a self-affecting subject then no distinction to the reality of the external world can be drawn. This would amounts to subjective idealism (Lavigne, 2009).

any kind of *outwards*-reaching action they could be still relating to the world *and* importantly also to themselves. Both forms of relatedness could be seen as a form of bodily, non-movement based action. We have seen however, that embeddedness requires an update of our bodily capacities and that the question remains whether the world-directed "I can", as a kind of inner action can be preserved without further world-involving engagement on behalf of the patient.

One could maintain that this worry would not hold for the patient's relation *towards herself*. LIS obviously limits actions directed towards the world. But why should we assume that the patient would stop to self-affect herself? One could argue that since self-affection counts as an aspect of the lived body and to be alive is to be self-affected then for as long as she is alive the patient would still be bodily active.

This line of reasoning is closely related to the problem we encountered with regards to the habit body. It implies an acceptance of Henry's premise that self-affection is completely independent from any engagement with the world. What Henry thereby does not consider is how the body (or life) originally comes to be self-affective. It seems that he implicitly departs from the fully developed embodied phenomenology of an adult human being. From this perspective it is obviously possible to distinguish different aspects of embodied experience, such as the directedness to act from a feeling of being immediately and immanently affected by one's own bodily existence. But the fact that we can at some point in our development distinguish a component of embodiment that does not instantaneously involve intentionality from a subjective viewpoint does not mean that this must be constitutively, or generally, so. Starting from the phenomenological experience of an adult being avoids the question of how the different versions of embodied experience have developed.[66]

One could try to defend Henry's position as follows: if self-affection is necessarily coinciding with life itself, then it is not only as adult beings, but as also soon as we are born and alive that we are immediately self-affective. This would not involve any engagement with the world. However, even at birth we are not isolated beings, independent from touch, smell or other impressions from the environment. Prior to being born we are already deeply, indeed physiologically connected with our mother. We are literally embedded in the womb and thus always potentially affected by some kind of external world. If we accept Henry's claim of the absolute immanence of auto-affection, then this would imply accepting a conception of life itself as entailing a radical immanence and independence from the environment. However, from a more enactive angle life is never independent from the world in which it is alive and other lives that live in it. Henry's concept of self-affection clearly enriches our concept of embodiment: Bodily existence is not only related to the world but also to itself. However, it remains an open question whether or not there is a boundary, i.e. make a clear-cut separation, between self-affective states on the one hand and intentional bodily experiences on the other. I discuss this question in the following chapter.

5.5 Summary and Conclusion

Remember the one question lying at the heart of the challenge posed by LIS and BCI: to what extent is the body required for cognition and consciousness? Faced with that question we saw in Chapter 4 that both sensorimotor and extended cognition appeared to provide a deflating answer. While extended cognition considers the role of the body to be trivial, the sensorimotor approach runs the risk of limiting the relevance of bodily action to a mere causal factor in the development of the cognitive or to restrict it to biological structures.

---

[66] Even though Henry acknowledges the existence of a difference between self-affective (immanence) and hetero-affective experience (transcendence), this difference seems to dissolve when he says that world-directed experience is only revealed through self-affection. If everything is based on the experience of a self-affecting subject then there is no difference to the reality of a world. This amounts to subjective idealism according to which everything is a manifestation of subjectivity (Lavigne, 2009).

The phenomenological concept of the body, in contrast, offers a non-trivial and non-developmental answer to the question. It can account for the physiological restrictions in the sensorimotor approach while also providing a plausible explanation of how external tools can play a role in cognition in terms of incorporation into the *habit body*. It expands the concept of the subjective body by a second dimension of bodily experience, i.e. self-affection. Based on these we can add the following kinds of actions to the refinement of bodily action from the previous discussions in Chapter 4 (for the general overview of the refinements, see Figure 5.3).

1. Movement-based and artificially constituted: In contrast to the sensorimotor approach, the phenomenological concept of the *habit body* allows for non-physiological, external objects to become incorporated into the body. This can accommodate the use of BCI in LIS patients. BCI are seen as a bodily means allowing for actions that are based on movement and at the same time artificially constituted. In contrast to an extended cognition version of this claim, the phenomenological approach emphasizes that the incorporation of artificial tools relies on subjective experience.

2. Not movement based, internal, but directed towards the world: The "I can" and directedness towards the world can be seen as a form of covert bodily action that precedes or parallels external actions (whether or not they are movement-based). One could argue that while the LIS patient is unable to engage in neuromuscular movement-based actions she still continues being directed towards the world in this way.

3. Not movement-based, external, directed towards the world: One example of this type of action is communication with BCI. In communicating via a screen the patient performs a bodily action that is external and directed towards the world, yet does not imply any kind of movement.

4. Not movement-based[67]*, internal, passive:* This widens the action-concept so as to account for cases of LIS patients that may not be directed towards the external world, but are still bodily active in that they continue to passively self-affect themselves. However, since Henry assumed that self-affection is independent from the world, cases of TLIS and BCI communication pose a problem for the phenomenological concept of the body. Without knowing how self-affection is constituted, we are unable (pace Henry) to say whether or not self-affection involves any kind of interrelation with the external world, and thus to say whether and to what extent BCI plays a role for the LIS patient.

[67] Note that Henry considers self-affection to be a kind of movement, a self-movement of life (Henry, 1999, p. 352). This notion of movement differs from the definition adopted here according to which movement is an action transcending the organism's boundary and involving the displacement of body parts (physiological or artificially realized) and objects.

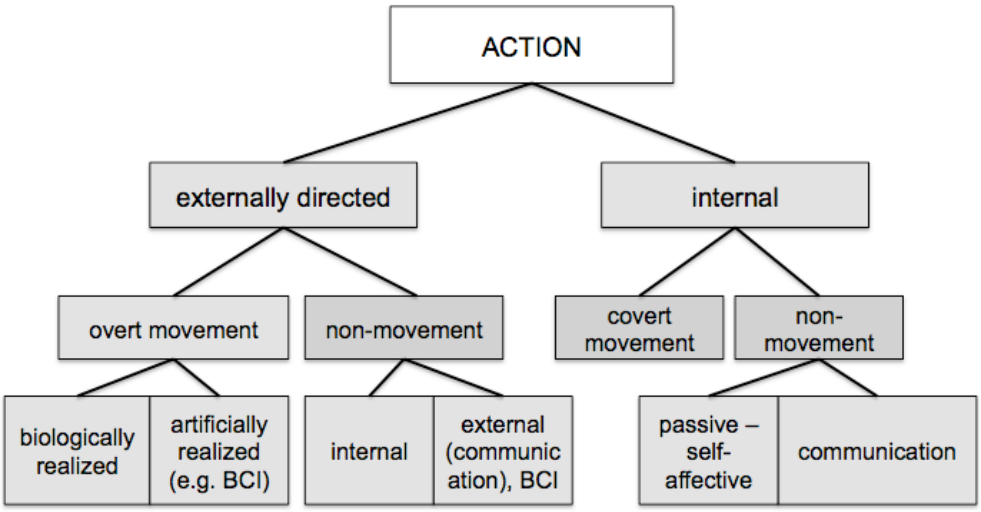


Figure 5.3: Varieties of bodily actions. The diagram provides an overview of the refinement of bodily action derived from the discussion of sensorimotor, functionalist and phenomenological approaches in the context of LIS and BCI. While bodily action is usually equated with visible movements of the biological body, this diagram adds two further dimensions. On the one hand, the concept of bodily action now comprises processes that do not transcend a cognitive system's organismic boundary (internal and externally directed types of actions). On the other hand, we find that actions can be non-movement-based, as is the case with some use of BCI.

These refinements make the phenomenological approach to the body a strong candidate for assessing the role of embodiment in cognition and consciousness in LIS. However, there appear to be two further problems. Firstly, it seems that there is a tension between two dimensions of the living body, namely the living body in relation to the world and in relation to itself. From a phenomenological perspective it remains unclear how hetero- and self-affective bodily experiences are interrelated. I suggest that this issue could actually be part of the body-body problem, i.e. the question how material and subjective body interrelate (Hanna und Thompson, 2003, Thompson, 2007, see Section 2.3.1). To resolve the body-body problem successfully also involves understanding how different dimensions of bodily subjectivity (self- and hetero-affection) are related.

Secondly, the phenomenological body concept discussed in this context reaches conceptual limitations when confronted with BCI-use for communication in CLIS and TLIS patients. We have accounted for the role of subjectivity in the patients themselves by linking subjectivity with a continuous embodied engagement. We did however not explicate the role of subjectivity in the patients' relation with other people. To communicate with other people is a bodily activity that involves not just objects, but other subjects.

By confronting the phenomenological body with LIS patients who use BCI we thus found two additional refinements of the challenge:

1. How can we explain the relationship between bodily experience that is directed towards the world and bodily experience that is self-affective?
2. How can we account for the role of embodiment in interactions with a social environment?

These questions are addressed in the following chapter where I discuss and formulate an enactive approach to embodiment, which is the last approach to embodiment discussed in this thesis. I show that the enactive approach can incorporate aspects of the phenomenological body-concept while at the same time accommodating some of its shortcomings.

Chapter 6.

LIS and BCI – a Challenge to Embodiment, III:

The Enactive Approach

> It is not enough to say that the mind is embodied; one must say how. (Edelman, 1992, p. 15)

Throughout the last two chapters I have discussed three approaches to embodiment: the sensorimotor, the functionalist and the phenomenological. At every step of the discussion the confrontation with LIS and BCI appeared to reveal particular conceptual limitations, which led to a refinement of the originally *crude challenge* (see Section 4.1.1). It lead thus to a stepwise elaboration of the body concept and what bodily action is (for an overview of the refined challenges see Appendix). In concluding the previous chapter I suggested that the phenomenological approach helps to account for the shortcomings of the sensorimotor and functionalist approach, such as the under-determined role of continuity of bodily engagement to sustain habits and the neglected role of consciousness in the incorporation of non-biological objects into the body. However, the phenomenological approach does not offer an organizational explanation from which the constitution and interrelation of the two dimensions of bodily subjectivity could be assessed. It does neither satisfactorily account for the body's role in social interaction. I suggest that these limitations can be accommodated by adopting an enactive approach to the body, to which I turn to in this chapter.

First, I discuss how the enactive approach can integrate aspects of the previously discussed perspectives on the body (the sensorimotor, functionalist and phenomenological). While these approaches might be seen as rather independent accounts of the body concept, the enactive approach respectively implies aspects of each of them and offers a way to relate them in a holistic way. We have seen that the enactive framework to cognition is an integrated approach to cognitive science, grounded in the concepts of autonomy and sense-making. We have also learned that, methodologically speaking, enactivists contend the mutuality of first- and third-person explanations. These two general properties, the integrating attitude and the mutuality of first- and third-person description, provide a theoretical toolkit which allows to synthesizing tenets and explanatory contributions of the three previously discussed proposals that so far appeared incompatible.

Second, I provide an elaborated outline of an enactive body concept, which can account for the refined challenges (the issues of physiology, movement, inner-outer bodily actions, the body-body relation) and offer a response to the first part of the general conceptual challenge to cognitive science (the clarification of concepts, i.e., of the body in the present case, see Chapter 3).

Third, I confront this elaborated concept of the enactive body with cases of LIS and BCI. It accounts for the first of the two challenges to embodied cognitive science (a clarification of the body concept), but similarly to the phenomenological approach is does not yet address the role of embodiment in social interactions. For that reason is unable to determine in the context of LIS and BCI what the body's role in social interaction is. I thus finally suggest that we should transcend considerations in terms of *individual* embodiment and move towards a more encompassing view, which approaches embodiment from a socially informed perspective (see Chapter 7).

6.1 The Enactive Body and Phenomenology

Both enactivism and phenomenology share basic assumptions that allow to setting some common ground for their concepts of embodiment. This is clearly the case with respect to their emphasis on the role of subjective experience. Both accounts hold that we cannot consider the body without at the same time accounting for the first-person perspective, i.e. the phenomenological dimension. Consciousness and body are inextricably tied in Merleau-Ponty's as well as in Henry's notion of embodiment on the one hand, and in the concept of autonomy and sense-making in enactivism on the other. Recall from Chapter 2, that by being and sustaining an identity, a cognitive system generates a centre of activity (autonomy). It thereby evolves a specific, subjective viewpoint on the world, from which encounters with the environment are evaluated as meaningful (sense-making). The conscious body can be seen as constituent and purposive of that centre of activity; it is the condition of possibility for the cognitive system to exist. A further general similarity between Merleau-Ponty's *habit body* and the enactive perspective is that the living body cannot be treated separately from the environment, i.e. its interaction with it. My suggestion is that the sense-making activities of the autonomous system can be referred to as subjective experiences of hetero-affection (experiences of the world), whereas self-affection can refer to the self-awareness of the autonomous system. Using the concept of operational closure as an example I indicate in the following how enactivism can integrate these two aspects of phenomenology within its underlying background epistemology.

The relation between enactivism and these phenomenological approaches, in particular of their concepts of embodiment, is a new research question. It would be a project in its own right to fully account for their linkages. Here I can only begin to argue for conceptual parallels between the two views and illustrate by way of example how the enactive notion of embodiment can be informed by insights from phenomenology and vice versa.

6.1.1 Habit Body and Autonomy

According to enactivism, cognition depends on the interaction dynamics of the autonomous system with its environment, where the relevance of aspects of the environment depends on the norms and goals of the autonomous system itself (see 2.2.2). An autonomous system is always embedded in an environment, and what this particular environment is, i.e. "the world" for the system, depends on the structure of the autonomous system – both mutually shape each other.

Reconsider the definition of autonomy from Chapter 2: an autonomous system is a "system composed of several processes that actively generate and sustain an identity under precarious circumstances" (De Jaegher and Di Paolo, 2008, p. 35). It turns out that similar ideas hold for Merleau-Ponty: we are already and always embedded in the environment, and we experience this fact through the body in movement. The environment is not objectively given, but becomes known through experiences of the enactment of our bodies. To clarify this point, we can consider von Uexküll's distinction between world and environment. According to Uexküll, what counts as the environment of a particular system is not the same as the "world seen simply through the lenses of physics and chemistry" (Thompson, 2007, p. 153). What an environment is depends on the animals' physiological structure and action capacities. Von Uexküll illustrates this for the tick. The tick can perceive butanoic acid and it can distinguish warm and cold. It starts acting – falling down, biting and sucking blood – only if it encounters an endothermic animal that produces traceable butanoic acid. Depending on its bodily structure and what it can perceive and do, the tick evaluates only particular aspects of the world as meaningful and what is meaningful becomes an environment. The rest of the world is simply not part of its environment (Uexküll, 1958).

We find a similar perspective on the relation between the body and the environment in Merleau-Ponty's approach. In describing the *habit body* he writes:

> If man is not to be embedded in the matrix of that syncretic setting in which animals lead their lives in a sort of *ek-stase*, if he is to be aware of a world as the common reason for all settings and the theatre of all patterns of behaviour, then between himself and what elicits his action a distance must be set, and as Malebranche put it, forms of stimulation from outside must henceforth impinge on him 'respectfully'; each momentary situation must cease to be, for him, the totality of being, each particular response must no longer fill his whole field of action. Furthermore, the elaboration of these responses, instead of occurring at the centre of his existence, must take place on the periphery and finally the responses themselves must no longer demand that on each occasion some special position be taken up, but they must be outlined once and for all in their generality. Thus it is by giving up part of his spontaneity, by becoming involved in the world through stable organs and pre-established circuits that man can acquire the mental and practical space which will theoretically free him from his environment and allow him to *see* it (Merleau-Ponty, 2002/1945, p. 100, my underlining)

By producing and sustaining an identity the organism thus brings about a physiological boundary between itself and other elements of the domain in which it is embedded. At least at the level of the single organism this seems pretty much in accordance with the enactive concept of (biological) autonomy. It is expressed in the above phrase "between himself and what elicits his action a distance must be set". To become aware of the world, the autonomous system needs to be distinguished from it. This is possible through the physiological structures of the organism (in Merleau-Ponty referred to as "stable organs and pre-established circuits") and the *operational closure* of the cognitive system (see 2.2.1)[68]. Through operational closure, the system establishes a border between itself and system-external processes that thereby become extrinsic and thus form the system's environment. This means that the system's (biological) autonomy can be sustained only if the interactions with the environment do not violate its organizational principle. A similar relation seems to hold in Merleau-Ponty's concept of the habit body when he says: "stimulation from outside must henceforth impinge on him 'respectfully'; each momentary situation must cease to be, for him, the totality of being". While the body clearly is interacting with the environment, these interactions do not affect it as whole, which in enactive terms is to say that they perturbe but do not determine it.

Furthermore Merleau-Ponty suggests that the physiological body "will theoretically free him from his environment and allow him to *see* it". Our body is the very means by which we can exist in the environment and which therefore allows us to perceive it. This reveals the aforementioned characterization of the body as intentional and directed at the world. This directedness towards the world, i.e. the body's capacity to provide experiences of it, can be associated with normativity and the sense-making activity of the autonomous system: having a bodily identity allows the autonomous system to create a perspective on the world, and to evaluate its interaction with it in a way that respects the goals and norms given to it by itself.

6.1.2 Operational Closure and the Relation of Self-Affection and Hetero-Affection

The enactive concept of autonomy, and in particular the notion of operational closure, can be linked to Henry's notion of self-affection. Henry urged us to acknowledge that prior to being directed at the world, humans are passively, non-intentionally and immanently related to themselves. He writes:

[68] Though we will later see that in more complex forms of autonomy this boundary is not longer constituted by the organism, but is also relying on its interactions with the material as well as social environment.

> [L]ife is an immanent process, the eternal process in which life comes to itself; life pulls itself in against itself and plays with itself, producing its own essence insofar as this essence consists in and is complete in this self-enjoyment. Thus the process in which life reveals itself to itself is identical with the process of its generation insofar as this is understood as self- generation. (Henry, 1999, p. 352)

This description of self-generation of life appears to have a conceptual parallel to the concept of the autonomous system as a self-generated and self-maintained system. In order to maintain its identity, i.e. that which makes it distinct from the environment, an operationally closed system is involved in a constant process of regeneration and of becoming itself. The system is thus in a constant dynamical relation to itself. Through monitoring itself in the adaptive regulation of its sense-making the system undergoes an ongoing self-change (Section 2.2.3). Since the system strives to maintain its identity as a system, this constant self-monitoring relation to itself is something that it cannot avoid. It can be seen as the basis for the system's awareness of itself. Here lies a possible avenue for exploring a linkage to Henry's idea that life is self-affection. The autonomous system is aware of itself, not in the sense of positing itself as an object (a "what") but rather in the sense of monitoring *how* it fares in the process of constant self-generation. In order to be an adaptive system (to gradually evaluate interactions with the environment) the autonomous systems requires a subjective viewpoint. If we add Henry's concept of self-affection, then a subjective viewpoint would not merely be an experienced perspective on the environment but also include an immediate bodily self-awareness. Self-affection could thus be conceptually linked to enactivism in that it offers a phenomenological description of an autonomous system as it experiences itself as a self-maintaining, self-generating and operationally closed system.

Let me now take the opposite perspective and consider the phenomenological perspective from the viewpoint of enactivism. This can help to address the apparent tension between Henry's self-affection and Merleau-Ponty's concept of hetero-affection. It might shed light on a particular aspect of the body-body problem, namely to determine not only the relation of material and living body, but also elaborate on what a living body actually is.

Following Henry, self-affection is the experience of self-construction that allows the system to experience objects as external and different from itself. It thus seems that logically the capacity for self-affection must be primary (Zahavi, 1999). Yet Merleau-Ponty argues that it is only through the engagement with the environment that the body can have experiences (see Section 5.1). This tension can be resolved by looking at a formally similar problem in the concept of autonomy, i.e. the relation between self-construction (identity-constitution) and interaction with the environment. Presuming a primacy of self-affection would in enactive terms mean that in order for a system to interact with the environment it first needs to be a systemic identity.

This entails however a very limited notion of operational closure. Recall that autonomy should not be confused with autarchy. In in order to construct and maintain a bodily identity the autonomous system also requires resources from the environment. This dependence on the environment was referred to by the concept of structural coupling (see Section 2.2.1). However, there is also a more sophisticated level of autonomy at which identity no longer coincides with the material boundaries of the organismic whole. At this level interactions with environment matter not only for providing energy to fuel metabolism but also in the sense that behavioral processes, i.e. interactional regulation processes can contribute to the network that make up the system's identity as a whole. The concepts of operational closure and interaction here thus reveal a much deeper interrelation. In order to exist as an autonomous identity the system is also engaged in interactions with the environment that transcend the boundaries of the organism, i.e. that are increasingly more mediated (see Section 2.2.4). Accordingly, the view that operational closure comes before any interaction with the environment or that it is even independent from it would seem to oversimplify the interrelation between the two notions. At a higher level of description the system's constant relation to itself should not be considered as independent from particular interactional processes and vice versa. Both shape each other: self-awareness involves

some engagement with the environment and the engagement with the environment implies that the system is aware of itself, i.e. monitors itself. Translating this back to the phenomenological perspective, it now seems less obvious that self-affection must necessarily precede or has to be entirely independent from hetero-affection. Adding the phenomenological approach to the body to the idea of operational closure, we can make new sense of the interrelation of the notions of self-affection and hetero-affection: From an organizational point of view operational closure constitutes a distinction between a cognitive system and its environment. It is the basis for the system's awareness of itself in the constant struggle to keep this distinction. However this distinction is not merely established by the organism itself, but also in interactive engagement with the environment.

Consider Zahavi's recent interpretation of Henry's body concept:

> Henry insists that a phenomenological clarification of the ontological status of the body must take its point of departure in our original non-objectifying body-consciousness (Henry 1965, 79). When I am conscious of my bodily movements and sensibility, then I am conscious of it by virtue of the body itself; more precisely, by virtue of the very self-affection of bodily life (Zahavi, 2007, p. 145).

Following the present considerations, 'bodily life' in this quote should then entail not only the body as organism, but also the body as interacting with the environment. Self-affection and hetero-affection can be seen as referring to two subjective perspectives, reflecting either the system's relation towards itself or towards the world. These perspectives are not independent but rather deeply linked or co-determined in a complementary fashion. In this way, self-affection can, e.g., clarify an aspect of hetero-affection that Merleau-Ponty describes as setting a distance between the physiological body and the environment, which allows him to see it (Merleau-Ponty, 2002/1945, p. 100). The fact that I have experiences of the environment requires also that I am distinct from the world and that I am thus related to myself. But hetero-affection also impacts on self-affection: in order to experience myself as distinct I need to have experiences of my engagement with the world. Our body, as an operationally closed system, is distinct from the world, and this distinction creates a perspective on the world enabling experiences of objects as independent and external to us. This distinction established by operational closure results in a subjective perspective of the opposite direction, it should influence the relation that we maintain to ourselves (self-affection).

In light of this, we can now see that to describe self-affection as entirely independent from hetero-affection is actually adopting only a partial and momentary observational stance. It is the subjective perspective of an embodied subject that *now* monitors herself and that from this momentary perspective can obviously have an experience of herself as being clearly independent from the world. [69] From an organizational perspective both dimensions of bodily subjectivity are however linked in an active world-directed existence.

6.2 The Enactive Body and the Sensorimotor Approach

In the previous section I have offered an example of how the enactive concept of the body can be informed by the phenomenological approach. By grounding hetero- and self-affection in autonomy, we can gain a new explanatory level at which we can begin to intelligibly address their interrelations. Something similar could be done with the sensorimotor approach. In the context of LIS and BCI we have seen that the problem with this approach is not its central tenet that cognition relies on sensorimotor interactions, but rather that the notion of sensorimotor interaction remains as it were restrictive and under-determined. If bodily action is solely explained in terms of a kind of sensorimotor interaction that is based on movement and which is (at least arguably) biologically realized then there would be no possibility of explaining the use of

[69] I am going to elaborate on the same idea from a social perspective in the following chapter.

BCI in cases of global paralysis in terms of sensorimotor contingencies. It seemed similarly that the original conception of the sensorimotor approach underdetermined the role of sensorimotor interactions themselves, in ascribing to them only a historical role. However, grounding the concept of the body in autonomy may account for these issues, as I indicate now.

Let me begin with the latter – the limitation of sensorimotor interaction in regards to a developmental factor. From the enactive perspective, an explanation in terms of sensorimotor interactions appears to involve an organizational perspective on how the body as a means of being in the world can be constituted. Roughly speaking, sensorimotor interactions ensure both that the cognitive system can maintain its identity as a system (by enabling it to feed itself on resources required to be alive) and that it can engage in sense-making activities.

Now if sensorimotor interactions are linked to the maintenance of a system's identity, and to its sense-making engagements with the world, then they cannot be considered as only developmentally relevant because this would imply a rupture between these bodily sensorimotor abilities and the world and thus also the system's conscious experience (see Chapter 4).

In enactivism, cognitive systems are not cut-off from the environment, but emerge as a result of dynamical and mutual interaction with it. If the body is a necessary condition for a cognitive system to realize this mutual shaping, then as long as the cognitive system exists as a cognitive system it will have to be bodily engaged in the world. This does not mean that the cognitive system has to engage in sensorimotor interaction constantly at every moment of its existence. But it requires at least, as suggested in the second to last chapter, some sort of continuity and update (see Section 4.2.2). Sensorimotor action cannot be merely a historical factor for some cognitive capacities of the cognitive system but is an existential precondition for the system to be cognitive at all. On the other hand, we have learned that enactive autonomy is necessarily linked to subjectivity (Section 2.2.5). Integrating sensorimotor interactions with the enactive concept of embodiment would thus mean that these interactions would be intrinsically related to the subjective perspective of the cognitive system itself.

These considerations are reflected by Thompson's analysis of the relation between enactivism and the sensorimotor approach: the sensorimotor approach considers subjective experience in phenomenal object perception but not subjectivity in general. This is similar to saying that hetero-affection is only concerned with the body being directed towards the world. World-related bodily experiences are only a subset of subjectivity; they have to be grounded in a more general understanding of subjectivity as a basic property of the embodied system.[70] By grounding the phenomenological concepts of hetero- and self-affection in autonomy (and operational closure) we could thus also begin to address Thompson's task and determine the role of sensorimotor interaction for embodied subjectivity itself.

The second limitation of the sensorimotor approach, the restriction to biological structures, similarly applies to the extended functionalist approach. I thus address it as part of the following discussion of the relation between the functionalist and the enactive approach to embodiment.

6.3 The Enactive Body and Extended Cognition

Enactivism and extended cognition can be equally delineated from a strong sensorimotor perspective as both reject the claim that cognition is necessarily dependent on the physiological structures of the body and allow extra-organismic tools to become co-constituents of cognition.

In extended cognition the concept of the body is not properly defined, but only accounted for by its functional and contingent role within a greater information processing system. This, I have argued, neither explains what the body is, nor how it may be related to BCI

[70] This is subjectivity as "creature consciousness", i.e. the very capacity of living systems to be conscious, as Thompson and Cosmelli (in press) have recently put it.

technology (see Section 4.3). Yet to defend a kind of "extra-organismic liberalism" does not have to necessarily lead to a deflationary conception of the body. A trivialization of the body can be avoided by grounding the concept in enactive terms and conceiving of the body as that which makes an autonomous system a centre of identity and action. The body as an autonomous (cognitive) system does not realize a functional role in order to resolve a cognitive task. It is itself cognitive insofar as it contributes to the maintenance of the system's identity and its sense-making interactions with the world. Whether or not non-organismic processes can be treated as part of the bodily structures system then depends on whether they contribute to this task of self-maintenance or whether they enable sense-making with the world. An elaboration on the contribution of external objects would furthermore involve a consideration of subjectivity. This is because tool incorporation, at least from a Merleau-Pontian perspective, is based on a habitual acquaintance which involves conscious experience. If the system does not experience a fit between intended use of the tool and the aim of its use no incorporation into bodily structures can be achieved (see Section 5.2.3).

The claim that the body is a means for self-maintenance of autonomous identity and its sense-making interactions bears similarity to the claim of extended cognition that the body functionally serves to solve a particular problem. In order to understand the difference to this approach we have to consider the concept of autonomy as epistemologically primary. The engagements of the cognitive system with the world depend directly on the system's autonomy and a continuous drive to sustain this autonomy. The intrinsic goal for self-preservation is thus itself grounded in the body. In the same time the body is the very basis from which aspects of the external world are seen as relevant or not with respect to this goal. Without an embodied intrinsic normativity a system would not relate any aspect of the world to itself and consider it as "problematic". Despite the fact that the enactive approach thus transcends a strong commitment to physiological structures it does not assume a stance of "anything goes". It sidesteps the danger of trivializing the body concept: the body is primarily the organism, but it can comprise non-physiological elements if they contribute to the system's general ability to be a centre of activity in the world and themselves begin to be conserved as part its autonomy[71]. A body is not contingent, but a necessary requirement for realizing the very identity of a cognitive system and also for appropriating external instruments. Since it is through our body that we can become affected by the world, the body is also the purpose of cognition itself.[72] It is the reason for why anything can matter to us and for why we act at all.

There is a sense in which the extended functionalist and the sensorimotor approach are in agreement. In objecting that the sensorimotor approach is too restrictive, extended cognition does not seem to question that *if* we were to speak of 'bodily structures' then these would be, indeed, biological and movement-based. This misses that bodily structures might not be biological or movement based and thus overlooks the possibility that the concept of the body may require an elaboration beyond physiologically and movement-based sensorimotor contingencies. The extended functionalist criticism of the sensorimotor approach has thrown the baby out with the bathwater – rejecting that cognition requires particular *biological* bodily structures does not have to imply that bodily structures are irrelevant *per se*.

---

[71] This implies that in order to count as a part of the body something must be more than a tool. We can also interact with objects as tools that are not fully incorporated. According to Heidegger's circumspective account of tool-use, for instance, a tool always and already refers to what it used for ("um zu") and also the action that is involved in its use (Heidegger, 2001/1927, pp. 66–71). A hammer is not understood theoretically but practically and already evokes the action of hammering and the purposive of that action (e.g., building something or attaching a picture to the wall). Such tool-use refers to actions that are not of immediate concern for the person herself, they are used to build something external. In contrast, in the present approach incorporation of tool means that that the tool is helping to conserve a (new) way of autonomy.

[72] I argue later that the notion of autonomy is not exhaustively described in terms of embodiment if embodiment does not also entail a social dimension. Another "means" of ensuring our existence in this world is provided by in social interaction dynamics.

However, despite the extended functionalist's de-emphasis on the body, the precondition of something close to an autonomous body seems actually at work in its explanation of how external tools can become part of the "cognitive machinery". In the previous chapter I objected in contrast to Clark and Chalmers' marginalization of the role of conscious experience in tool incorporation that it is only because our habit body provides us with the experience of harmony and flow in the engagement with the tool that a tool can become a part of the cognitive system (Section 5.2.3). The obvious implication of this is that conscious experience of a match of intention and the use of a tool involves the body. This has a consequence for the present context: by underrating the role of conscious experience in incorporating external tools, extended cognition has not only underrated the role of the body. It is at the same time constitutively relying on it. Without a conscious body there could not be an extension of cognition because no external tools could be appropriated as means to drive cognition. From the enactive stance, they only drive cognition in so far as they enable us to realize our embeddedness in the world, which is exactly what a body does. Incorporated into the habitual structure of our embodiment, external tools sustain ways of relating ourselves to the world. The claim that certain external tools are a part of the cognitive mechanism appears therefore entirely unintelligible when it is not also related to a non-trivial notion of embodiment.

6.4 The Enactive Body – An Overview

In this section I summarize the above discussions and suggest four basic steps towards an enactive concept of the body. This concept is a synthesis based on the enactivist background epistemology, expanded by the incorporation of aspects of the phenomenological approach and further informed by aspects of the sensorimotor and functionalistic approaches.

First, since in enactivism the cognitive system *is* the autonomous system, every cognitive system is therefore necessarily a body, or in other words: a living body is a cognitive agent. The enactive body enables the maintenance of its autonomy and sense-making. The body is the very possibility of a cognitive system's existence in the world, using itself as well as external objects as media for it. On one hand the enactive body enables the autonomous system to create and maintain an identity and an operational closed centre of existence. But on the other it is also defined by what it actually allows the cognitive system to do.[73] Cognitive systems are embedded into a world and they are intrinsically involved with it. They are not immobile or merely reactive to the environment but navigate through it and interact with it in a flexible manner. A body is a body if it enables us to realize such flexible being in the world. This implies that a body that is not also engaging with the world, whether by moving itself, or interacting with objects, would cease to be the means by which we are in the world. In order to maintain bodily capacities they thus need to be continuously put to practice. The enactive body is thus constitutively a *body-in-action*.

Second, according to enactivism, the body is not only the means of being an autonomous system but also its purpose. A cognitive system strives to maintain its identity and engages in sense-making interactions with the world that are evaluated with respect to this basic goal. This striving is inscribed in the body. On the one hand only the system's embodied existence causes the cognitive system to strive to evaluate interactions with the world so as to ensure that existence. On the other hand the only way of doing this is by being an active embodied existence.

Third, in accordance with the enactive presumption of mutuality of first- and third-person perspective, the enactive body is approached from two different, yet inextricably linked, angles. It is considered from an organizational point of view, allowing a determination of the processes, components and their relations that contribute to it. This is the body as a living system. However it is also considered from the subjective perspective of that system which just

[73] A similar claim had been made by Uexküll (1958).

has, or better, *is* that body, the body as subject. This is the body as lived system. Bodily subjectivity could imply two basic kinds of experiences: experiences of the external world (hetero-affection) and experiences of itself (self-affection). Since the body is both a centre of identity and of interaction, self-affection and hetero-affection should not be seen as independent. They are both means of "telling" the autonomous system how it fares with regards to its goal to maintain an identity and to evaluate sense-making interactions accordingly.

Fourth, experiences of relatively stable and repeated patterns of interactions with the world allow organismic-transcending processes to co-contribute to the constitution of the interactive body. This is achieved through incorporation of non-physiological elements. External tools are incorporated to the extent that they contribute to the body's organizational structure as a means of interaction with the world and at the same time alter the normative value of these interactions (e.g., by helping create new goals). This incorporation requires a habitual (experienced) acquaintance with the tool.[74]

A basic implication of the enactive perspective on embodiment is that the body is the primary means by which cognitive systems can be cognitive systems. While neuronal-based processes can count as bodily processes, they would not serve this purpose on their own as they do not provide the distinction and the operational closure needed in order to constitute a center of activity and to be able to relate to the outside world. From an enactive perspective, there is no other primary source, which could be driving cognition than the body as a whole (comprising identity and interaction autonomy). It is not that the body is thus 'envatting' a controlling brain as the actual driver of cognition (Fuchs, 2011, Thompson and Cosmelli, in press). The brain "is certainly a central organ of the living being, but it is only an *organ* of the mind, not its seat" (Fuchs, 2011, p. 197)[75]; it rather mediates this bodily autonomy. As a consequence of this, there is no clear-cut distinction between neuronal and bodily in so far as neuronal processes can be a part of covert bodily action. Since in bodily autonomy identity and interaction are inextricably linked, both neuronal and bodily processes are not separable from the environment. This also explains why, in contrast to the sensorimotor approach, the enactive body concept does not assume a developmental response. Covert (inner) bodily processes are not entirely independent from overt processes because they require some sort of active update through active engagement with the environment.

Before I discuss the enactive approach to the body in the contexst of LIS and BCI let me clarify the relations that hold between the enactive body and the environment as well as the different kinds of bodily action that the enactive body comprises. Based on an autonomous perspective on the body we can determine three types of relations between the enactive body and the environment:

1) an independence from the environment with regards to the body's operational closure
2) a dependence on the environment with regards to resources needed to sustain its (organismic) identity
3) a dependence on action with regards to the body's ability to being an interacting body and to appropriate external instruments (practice feeds practice)

The three are closely interrelated. In order to provide operational closure under precarious conditions the system needs resources from the environment (2). Operational closure then allows for a particular distinction and thereby for an organizational independence from the

[74] Repeated active engagement can also become a habit that directly impacts on the physiological structure of the body. A body-builder's physiology, for instance, changes through constant training. The body builder may increase the amount of muscles and thus her body weight. This can change ways in which the body is realizing the athlete as an organismic individual and centre of activity.

[75] Though as I will argue shortly to conceive of the cognitive system in terms of (even an enactive form of) embodiment is still restrictive and turns this assertion to a subject of open discussion.

environment: it can now engage with it, have a perspective on it, without being affected (existentially) as a systemic whole (1). While the body thus needs to act in order to be a reliable means of being in the world (3), this action does not affect its core, being a center of identity. One might thus say that the autonomous body is *independent in dependence*.[76]

Let me now reconsider the different kinds of bodily action implied in the enactive perspective on embodiment. If the body is genuinely the means of being in the world and comprises the *body as identity* (autonomy with regards to identity) as well as the *body as interactive* (autonomy with regards to sense-making) then the term "bodily action" should comprise both, actions associated with the body as identity and actions associated with the body in interaction. These actions can be either overt (externally observable) or covert, or both at the same time.

The distinction between overt and covert is not parallel to the distinction of the body as identity and the body as interactive. It is a distinction between inner and outer processes, whereby inner/covert refers not only to neurobiological processes (as suggested by Hanna and Maiese, see 4.2.3) but also to bodily processes taking place *within* the organism. These are processes that are not (at least not without additional measurement and methods) accessible from the outside. Overt or outer processes can then refer to either the behaviorally visible action of the organismic body itself *or* to processes of the organismic body *in interaction with other objects* (incorporation of tools). As a consequence, the enactive body comprises four types of bodily action:

1) covert (inner) bodily action, associated with the body in interaction
2) covert (inner) bodily action, associated with the body as identity
3) overt bodily action, associated with the body as interaction
4) overt bodily actions of the body as identity[77].

6.5 The Enactive Body and LIS

Recall the crude version of the challenge posed by LIS to embodied cognitive science: How can there be embodied cognition in LIS patients given that they cannot move at all? In a first step we refined this challenge by asking more specifically what people understand when they say "the body" (throughout Chapters 4 and 5). In concluding Chapter 4, I claimed that the sensorimotor approach to cognition is too restrictive and the functionalist approach too loose to fully account for the impact of LIS and the use of BCI. In Chapter 5, I suggested that a better approach is the phenomenological one; yet since it is restricted to mere descriptions and does not offer an organizational or mechanistic explanation of the body – it is, while enriching, not sufficient to account for LIS and BCI.

In this section I apply the enactive body concept to the context of LIS and BCI. I show how the phenomenologically refined conception of the body can explain different kinds of bodily action in LIS and that it thus accommodates almost all of the challenges outlined and refined throughout the previous chapters.

[76] These considerations are mirrored by Jonas (1966) who describes the nature of the living as dialectical, possessing an "organic freedom" which in the same time has to be "...balanced by a correlative necessity" (p. 83). The metabolism is able to sustain the organism as a living system (by exchanging matter) but in the same time it also is required to do so and it can realize this exchange only by relying on resources from the environment.

[77] An example for this could be found in various yoga practices. In the dynamic change of different yoga positions (asanas) the body is active in movement but the purpose of this bodily activity is not instrumental and directed towards the world. Instead it supposedly affects the person's identity by improving the sense of self (Boudette, 2006).

6.5.1 The Enactive Body and Classical LIS (CLIS)

From the enactive perspective, the body is an autonomous center of existence, ensuring both the preservation of the cognitive system's identity and its cognitive engagements with the world (sense-making). Bodily action is not limited to movement, or other behaviorally visible action and it is not only realized by biological structures, but can also comprise tools and technology.

Based on the two dimensions of the enactive body, i.e. the body as identity and the body as interactive, we can now provide a more refined description of the role of the body and the possibilities of bodily action that subsist for patients in CLIS.

Since the CLIS patient is not able to exert overt movement-based bodily action in terms of the whole body, her sense-making interactions are severely restricted even though limited interactive engagement may still possible through communication via eye-movement. However, according to the above suggestions, the interactive body may also comprise bodily processes, which are not behaviorally observable (covert). One could thus assume that the CLIS patient can still rely on *covert* bodily processes to sustain the body as a means of sense-making. The patient is still sensitive to the world or can for example hear other people's voices. Since these engagements rely on activity that is partly covert, i.e. not involving muscular movement, they could arguably count as aspects of the interactive body that still subsist despite the paralysis.

The question however remains to what extent these covert bodily processes can still contribute to the maintenance of the autonomous body as interactive. The enactive approach as presented here presumes that the interactive body requires some kind of regular update process. A mere developmental role of bodily engagement is ruled out. Once again this does not mean that overt bodily processes need to be constantly enacted in order for covert bodily process to persist. But one can hypothesize that in the permanent absence or a severe limitation (to eye-blinking movements) of overt movement-based actions there will be a limited experience of a match between a bodily intention and its effect in the world. This could also affect world-directed bodily action that is covert, such as sensitivity to touch and auditory stimuli.

A second question is how covert bodily processes can sustain the body not with regards to interaction and sense-making, but as an autonomous identity, i.e. a distinct center of existence. If bodily identity was not independent from the body in action then one could raise the worry that a decrease of bodily interaction would also lead to an imbalance between interaction and identity dimension of autonomy. It is conceivable that due to a continuous lack of interaction the sense of identity becomes increasingly emphasized. This would provide a route towards an explanation of the phenomenon of bodily alienation that is frequently experienced by subjects with bodily paralysis and other health problems. It could be because being a body is not only being an identity, as a distinct center of existence, but crucially also an interacting center of existence that in cases of severe disturbances in the interaction process, the body can feel uncanny, instrumental and be experienced as not properly belonging to the subject. Such experiences could hint at the body's intrinsic purpose of engaging with the world. Such imbalance between the body in identity and interaction might also account for the increase of "mental" activity and imagination in LIS patients that Dudzinski has observed (3.2.2). This issue appears much more evident in the case of TLIS, as we will see in due time.[78]

6.5.2 The Enactive Body and BCI-use in Classically Locked-in Syndrome

Let us first reconsider the status of BCIs for patients with CLIS. According to an enactive perspective, bodily action need not be physiologically realized. In the previous chapter, I suggested with Merleau-Ponty that the BCI, just like the stick of the blind person, can become incorporated into the habitual body of the LIS patient (Section 5.2.3). This is to say that the BCI

[78] In chapter 8 I come back to the phenomenon of alienation in the context of schizophrenia and elaborate the idea that it could hint at an imbalance between constructive and interactive processes in autonomy.

helps the patient to accommodate the paralysis in so far as some of the processes that ensure the maintenance of bodily autonomy are now realized by covert (neurobiological) and overt artificially enabled bodily action that involve no muscular activity at all. The BCIs ensure the continuation of the body's sense-making activities and thus provide an alternative way to enact the interactive body. Once again, understanding how BCIs can realize the body's capacity for sense-making still requires a clarification on the role of BCIs with regards to maintenance of the body *as identity*. How does a change in interactive bodily structures affect bodily processes that contribute to identity construction? Light needs to be shed on how both dimensions of the body, identity and sense-making are generally interrelated.

6.5.3 The Enactive Body and TLIS

Above I have suggested that the CLIS patient's embodiment while being restricted has not been affected at its core because minimal sense-making engagement is still possible. However, in the case of a TLIS patient who does not use BCI and for this reason has no way to enact any kind of *interactive bodily* action, the enactive body concept seems to lead to a different account. If we assume that the interactive body constitutively relies on some sort of continuous wordly engagement then a complete cessation of interactive bodily action should in principle lead to some kind of diminishment of the body in interaction itself.

Clearly, to suggest that the TLIS patient can no longer rely on the body as a means of sense-making does not imply that the patient would *immediately* cease to be involved in any other bodily activity. If self-affection counts as a kind of bodily engagement that does not involve any worldly interaction then even a patient who ceases to be directed at the world could still continue to be bodily active – she could arguably still engage "with" herself (see 5.4). However, I have also pointed out that while there might be an experienced separateness of world-directed and self-affective bodily action from the perspective of the person being the living body, the same need not be the case for the external observer. From an organizational perspective, it seems that we cannot clearly separate both aspects of bodily autonomy. To say that the body is the means and purpose of being embedded and engaging with the environment implies that the two aspects of embodiment (sense-making and identity) are deeply intertwined. They do not exist completely independently from each other – being a body requires both being a distinctive entity *and* the engagement with the world. According to this, it would seem that a TLIS patient is "reduced" to embodying a distinction without actually being distinct from anything. In other words, if the patient is not able to engage with the world, not just temporarily, but *ever*, how could she remain a distinct systemic whole?

One way to overcome this problem could be to insist that the patient still possesses, or rather is, a functioning *organism*. Though there may be a complete and permanent cessation of overt bodily action that affects its interaction with the world (and thus the capacity for sense-making), the organism on its own could arguably sustain a bodily autonomous identity. It could constitute the necessary distinction from the world simply by providing a material boundary to the environment allowing the patient to remain minimally sensitive to particular environmental changes. Recall however, that we should not equate embeddedness with being spatially located (Section 5.2.2). I believe most people would be intuitively very reluctant to say that the TLIS patient is only located, but not embedded or embodied anymore.

Unfortunately however, even the maintenance of the patient's organismic identity is not entirely a matter of the patient herself. The patient requires substantial help from the environment through artificial nutrition and respiration. She thus "relies on" an immobile organism whose very survival depends directly on the continuous aid of the external environment. As a consequence, holding that the patient continuous being an identity because she still is an organism has the undesirable implication that bodily identity is equated with the mere presence of a living, or rather, kept alive, organism. With regards to human beings this seems at least highly counterintuitive.

To understand bodily identity in the case of TLIS is extremely difficult, not only for the pragmatic reason that we cannot ask the patient herself (Kurthen et al., 1991, see Chapter 3). The present discussion reveals also a much more deeply rooted epistemological reason: It seems that when speaking of bodily identity – at least in the case of human beings – we may actually have never only referred to identity in terms of being an integrated *organismic* system but rather all along have taken it to imply something that transcends organismic boundaries.

The enactive account of TLIS seems to reach its limitations when considering how exactly the two components of the body as autonomous identity and as involved in sense-making interrelate. Clarifying the relations between different *phenomenological* descriptions of autonomous identity and interaction (Section 6.1.2) by linking embodiment to autonomy could be the beginning to account for this problem in. But it seems unclear however how it might account for a very particular aspect of this linkage, namely whether and how it is that identity and sense-making processes co-constitute each other. It seems that in order to address this question we first need to understand what contribution the organism actually makes to the (bodily) identity of the LIS patient. This as I argue shortly requires that we refine our perspective on cognition not only with regards to the bodily, but also the social dimension of cognition.

6.5.4 The Enactive Body in TLIS and BCI communication

In this section I reconsider the case of TLIS patients who rely on BCI for communication. While in the previous section attention was drawn to the identity-dimension of embodiment I now focus again on the interactive body. BCIs were considered as a means to accommodate the physiological restrictions of the CLIS and TLIS patient by ensuring a continuation of bodily interaction with the environment. However, in this discussion no further attention was given to the particular kind of interaction the technology actually enables.

We saw in Chapter 5 that there seems to be a fundamental difference between interactions involving *objects* and those involving *subjects* (Section 5.2.3). Other than in object-manipulation the action structure in communication is not uni-directional, but interactive. Though it is "through the body that we communicate with others" (Cole, 2010, p. 669) the communication process is not reducible to the individual body. This is because the realization of an intended act (of communication) is not a matter of the patient's intentional action alone but requires at least one additional step: the engagement of another person. Thus remains the challenge derived from the discussion of the phenomenological body in the context of BCI and totally locked-in syndrome at the end of the previous chapter: How can we account for the role of embodiment in interactions with the social environment?

Unfortunately it turns out that the previously outlined concept of the enactive body is equally unable to account for this question. The reason for this is linked to the problem of understanding the relation of bodily identity and sense-making that I just highlighted in the previous section.

The problem with the previous analysis of bodily sense-making is that it involved not only a focus on the object-environment but crucially also an *individualistic* perspective on the body. This individualistic perspective is inherent throughout the entire body discussion of the last chapters. We have always considered the patient as an isolated subject that interacts with a world that remains distinct from her.

As an immediate response to this worry one might wish to point out that we already allowed for bodily processes to transcend the *organism* – in interaction with the environment and the inclusion of non-physiological processes. The boundaries of the body can accordingly not be equated with the boundaries of the organism. Why would I still claim that this involves an individualistic perspective? The reason is that this case and my above statement imply several different notions of boundary. In the case of bodily action that transcends the organism, 'boundary' refers to the distinction drawn between those processes that belong to the organism and those external to it. However, when looking at the body we also find a different kind of

boundary, namely the boundary between the autonomous body *as a whole* that comprises interactive and non-physiological processes and the environment. In this case 'boundary' thus refers to a delineation between non-organismic in conjunction with organismic processes (both belonging to the autonomous body) from those processes that are not, i.e. that are part of the body's social and material environment.

Therefore, when I claim that throughout the previous chapters the body has been considered only from an individualistic perspective, I refer to the second notion of boundary between the autonomous whole of the body and the world. What I mean by 'individual body' is not the individual organism but the body that already integrates both organismic *and* non-organismic components.

This offers a more fine-grained reason for why I believe the enactive body concept is unable to account for BCI in communication of LIS. Even though the elaboration of the body concept allowed the constitution of bodily action to transcend the *organism,* this did not consider the possibility of the body as individuated against the background of an *environment of others*[79]. To claim that bodily identity in the non-communicative TLIS patient is preserved by means of her organismic boundaries is odd when considering that embodiment is never only a matter of organismic boundaries. It is also odd, however, because even if from a mere organismic perspective we are not only distinctive entities against the background of material objects but also individuals in a world of others.

6.6 Conclusion

Where do the above considerations leave us with regards to the basic challenge posed by LIS and BCI? I have suggested that LIS is of general interest to cognitive science because it puts into question not only the body, but also the body in relation to the self and the social environment (see Chapter 3). The fact *that* these concepts are related has been taken for granted in the philosophical approaches to LIS and I provided a summary of the phenomenological descriptions of some of these interrelations (see Section 3.2). However, none of the approaches explicates *why* and *how* these concepts are interrelated. I thus claimed that the relation of body, self and sociality is an open issue for cognitive science. And I suggested that the way to address this issue would comprise two strategies: On the one hand, a focus on the concepts at stake. But on the other hand, a focus on the background epistemology, i.e. a greater framework in which the interrelations between the concepts are determined. The previous discussions have illustrated that these foci are not independent from one another because an explanation of a particular concept is always affected by the presuppositions made about the relation it bears to other concepts.

In the ensuing Chapters 4 and 5 I have then focused on one of the concepts challenged by LIS, namely the body. We have found that none of the introduced approaches to embodiment (the functionalist, sensorimotor or phenomenologist) on their own could account for LIS in terms of embodiment. However, by grounding the body concept in the enactive notion of autonomy we were able to outline an approach to the body that integrates important aspects that were beforehand treated more or less independent from each other. The body as autonomous system, and not as an instrument of the brain, has been characterized as the minimal system capable of cognition. The explanatory benefits of the enactive perspective on the body were 1) an

[79] We already saw that recent work in enactivism has sought to account for this individualism by proposing that individual agents engage in so-called *participatory sense-making* and thereby enlarge their realm of sense-making capacities. However, I argue in more detail in chapter 7, that the present discussion reveals another and novel form of mutuality. The notion of participatory sense-making as it is most frequently referred to can not fully account for the problem because it is primarily concerned with the role of the individual for the social. My point here is that in addition to considering how several agents engage in social interaction which constitutes a novel kind of autonomy (not reducible to the individual agent) we should consider how social interaction figures in the constitution of the individual agent.

elaboration on the different kinds of bodily action 2) an account of the interrelation of cognition, body and consciousness and 3) a non-contingent stance concerning the role of the body beyond developmental purposes (see Figure 6.1).

	Sensorimotor approach	**Extended Cognition**	**Phenomenology**	**Enactivism**
Split Consciousness-Body	an implicit split between experience and bodily engagement	explicit	no, consciousness is inextricably linked with embodiment via existence in the world	no, consciousness is inextricably linked with embodiment via normative existence in the world
Consciousness	only object based, no account of general subjectivity	no relevance	yes, in world-directed hetero-affection and immanent self-affection (unclear how they interrelate)	necessary aspect of autonomy, establishes viewpoint on world and autonomous system itself based on which system evaluates actions
Movement	yes, but not necessary	-	yes, necessary	yes, but not necessary
Physiological Necessity	yes	no, external tools can substitute bodily functions/extend the mind, based on multirealizability claim	no, by becoming part of habit body external tools can act as non-physiological body component	no, as long as the tool is incorporated into the body by linking it to preservation of autonomy
Split Body - Social	yes	yes	yes	yes, in an individualistic concept of the body no, in so as far as enactivism offers basis to accommodate social dimension

Figure 6.1: A comparison of the body concept in sensorimotor approach, extended cognition, phenomenology and enactivism. 1. The sensorimotor approach: it associated the body with movement and physiology and is thus too restrictive. 2. The extended functionalist approach: here the body plays a contingent role for cognition only. This notion turned out to be too lose. 3. The phenomenological approach: this approach encompasses the previous two conceptions by allowing e.g. for non-physiological objects to be incorporated into the body. It turned out to be limited as it offered no explanation of how different aspects of bodily subjectivity interrelate and could not account for the role of the body in social action. 4. The enactive approach: it offers an integrative notion of the body grounded in the concept of autonomy. While it was able to address almost all of the refined challenges, it also reaches conceptual boundaries with regards to social (inter)action.

With these steps towards an enactive approach to the body I tried to indicate a true alternative to the orthodox view of cognitive science. In this proposal the body is not only a matter of adding something to the cognitive mix. We have taken seriously the idea that the cognitive system is neither like a computer, nor to be found in the brain, but rather like a living system. The body is the non-negotiable purpose and means of being an embedded cognitive system. This led to a body concept, which involves external processes and is inextricably linked with consciousness from the very beginning. We do not approach cognition and consciousness by starting with a brain and then add a body to it, but the other way around. The appropriate epistemological lens for understanding cognitive processes is the whole embodied cognitive system as it evaluates its action with regards to the goal of self-preservation.[80] This conception

[80] Once again, this is clearly not to say that the brain plays no important role in cognition. But the brain is associated with the task that the body is pursuing: to enable a cognitive system to be and relate to the world (Fuchs, 2011).

could save embodied cognition from positing a restricted or merely trivial notion of the body and also from the danger of falling back into the classical position it had sought to transcend.

However, even though the enactive account of the body assumes that the body is crucially a subjective body (with experience of hetero- and self-affection), it does not yet consider that – at least in humans – an individual body is embedded in and distinguished as an individual in a world of other subjectivities.[81] Cognition in the enactive perspective is not limited to the individual (as seen in the participatory sense-making approach). However, body and bodily action remain so far considered from an individualistic point of view.

This has consequences for the present context. If, as the case of BCI communication suggests, not all cognitive action can be explicated in terms of the body or the individual body on its own then the previous refinements on embodiment cannot fully account for the role of the body – and therewith also that of the organism – for cognition in LIS and BCI.

There thus remains another pressing question for cognitive science: how is the cognitive system affected in its identity and sense-making autonomy when it is considered as a cognitive system embedded in a world of others?

If cognitive science is to answer this question it should obviously not adopt an orthodox view of cognition. But I also believe that it cannot remain with a "mere" individualistic embodied view. The orthodox view of cognition view implied a split of body and mind, which embodied cognition then reformulated as a split between the physiological body and subjective body (the "body-body problem", see Section 2.3.1). We elaborated this issue by adding the question of how world- and self-related experience are interrelated. Unfortunately the body-body is not resolved by understanding how subjective and objective aspects are integrated in the individual organism's existence (as, e.g. Hanna and Thompson, 2003 and Fuchs, 2012 have it). This resolution still runs the risk of falling into another split, namely that between the individual body and the world of others (or other bodies). The body does *not* shape the mind through "motor abilities, its actual movements, and its posture" (Gallagher, 2006, p. 8). It also *contributes* to shaping the mind by enabling participation in social interaction dynamics. The actual problem I thus conclude is not merely a body-body but a body-body-social problem.

In the following chapter I make a proposal that envisions a more encompassing perspective on the cognitive system. From this perspective the social world matters not only for a cognitive system's sense-making interaction but also in that it contributes to the constitution of its very identity as a cognitive system. This will offer a way to account for the body-body-social problem.

[81] At least as it has been discussed in the present thesis. That the body is linked to the self and involved in social interaction comes without doubt. Yet how it is linked cannot be assessed from within the conceptual boundaries of embodiment.

Chapter 7.

Steps Towards an Enactive Approach to the Self

> [S]ociality is an indispensable element of the life–mind continuity thesis and of cognitive science more generally.
> (Froese and Di Paolo, 2009)

In the previous chapter we saw that the basic challenge outlined in Chapter 3 – to account for the relation of body and self in LIS patients as well as the role of BCIs – could be partly addressed in terms of a refined enactive approach to the body. However, whereas movement and interactions with the material environment clearly matter in that approach, the fact that human beings also engage with each other does not seem to find its way into the characterization of the basic nature and constitution of the human mind. Recent cognitive science has overcome the clear-cut separation between the individual cognitive system and the environment. The environment is not an objective or independent given and cognition does not consist in solving problems imposed from the outside. Cognition is rather based on a dynamic interrelation of individual and environmental processes. From the enactive perspective this has been specified in that the boundary between system and world emerges based on the embodied activity and structure of the cognitive system and the evaluation of its engagements according to an intrinsic norm. This offers a clear alternative to the classical input-output model of cognition.

What is however widely ignored in the attempt to overcome the cognitivist split is that at least with respect to cognition at the human level there still exists a particular version of the split: a dichotomy between the embodied and active individual subject and its social surrounding. I believe this dichotomy is one reason why there remains a body-body-social problem and the question how global paralysis or the use of BCIs impact a patient's life has hitherto remained unresolved. In this chapter I make a proposal for resolving this issue.

We learned that according to enactivism a full account of cognition requires taking into consideration the system's subjective perspective on the world based on which it evaluates its sense-making. This evaluation is grounded in the system's intrinsic goal to preserve its identity. Enactivism thus "looks upon cognitive systems as self-referential systems" but crucially it is also itself a "consistent self-referential theory, because it starts and ends with the operations of an observer" (Baerveldt and Verheggen, 1999, p. 190). Phenomenological enquiries are thus part and parcel of the enactive methodology. To understand how humans evaluate their sense-making interactions with the world and what kind of identity they strive to preserve requires the researcher to take a phenomenological perspective herself (Froese and Di Paolo, 2009).

For this reason in this chapter I shift the focus from the concept of the body to the self and how it is experienced from a subjective perspective. I provide a couple of everyday scenarios that explore from a phenomenological perspective what it means for a human being to be an autonomous identity. These examples focus on human identity not as concerned with the individual body but as it is experienced in a social context. Who we are as individual cognitive systems is not experienced only in interaction with a material environment but rather co-determined by the existence of others, by their perspective on us and by our directedness towards them. Our experience tells us that we are systems that strive to be different from others while still being able to engage with them. I take these experiences to be of vital importance for understanding what it is for a human being to be an identity.

In the second part of this chapter I thus explore in more technical terms of the enactive notions of autonomy, precariousness and sense-making how the phenomenological observations

help to inform a more general account of human autonomy. My hypothesis is going to be that social interaction processes play a vital role for human beings, not only in the obvious sense that humans have social cognitive capacities but also in that social engagements construct and shape the individual human being itself. From an enactive perspective, human beings are considered as embodied systems that are also primarily social. My suggestion is that at this level of the life-mind continuity the notion of autonomy converges with the phenomena associated with the self.

Based on these explorations I come back to the challenges outlined in the previous chapters and show how an enactive approach to the self can inform cognitive science's perspective on embodiment and thus provide ways to account for them.

7.1 Self in Context

I would like to invite you to consider the following scenarios and imagine how you would experience these or maybe similar situations from your own perspective.

1. A Walk in the Street
Imagine you are invited to a party and you dressed up for the occasion. The location is not far from your place and you decide to walk there. People are walking by. Suddenly you notice that a person approaching you, just before passing by, clearly looks at you. You notice that she is scanning you from top to bottom, seemingly evaluating your outfit. At this moment you become aware of yourself and wonder whether you are dressed appropriately. In this situation the gaze of a stranger makes you aware of the fact that you are there, as somebody other people can look at, and judge. Your self-awareness is raised by the other person's look.

2. Job Interview
You are invited to an interview for a job that would mark a big step for your professional career. You enter the room to find an evaluation committee of three people and you are asked to take a seat in front of them. You sit down and await the questions. One of the jury members asks you to be patient because he wants to find your application in the pile. At that moment you become very aware of the entire context and your role in the situation. You know that they will ask you questions concerning your previous work and that they will probably test in some way or other whether you really are the right person for the job. You are in this room because you have to convince them that you are, in fact, the expert they are looking for. Again, your sense of self is heightened.

3. Being the Stranger
You are visiting a foreign country. You have virtually no language proficiency. After a long day of sightseeing you decide to get yourself some food and drink and go back to your hotel. At the supermarket you are not able to find what you need and you approach the shop assistant to ask for help. You do not know the appropriate words but you make great efforts in finding alternative ways to describe what you are looking for. Unfortunately, the assistant keeps giving you a puzzled look. An easy thing to do at home is suddenly effortful and complicated. With your last attempt to explain what you need she even seems to get stressed and annoyed with you. The assistant looks at you expecting you to express yourself but she does not understand you. She beckons to another sales assistant, but that does not improve the matter. You begin to feel bizarre and separated from them. You leave the shop frustrated, and the city you had enjoyed all day long suddenly becomes a strange place to you.

4. Letting Go in Tango
You are at your favorite dance location. On this particular day you are having an extraordinary experience with your dance partner. Your dance is so smooth and effortless that you are forgetting to watch your steps or control the movement. You are giving yourself to the present moment, you are connected, in direct touch with your partner's body. You experience the dance as flight-like, a flow of movement in rhythm and perfect harmony. It feels as if the boundaries between you and the other have disappeared, you do not guide, but let yourself be guided. It almost feels as if you have become one. (If you have no experience in dance, you might think of a similar experience of harmony in a sexual encounter with somebody instead).

5. Falling in Love
Remember what it is like to happily fall in love with someone. A part of the process is that you are fascinated by each other and cannot wait to explore the world together. You are awaiting the moment to see the other person again. Excited like a child. Every encounter makes you feel light, happy and deeply cherished. Everything about the other person is magical and interesting. Especially if the affection is mutual you feel wide open and ready to let yourself be moved by the other and to absorb her presence fully.

6. Being at a Concert
Imagine being at a concert of your favorite musician or band. The concert is sold out. The hall is full of cheering, happy and dancing people all tuning in the with artist's performance. There are moments in which you experience yourself as part of something bigger than you. You feel connected and in sync to perfect strangers.

7. The Birthday
It is your birthday and to your surprise friends have organized a party for you. You come home to find a bunch of very happy people singing you the birthday song and offering you your favorite cake. In the beginning you are shocked, unsure what to think and do about the situation. All of this is happening because of you. You become aware of yourself as everybody is looking at you, expecting you to be happy, and to say something. You feel awkward in the beginning. However, as the party continues you are loosening up, immersing into small talk with your friends, laughing, cheering, making jokes. You forget about the fact that the entire context was set up for you and just enjoy being in good company.

8. A Conference Talk
Your paper got accepted for a conference. You are a little nervous because the ideas you are going to present are new. During your talk you get the impression that people are not able to follow you or that they are bored. You become aware of yourself, you feel a little awkward and at a distance to the audience. But, as soon as the discussion starts, several people are raising their hands and you are asked questions that inspire you and let you elaborate or emphasize particular aspects of your work. You are relieved. Now, fully engaged in the discussion there is no separation to the others and you feel that your work is appreciated.

9. A Fight
You are talking to your partner and try to make her to understand a matter of deep importance to you, but she keeps misunderstanding you. While explaining your problem you feel open and optimistic that you will soon find a solution. Usually your partner is on your side. However, today the situation is different, she is insisting that you are mistaken. You feel a bit sad and separated. Nevertheless, you are making further attempts to express your concern. When your partner keeps misunderstanding you, and even begins to talk about something entirely different, you become angry and disappointed. Throughout the conversation you experience a back and forth between

wanting to engage and connect to your partner and feeling separate with your concern not being shared by the other.

These scenarios are examples of social interactions as they are experienced from the individual perspective by one of the participants. In all of these situations we can observe that how we experience ourselves is nothing that happens in isolation of an engagement with others. However, they also highlight two general qualities of that self-experience. On the one hand we find that we experience a kind of awareness of ourselves as an individual, as being separate from the other. Yet, on the other hand we also experience that we are engaged with others, directed to them, open and ready to be affected by the interaction with them.

Scenarios 1–3 exemplify social interaction processes in which the individual comes to experience herself with an increased awareness of being an individual. In the first scenario, this is due to the other's evaluating look; the experience of alienation in case of the tourist visiting a foreign country is intelligible only against the background of the failed communication with the sales personnel. Interestingly thus, an experience of being oneself seems to not only reside within the individual subject, but also involves another person's action being directed at her. What makes us feel self-aware and being a distinct individual is also the result of something happening between subjects.

Scenarios 4–6 illustrate social engagements in which this quality of being separate, individual or distinguished from the other is less prevalent. They rather show that we also experience ourselves as being connected to other people. The quality of self-awareness in such cases could be described as a particular feeling of openness and a readiness to engage and participate. When we fall in love, e.g., we are not so much concerned with us as a separated being, but experience ourselves as open, eager to meet and connect with the other.

Scenarios 5–8 finally illustrate that these two qualities, of feeling oneself as an individual and as somebody willing to engage with others, are not entirely independent or clearly separable from each other. They can precede and follow each other, they can even overlap. There are situations in which we experience the shift from one quality to another quite clearly. If, e.g., in a difficult discussion our partner finally seems to understand what we want to say, a relief or a sudden relaxation may appear, upon which we begin to feel less separated from the other and begin to experience a readiness to be open again. In case of the surprise birthday party we could observe a shift from feeling very much like a distinct person to just enjoying being there with others. However, even in this first instance of feeling separated and the awkwardness that might accompany this, one can already experience the wish and need to just be able to "be oneself" in the company of others. Similarly, at a conference we can experience both a sense of separation (for instance because of nervousness in the face of criticism) and a sense of eagerness to engage with others (because we would like to present the new ideas) at the same time.

With regards to the present context, I want to suggest that these two qualities of experiencing one-self in social interactions with others, i.e. the sense of being someone individual and someone who can engage with others, reflect something fundamental about the nature of the human beings. They indicate a tension in humans between striving to be distinguished from others while in the same time striving to be connected and engaged with them. The suggestions is that this tension is at play even in less explicit situations and whether or not we are fully aware of it. It matters for human cognition in general. We are not isolated beings. We are embedded in a socially structured world of other people and we are involved with them in a fundamental way. Who we are as individuals is to a great extent enabled by and vitally dependent on social interaction processes (Mead, 1934).

With regards to understanding the human mind this would mean that intersubjectivity matters for cognition and consciousness not only at a particular developmental stage of the adult and mature human beings, but also that individual human mind is co-constructed from the very beginning and that humans thus need to constantly negotiate their existence against a social background.

In the following sections I explore this hypothesis in terms of the enactive framework of cognition.

7.2 Towards An Enactive Approach to Human Autonomy

The basic insight from these observations of self-experience in a social context is that the individual human mind should be considered from a social perspective. This perspective will lead to a different view of human autonomy. It is a genuinely social autonomy. According to this, social engagements seem to play a role in the very organizational network that makes up human autonomous identity. They strive to remain individuals while also engaging with others. This proposal will result in adding a new dimension to the higher forms of autonomy on the scale of mediacy (see Section 2.2.4).

The following considerations are exploratory in nature and do not provide a fully defended position. To develop such a position, i.e. a coherent enactive approach to the self, would require a detailed elaboration of the basic concepts of enaction. We would need to reconsider how they interrelate in light of the idea that human cognition is at heart socially mediated.

7.2.1 Cognition from a Social Perspective

That sociality matters for cognition is not a new idea. But though there exists a range of research that is also concerned with so-called social cognition, such as theory-theory and simulation theory (Gallese and Goldman, 1998, Carruthers and Smith, 1996, Vogeley et al., 2001) these approaches seem to entail an "objectifying" attitude towards the existence of other human beings. The other is relevant in so far as she presents a problem that needs to be resolved. She is an object in the world, whose behavior needs to be manipulated (Frith and Frith, 1999). According to these approaches to social cognition understanding the other is to represent their mental contents (see Gallagher, 2001, Fuchs and De Jaegher, 2009, Di Paolo et al., 2010, De Jaegher, 2010, De Jaegher et al., 2010 for recent objections,). This representationalist stance on intersubjective engagements presupposes that there is a gap between self and others (Hutto, 2010).

From a phenomenological perspective however we understand that subjectivity does not end at the borders of our individual bodies. We are not islands of embodied subjectivity surrounded by manipulable objects, but embedded in a world of other subjects with goals and perspectives of their own. Our environment is a lived world, socially structured from the very beginning (Schütz and Luckmann, 2003, Gurwitsch, 1964). Others exist not independently from us and they affect how we experience ourselves. This should be mirrored in our description of human mind. My proposal is therefore that human autonomy should be defined in social terms. Such a proposal implies that human cognition at heart involves intersubjective processes.

This idea has been already explored by recent work in enactivism. Recall that enactivism aims at an alternative explanation of the dynamics of low-level cognitive phenomena, but also at providing a pathway to the "highest reaches of human cognition" (Froese and Di Paolo, 2009, p. 439). It is here that the inclusion of the social realm becomes crucial as it arguably helps to bridge the "cognitive gap" between lower and higher levels of cognition (see Chs.2 and 4). Froese and Di Paolo have suggested adopting the following two-fold strategy (ibid., p. 443): Firstly, beginning at the lower level of complexity one can show how the immediate domain of cognitive systems can be expanded through interactions between several cognitive systems. Secondly, starting at the higher and phenomenological level, we understand that in humans cognitive processes are shaped by intersubjective engagements.

An application of this strategy is illustrated by the notion of participatory sense-making: The ongoing coordination of the behaviors of individual autonomous systems engaged in social interactions can bring about a new systemic entity that can be described as a form of autonomy

with "self-maintaining tendencies," namely the interaction process (De Jaegher and Di Paolo, 2008, p. 38). In connection to this, participatory sense-making also establishes new "domains of sense-making" in that the newly emerging interactive autonomy opens up a realm of individual sense-making of the participants, "unavailable to each solitary individual" (ibid., p. 41). The autonomy of interactive processes thus impacts on the *individual* autonomy in that individual sense-making capacities are augmented or modified by participation in social interactions.

So far enactivists thus discern two types of autonomy that involve social interaction: autonomy as referring to the individual agent engaging in social interaction, i.e. "*agent-autonomy*" and the new kind of autonomy resulting from the interaction of different agent-autonomies, i.e. "*interactive autonomy*" (Colombetti and Torrance, 2009, p. 32).

In this chapter I suggest considering an *additional* relation between sociality and the individual cognitive system. The idea is that social interaction processes matter for the individual autonomous system not only with regards to its sense-making capacities, but also for the construction of its very identity as an individual. Even though it is not made explicit, this relation is already entailed by the previous considerations. Consider the definition of social interaction suggested by De Jaegher and Di Paolo:

> Social interaction is the regulated coupling between at least two autonomous agents, where the regulation is aimed at aspects of the coupling itself so that it constitutes an emergent autonomous organization in the domain of relational dynamics, without destroying in the process the autonomy of the agents involved (though the latter's scope can be augmented or reduced). (De Jaegher and Di Paolo, 2007, p. 493).

For someone who presupposes that human individuals are constituted only by means of the individual this definition could be interpreted as saying that social interaction matters only in so far as it presents a context in which "ready-made" autonomous individuals are interacting with each other. However, even though the above definition focuses on the role of social interaction for the sense-making capacities of the individual agent, it does not rule out that social interaction also and already matters for the constitution of the individual autonomous system itself. To augment or reduce the "scope" of the individual agent's autonomy does not have to be limited to the scope of sense-making. It could include the agent's identity itself.

Therefore, when I suggest considering the individual cognitive system from a social perspective I want to focus on a specific dimension of this perspective. I do not focus on interactive autonomy as a result of coordinated interaction of individual autonomous systems. I also do not focus on the role of interactive autonomy for the individual's sense-making capacities, i.e. the behavioral aspects of human cognition. The present proposal rather focuses on the role of interactive autonomy for the *very identity of the autonomous system*. From this perspective, sociality matters for human cognition with regards to those aspects that constitute the cognitive system itself (its autonomous identity).

We have learned that the organizational principle of autonomy ensures two basic aspects, viz. the existence in a world as a distinguishable identity as well as the striving for sense-making interactions with the world (see Sections 2.2.1, 2.2.2). If we acknowledge that from the subjective perspective being a human cognitive system is experienced against a social background, then this should matter for both of these aspects of autonomy. As a consequence sense-making interactions at the level of human beings are primarily sense-making interactions with others, viz. social sense-making. Contrariwise, being a distinguishable identity means for a human autonomous system that it is distinguished and distinguishes itself from *other subjects*. My basic suggestion is thus to envision an approach to the operational principle of human autonomy in terms of a network of processes that are *socially* defined.

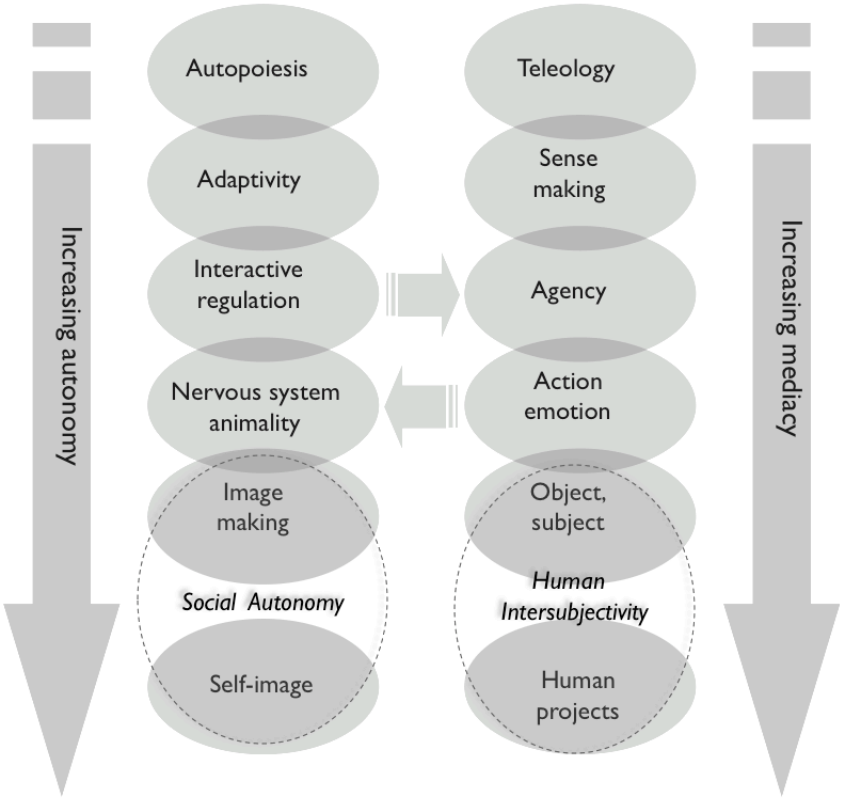


Figure 7.1: Social autonomy and the scale of mediacy. The two dashed ellipses (social autonomy and human intersubjectivity) illustrate a possible way of locating human intersubjectivity within the life-mind transition as it was proposed by Di Paolo et al., 2010 (see also Section 2.2.4). According to this, we find at the level of human beings a form of autonomy that is enacted through intersubjective processes.

Such perspective can be seen as an elaboration on the life-mind continuity thesis and contribution to understanding the higher levels of autonomy. Recall for this purpose the original scale of increasing autonomy and mediacy (see Section 2.3.1). The social perspective on autonomy would add a new stage for both, the dimension of increasing autonomy and of increasing mediacy. We would add a level of autonomy that is no longer defined in individualistic but social terms and that is mediated by intersubjective engagements (see Figure 7.1). Human beings would be situated at this level of increasing autonomy and mediacy.

This is of course not to say that the previously considered levels of autonomy are now irrelevant. Humans have a metabolism, a complex nervous system and a body, allowing for motility and a wide range of engagements that surpass basic concerns of survival. A description of human beings in enactive terms involves several different types and levels of autonomy. The suggestion is rather that at the social level of description human autonomy encompasses all of the above dimensions.[82]

In the next sections I expand on how the social perspective I propose could alter our understanding of human autonomy in terms of operational closure, sense-making, consciousness and intrinsic normativity.

[82] My proposal is not meant to imply that as soon as there is sociality we are in presence of socially enacted autonomy. Obviously, social interactions matter not only for human beings but also for other animals. My point is rather that sociality matters for human beings in a very particular sense, namely for their cognitive identity. Even though we might have to consider that there are grey areas (e.g. in great apes or in cases where an animal is raised by humans) in this respect, humans differ from other animals.

7.2.2 Closedness and Openness – An Overlap of Self-Construction and Interaction

In this section I reconsider the notion of operational closure in light of a form of autonomy that is socially enacted. I do this by drawing an analogy to the description of operational closure of the cell from Chapter 2 (Section 2.2.1).

In the example of the cell we have seen that each of the organizational processes making up the cellular network is enabled by at least one of the other process in the network while at the same time also being an enabling condition for at least one of them. This is the precarious operationally closed network of the cell organization. In the case of human autonomy this type of operational closure could be applied to capture the phenomenological observation that we experience ourselves as individuals separating and distinguishing ourselves from other individuals. Just like the cell, humans strive to maintain an autonomous identity based on which a kind of boundary arises making them distinguishable from their social environment. So far the intrinsic goal of autonomous systems is thus described as a striving to keep certain processes *outside* the network of their self-construction (see Section 2.2.2).

However, this seems not sufficient to fully capture operational closure from a social perspective. While striving to remain an individual subject, humans paradoxically also strive towards the contrary: They are ready to connect and interact with other human beings and let themselves be affected by them. To some extent they remain open to influences from the social environment.

In terms of processes constituting the organizational network of human autonomy this would mean that the processes constituting the organizational network have different qualities. At times they do not only relate to each other so as to ensure operational closure of the network but also allow social processes to contribute to the operational closure of that network. This entails a tension because operational closure usually implies that the system strives to exclude certain processes not in the sense that they are excluded at some point and then included at another.

Let me explore in what follows how this tension can be accommodated in the organizational description of socially enacted autonomy. So far we have distinguished between processes that contribute to the maintenance of a systemic identity, i.e. constructive processes on the one hand and behavioral processes, i.e. interactive processes (sense-making), on the other (Chapter 2). This would however entail a very limited reading of the term of operational closure. It seems that there are cases of autonomy for which the self-constructive and interactive dimension can no longer be clearly separated but appear as dynamically linked and overlapping. In order to account for the above-mentioned tension we should thus read the term operational closure in a more liberal way (we encountered this issue already in the context of embodiment, see Section 6.1.2).

Imagine therefore a simple organism for which we can clearly distinguish between processes that contribute to its construction (e.g., molecular, metabolic processes) and those that count as interactive processes (e.g., behavior, action, perception). Imagine now that the organism encounters a situation leading to a failure of the constitutive processes (e.g. in cases of illness or a genetic mutation) and thus its death. However, before the system actually dies it discovers a strategy, i.e. particular behavior to substitute this failure and to ensure its survival.

We find an example for this in a species of insects that survive under water by using air bubbles collected at the water surface (Flynn and Bush, 2008). The insects store the air bubbles in hair on the abdomen, which stabilizes the bubble mechanically and prevents it from collapsing under water. They then use the oxygen contained in the bubbles to survive longer periods under the water surface. For this reason one could say that the survival of the insect under water relies not only on its metabolism but also on a particular behavioral and interactive strategy. The closure of self-constitution, i.e. of what constitutes the identity of the insect, involves a behavioral arc – self-constructive and interactive processes appear not clearly distinguishable.

7.2.3 Distinction and Participation

The basic idea is that such kind of overlap between self-constructive and interactive processes is also at play in human autonomy. We have already learned that autonomous systems are not independent from the environment but enact their systemic identity while interacting with it. Clearly, such active engagement is also relevant for human autonomy. It is needed in order to feed the metabolism of the human organism. However, in accordance with the present proposal, human interactions are not limited to engagements of an individual organismic agent with the material environment.

From a social perspective we see that they also engage in inter-individual interactions. These social interactions are relevant for the agent's identity because they allow the agent to exist not merely as a separated organism, but as a *social* individual – an individual that distinguishes itself from others while in the same time striving to be connected with them. The social environment is not something that we can decide to engage with or not. It rather fundamentally conditions who we are. In enactive terms this would be to say that individual human autonomy is *socially enacted*, or co-constructed with others. This idea makes the socially constituted autonomous system a very special kind of autonomous system, for in order to be and become autonomous it vitally needs other autonomous systems.

It follows from the logic of this proposal that the overlap between self-constructive and interactive processes alters what we mean by intrinsic normativity at the level human identity. The fact that human beings are embedded in a social environment must be reflected in their basic goal of self-preservation. If human autonomy depends on social interactions then the autonomous system should care about whether it is able to engage in social interaction. Its striving to maintain a human identity must therefore entail a striving towards engaging in social interaction processes.

As a consequence, normativity in human autonomy differs from mere biological normativity because from this perspective on autonomy, the "direction" of sense-making in human autonomous systems would be towards a "self-generated identity" (Di Paolo et al., 2010) that is also and necessarily an identity that strives to remain open to structural change generated in interaction with others. It is in fact a *self-other-generated* identity.

To follow the basic goal of self-preservation thus means that humans do not only strive to be a separated individual. When it is my goal to be not only a separate individual but to also be an individual that can engage with others I have to ensure that in trying to realize either of these goals I do not lose track of realizing the other. If I only engage with the world in a way that makes me separated and distinguished from others I might find it difficult to be someone who can engage with others or who others can connect with. There are moments where I should be ready to let go of being a self-centered individual and participate more openly with others. Conversely, if I permanently engage in social interactions in which I dissolve myself I might lose my sense of being an individual in its own right. Because both goals are pointing towards opposite directions I have to ensure that they are somehow balanced – realizing one goal does not simultaneously ensure the other.

We find this situation mirrored in the division between self-affection and hetero-affection characterized as two basic aspects of human embodied experience in Chapter 6: We incorporate in our bodies two dimensions of significance, viz. that of the world surrounding us (hetero-affection) and that of ourselves and how we fare in bodily interacting with it (self-affection). However, hetero- and self-affection are not independent but mutually shape and enable each other (6.1.2). If the body realizes my engagement with the world I also need to have a sense of how I fare in realizing my engagement with the world.

Similarly, living in a social world requires that we understand both others and ourselves. But it is vital that the understanding of ourselves is at the heart also an understanding of ourselves *in relation to others*. Thus when we become aware of ourselves, this also involves some sort of awareness of how we fare as beings that are engaged with and directed towards others.

In the case of self- and hetero-affection we separate between processes belonging to the individual (self-affection) and those belonging to the individual as it interacts with the world (hetero-affection. However capturing self-experience at the level of social autonomy in organizational terms is different. If we follow the idea that in human beings the distinction between constitutional and interactive processes appears to dissipate, we face the difficulty that the conceptual boundary between an individual autonomous identity and its environment becomes hard to grasp. Social interaction processes are now both, part of the description of sense-making and also of the construction of an autonomous identity. For this reason I suggest amending the original distinction between identity-constitution and sense-making by a new conceptual distinction. This is to accommodate the overlap between behavioral and constructive processes and to capture the two-fold intrinsic strive of humans towards closing up against the environment while also remaining open to it.

When looking at the sea of processes that make up the network constituting individual human autonomy we find social interaction processes *in* the network that enable a closure of the network, thus establishing a distinction from the social environment. But at the same time we also find that these processes appear to open up to the social environment. We thus observe that the social interaction processes in the autonomous network can have different qualities, some pointing towards the system as a separate individual and some pointing to the opposite direction, i.e. the system as an individual that strives to engage with others. My suggestion is to capture these different directions as social interaction processes of either *distinction* or of *participation* (see Figure 7.2).

It should be clear that from the present level of description all processes constituting the autonomous network are interpreted as social processes and are thus co-constructed in interaction with other autonomous systems. Processes of distinction refer to social interaction processes that realize the goal of human autonomy to be a separate individual. Processes of participation are social interaction processes that enable the individual to be an individual that engages with others.

Both distinction and participation processes therefore involve processes that transcend the organismic individual system. This may be quite obvious in the case of *participation*. This tendency of processes could be associated with the participatory sense-making dimension already established by De Jaegher and Di Paolo (2007, 2008). We find it expressed by the phenomenological observations on dancing tango or going to a rock concert. In both scenarios we engage with other people and create a shared sense-making process that is based on the dynamics of the interaction and not reducible to the individual's action.

Note however that in the context of this proposal participatory sense-making does not (only) account for the behavior of the autonomous system but qualifies and contributes to its very identity as an autonomous system. Having a flow like experience in sex or dancing, becoming one with the crowd, or feeling genuinely open to the person we fell in love with clearly affects how we experience ourselves.

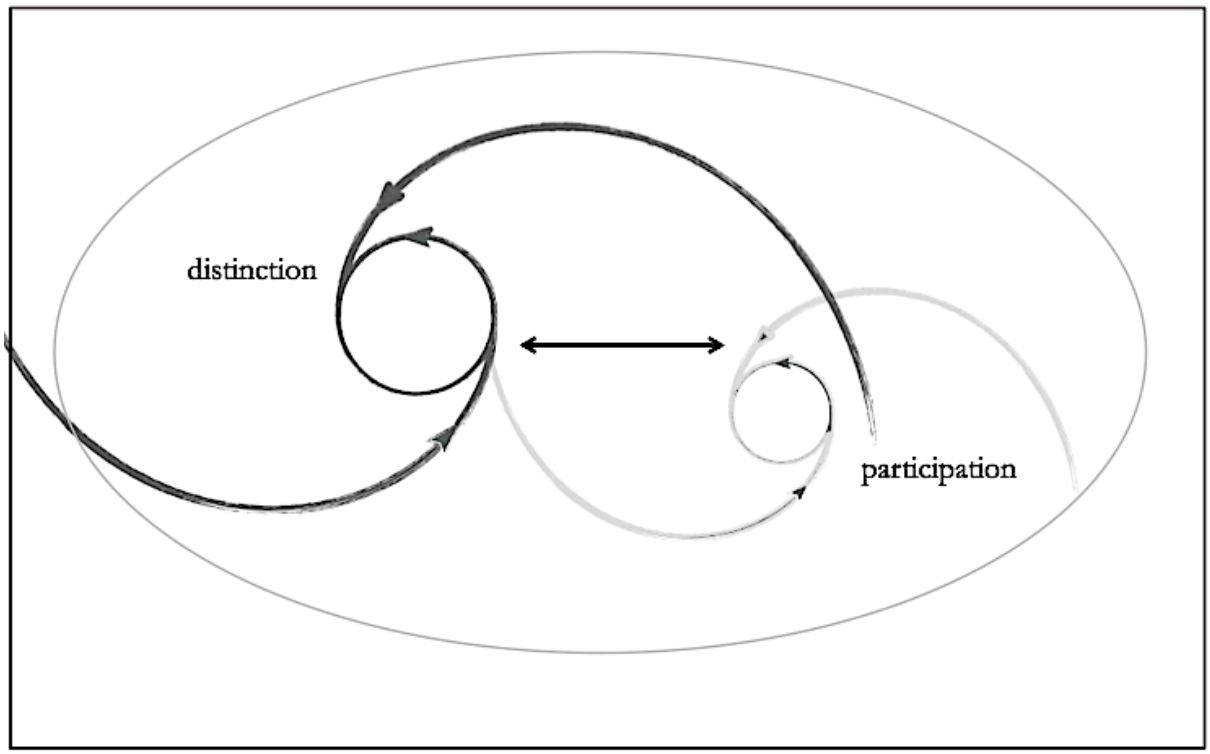


Figure. 7.2: An Illustration of Distinction and Participation. In both elements of the illustration the arrowed "arms" represent types of social processes that co-construct the individual autonomous system (the ellipse). They give rise to two types of self-constituting processes: distinction and participation. The circular figure on the upper left illustrates the pole of *distinction*. The circle is bigger than that of *participation* (lower right). This is to highlight that distinction processes contribute to the goal of establishing a boundary to the social environment and "increasing" the autonomous identity as distinguished from other autonomous individuals. The circle illustrating participation on the lower right is smaller. This is to illustrate the identity of the system as less re-affirmed and more open, to be affected by the social interaction and as literally reaching out. The arrowed arms interconnect the circles to illustrate that distinction and participation are precariously interrelated, i.e. they enable and are enabled by each other. The arrow between the two circles indicates that the poles of distinction and participation thus constitute a range on which the individual's actions can be qualified as being more or less participating or distinct.

Importantly, even processes of *distinction*, the processes that qualify the system as distinct social individual, are not limited to the individual organism itself but comprise social interaction processes.[83] Having a fight with your partner, being critically looked at by a stranger, or not understood by the sales personnel – these examples involve social interaction processes that increase your feeling of yourself as being separated and thus distinct from the other.

Particular processes or interactions of processes can also show tendencies of both participation and distinction. We have seen an example for this in the case of giving a conference presentation: you are aware of the fact that you are going to be the center of other people's attention but you are also striving towards the engagement with them.

Based on these considerations, a description of operational closure at the human level of autonomy would thus involve two basic steps:

> 1) We assume that neither the constructive nor interactive processes can exist as constituents of organizational closure on their own (individual autonomy and sense-making overlap).

This can be specified in terms of particular tendencies that these processes have:

> 2) the processes in the network have to be organized in a way that allows the network as a whole to realize a two-fold basic strive towards including the others (participation) and towards excluding them (distinction) (see Figure 7.2).

There is however another important reason why normativity in human autonomy differs from normativity in metabolic autonomy. We have seen that social interaction processes acquire their own autonomous dynamics and involve other agents (see Section 2.2.4). The quality of a given interaction in some cases can thus push me towards a more participatory situation (of the kind we find in examples of a concert or dance) even though as an individual I was striving to re-affirm my identity, e.g. by seeking recognition. Contrariwise, I might want at times to engage in a social interaction in order to connect with another person but find that the dynamics of the communication take a quality that prevents me from realizing this goal. The communication takes its own dynamics; it is not entirely controlled by the participants in the communication.[84] A good example to illustrate this is the fight with a partner. Even though I would like to engage with the other, the conversation can lead to misunderstanding or disagreement. This can evoke an experience of being different and thus separated from the other, even though – in the same moment – I actually want to be connected to her.

The fact that the maintenance of human autonomy partly depends on social interaction processes that are not controlled by the individual brings thus about an ambiguity in the processes that realize human identity. The very means by which we are able to maintain our social identity seems at times to conform to this goal while at others failing to realize it. The strategies humans have to apply in order to remain social individuals can turn out to be in genuine tension to that goal. This stands in contrast to what we have learned about the preservation of metabolic identity. In order to stay alive the organism needs food; sugar, e.g., in case of the bacterium. The sugar is evaluated as a valuable source of energy needed in order to fuel the bacterium's metabolism. However, it is not that at times the consummation of sugar can fulfill this purpose while at others it does not. The sugar simply remains relevant for the metabolism and consuming it as the right strategy to ensure that the bacterium survives.

As a consequence socially enacted autonomy involves a particular kind of precariousness. In order for humans to maintain autonomy they need to ensure that two goals are respected, the goal to remain an individual that is distinct from others and the goal to remain an individual that is open and able to engage with others. This means that human autonomy requires balancing two tendencies that point into opposite directions. If we would interfere with the processes in the network such that the negotiation of these processes is affected, for instance by creating a critical imbalance towards either of the poles of participation or distinction, the network as a whole will be expected to encounter difficulties in the maintenance of its identity. But the deeper reason for why maintaining this balance takes place under precarious circumstances is that the processes by which we attempt to realize it depend on social interaction dynamics that can at times impede or even prevent us from realizing that very goal.

That humans are embedded in a world of other autonomous systems becomes therefore existentially relevant for their identity as a cognitive system. Others can to some extent decide whether or not they want to engage in a particular interaction and they can co-determine what quality this interaction can take.

As a consequence, precariousness in human autonomy entails that in order to maintain the network through the negotiation of processes of distinction and participation a readiness on

[84] The German system theorist Niklas Luhmann (1992) went so far to say that individual agents play no role in the description of that dynamics. Humans cannot communicate – it is the communication itself that communicates (p. 31).

the part of others to contribute to this negotiation is required. Not only the individual has a say in the maintenance of its identity as system, other do, too.

7.2.4 A Consequence of Precariousness: Need for Recognition

As a consequence of a social perspective on precariousness human autonomous systems are considered as genuinely dependent on some kind of benevolence in supporting each other's development as individuals. For that reason I would like to suggest that in human autonomy there is an important linkage between precariousness and *recognition,* and accordingly also the lack thereof. To account for this linkage could be an important part of a future elaboration on socially enacted autonomy. At this point I will only provide some hints.

We find the present idea famously reflected in Hegel's account of self-consciousness in the *Phenomenology of Spirit* (1807/1986). According to Hegel, self-consciousness has two basic dimensions, the "an sich" and "für sich". On the one side it is a pure, immanent and absolute consciousness, it is for itself (*für sich*). On the other side it also a consciousness that is open towards the world, it thus exists for something outside itself (*an sich*). Hegel uses the "master-slave" dialectic of mutual inter-dependency to illustrate that these two aspects of consciousness are in ongoing tension. Roughly speaking, even though master and slave exist in their own right, *that* each can exist is dependent on the other in the sense that it requires each other's recognition. The master may have power over the slave, but the slave is appropriated with the world. Since the master cannot exist without being related to the world he thus needs to recognize the slave who has this relation (Hegel, 1807/1986, pp. 145–150).

Hegel's master-slave dialectic can be linked to the above-introduced tension between distinction-participation in a two-fold way. Humans strive to be related to the world of others and at same time being a proper subject on their own. In order to ensure the survival as an individual, they have to therefore acknowledge both aspects and strive to maintain them. Yet because distinction and participation are antithetic this striving creates a tension between them. To acknowledge the tendencies that ensure the individual's survival thus requires constantly negotiating the tension between them. This would be reflected by the struggle between master and slave (consciousness as absolute and consciousness as world-directed) that exists *within* each individual. It translates into the twofold intrinsic goal of social autonomy to be distinguished while being able to participate.

In addition, since social autonomy is considered as co-constructed, the negotiation of the tension can be achieved only together with other individuals. The tension within the individual is thus necessarily tied in with another, social tension, viz. between the individual and other subjects that themselves struggle to negotiate their own intra-individual tension. The master-slave struggle for recognition is therefore not only a struggle of the individual that needs to acknowledge both tendencies towards distinction and participation. It is also a struggle for recognition between individuals. Since they are dependent on each other in order to maintain their social identity they also need to mutually recognize that struggle in each other.

This idea is nicely expressed in the work of Jessica Benjamin who argues in line with Hegel:

> the other must be recognized as another subject in order for the self to fully experience his or her subjectivity in the other's presence. This means that we have a need for recognition and that we have a capacity to recognize others in return, thus making mutual recognition possible. (Benjamin, 1995, p. 30).

Processes of recognition are relevant for both dimensions of socially enacted autonomy – distinction and participation. For example, in order to evolve an individual identity the young

human being clearly needs to learn from a more experienced human that she is and, importantly, can be somebody on her own. However, a part of that is also that the child is able to engage in social interaction processes with others and to thus experience herself as an "able" participant. The task of anyone responsible for a child is thus to provide her with the resources she needs in order to maintain and negotiate the tension between processes of distinction and participation, which is the basis of her autonomy. Ideally, the child would be raised so that it comes to experience 1) *on her own:* that she is the source of her action and that she can choose and change her interactions with the world according to her own goals and 2) *in participation* with the social environment: that she is recognized as the author of actions, as somebody who is somebody in her own right and crucially, somebody who can engage with others and who can be engaged with. This two-fold structure of human autonomy is nicely mirrored in attention studies with children by Reddy (2003). She has shown that infants of about 2 months age already show a twofold awareness of being addressed as someone ("of self as an object of other's attention") and of the other as somebody that could be addressed ("of others as attending beings") (ibid., p. 397).

According to enactivism "life is need" (Barbaras, 2011, p. 93). In human life, however, this need is not only a need for food but also for recognition. From an enactive perspective the condition of precariousness is a fundamental characteristic of autonomy, so that the need for recognition would not only apply to an initial phase of development. Because human autonomy is embedded in a social world and only fully emerges as a result of interacting with it, it also remains subject to permanent perturbation by others. Recognition is vital throughout lifetime (Ikäheimo, 2009). In contrast to the early stages of development, the "maturing" individual autonomy might become more proficient with regards to the regulation of these processes, i.e. the mechanism of its identity constitution. We can develop skills so as to choose more flexibly who these others are, as we do not always depend on particular other human beings such as mother or other primary care givers. However, in the striving to get our own say in evaluating interactions and how we behave towards others we cannot escape the fact that we thereby have to consider the existence of others and that this will require an engagement and negotiation with them.

Processes of recognition can be enacted in various ways and occur in every social situation. We see them applied in everyday language and the use of pronouns such as "I" and "you" as well as in being addressed by a proper name. Whenever we say "I" or "you" we refer to ourselves or another person as a distinct someone and thus affirm that we or they are acknowledged as an identity (Stawarska, 2009, p. 172). Our personal names provide an "objective" label for others to recognize us as somebody even when we are not present. That prisoners are given numbers is in a way to prevent them of being recognized as being somebody free.

If human autonomy vitally relies on intersubjective engagement with others and crucially, these engagements require some form of support by others (through, e.g., recognition) then human autonomy is principally vulnerable to potential disturbances. Others can – intending it or not – fail or refuse to contribute to a person's identity affirmation. If, e.g., a person constantly faces a lack of recognition as a distinct and independent entity, or as somebody to be engaged with, then this could ultimately interfere with the very organizational network that constitutes human autonomy. Particular interactions (or the lack thereof) would lead to problems, either with regards to the individual's experience *as* somebody individual or with her experiences of being somebody that is *connected* with others.

An empirical example to support the hypothesized relation between human precariousness and the fundamental role of social recognition is seen in studies of social pain using the cyberball game. Eisenberger has shown that the dreadful experience of social rejection (being excluded from participating in the game) leads to the same activation of neuronal circuitry as in physical pain (in reaction to increased temperature for instance) (Eisenberger, 2011). According to Eisenberger, this explains why social pain can reportedly have the same distressing quality as actual bodily pain. Pain of rejection has an important evolutionary function. Humans

genuinely rely on "social connection" in order to ensure their survival. For that reason they have to avoid situations where they are separated. A social rejection indicates the danger of being separated from others, and is thus ultimately a threat to "survival" (ibid., p. 587). That social rejection hurts is therefore evolutionary beneficial as it helps to avoid risking further (life) threatening situations in which we find ourselves separated from others.

This chimes well with the proposed characterization of human autonomy and precariousness in terms of a tension between distinction and participation. Assuming that human autonomy vitally relies on the recognition of others and the potentiality of engaging with them, can explain why social encounters involving explicit rejection or indifference lead to negative experiences or even severe pain. Social rejection is a threat to that dimension of our identity, which strives towards being connected with others (participation). If this dimension was a vital aspect of our identity, then this threat constitutes at the same time a threat to the survival of identity.

The vital role of recognition may become more evident in situations where we feel ignored or overseen. In case of explicit rejection or social exclusion we are still addressed as somebody. This is identity-affirming, even if in a negative way. To reject someone bears at least the possibility that the person could be somebody to engage with. To ignore somebody means to deny that the other person (even though she is present) exists as a potential participant in a social interaction. The matter is not to choose whether or not the person could be engaged with, but rather an absence of a choice at all. In the experience of indifference one's identity is threatened to an even deeper extent because it implies a denial of one's existence as another social being. One is neither acknowledged as a distinct individual nor as someone who could be interacted with.

Note that this does not mean that the integrity of autonomy would have to be at stake in every single interaction or with every other encounter throughout its life. The principled dependence on others hinted at in the suggested approach to precariousness rather translates to a vital need for an "update" (as similarly argued with regards to embodiment in the previous Sections 4.2, 5.1 and 6.5). Human autonomy is not a given or a fixed entity, but requires some confrontation with others. This is not to rule out that, to be and to experience one's identity as distinguished also involves phases of retreat into the "privacy of consciousness" (Humphrey, 2007, p. 751), in which we separately enact our "individual environment" and can be alone. However, even in the absence of an actual shared world, we continue to be directed towards it, through, e.g., imagining possible dialogues or encounters (Hermans et al., 1992). This is because we are individuals only against the background of our existence amongst others. If this was not the case, solitary confinement would not be chosen as one of the harshest punishments. As studies with prisoners have shown, social isolation can lead to serious psychiatric disturbances such as paranoia and hallucinations (Grassian, 1983).

7.3 Towards an Enactive Approach to the Self

In this section I suggest that the approach to human autonomy envisioned in this chapter can serve as the basis for developing an enactive perspective on the self. This will allow us to re-appraise the basic challenge from Chapter 3 to account for the relation of body and self and assess the role of facilitating technologies such as BCI for the LIS-patient's life (Section 7.5).

The concept of self has been and still is highly disputed, and models and conceptions of the self are diverse. The most common assumption among these is that the self is what makes us distinct from others and the individuals that we are. But consider the following brief list of other assumptions about the nature of self: According to Strawson (1999), the self refers to the sense of being someone conscious who has experiences. Dennett describes it as a "center of narrative gravity" (Dennett, 1992). Dainton (2004) argued that selves are owned by systems that produce unified streams of consciousness. For Hume the self was a bundle of perceptions (1739) and

James described it as the experiencing knower (1890). Zahavi has recently suggested that there are several phenomenological selves, the minimal, narrative and intersubjective (Zahavi, 2005, 2010). Perlis (1999) assumed the self to be a function and identical to consciousness, Tani (1998) conceived of it as an open dynamical system, Hayward (1998) approached the self in terms of dharmas and for Metzinger (2004) the self is nothing but illusionary. People thus speak of different notions of the self: personal, social, biological, self-affective, minimal, phenomenal, neuronal based, and unconscious. Some assume it to be a substance, a thing, a concept, a narrative, a system or a function. And even others argue that there is no such thing as the self. This list is clearly not exhaustive but should suffice to make a point: there seemingly exists nothing like *the* concept of the self. Unfortunately one cannot understand the self or discuss a problem of the self, when there is no agreement about what the self is to begin with. As Olson has recently put it, the problem with the diversity of proposals arises because "philosophers believe in a concept that doesn't exist" (Olson, 1998, p.645). As long as there is no account of the self that is more basic than all of the suggestions above, and could thus integrate the various aspects they ascribe to it, Olson says, we are better off not using the word at all.

The motivation for the proposal of a social approach to autonomy is directly tied in with Olson's verdict and his implicit call for clarification:

> What we need is not just an account of self that would command wider assent than any of these, but one that would synthesize them and show them all to reflect a part of some larger, common idea (Olson, 1998, p. 651).

My suggestion is that the enactive framework can in principle offer the basis to develop such an integrated and synthesizing concept of the self. I have outlined this idea in a recent paper (Kyselo, 2011), in which I discuss philosophical work that is directly concerned with the impact of Brain-Computer Interfaces on the LIS-patient's self (Fenton and Alpert, 2008, Walter, 2009). Fenton and Alpert suggest that on adopting extended cognition as a epistemological background, BCI's could be seen as an extension of the LIS-patient's self (see Section 4.3). Walter had criticized their proposal and suggested that the better framework to pursue this task would be the enactive approach.

I believe this discussion illustrates Olson's point: The authors are pondering about a concrete "problem of the self", i.e. the relation of self to body and environmental components (such as BCI) while taking for granted that the concepts at use are clear and coherently shared among themselves as well as the reader. I believe, however, that such questions remain hard to address unless both extended cognition and enactivism provide a sufficiently articulated concept of the self. I have derived the following list of constraints that a better candidate for the notion of self would have to accommodate. I take it that the above proposal for an approach to human autonomy as socially enacted based can begin to account for them (adapted from Kyselo, 2011, msp. 6):

1) A Definition of Self.
2) Determining the constitutive processes and interactive dynamics making up the self.
3) Understanding the structures and conditions that realize self-constituting and interactive processes.
4) The role of consciousness and the subjective experience for having or being a self.

It is not my intention to provide an elaborated model of the enactive self. This will be the task for the future. But by way of summarizing the suggestions made for autonomy at the level of social human beings, I would like to at least provide some hints for it.

1) In accordance with the basic definition of autonomy, the *self* might be generally *defined* as: an autonomous system composed of several processes that actively generate and sustain an identity under precarious circumstances. These processes necessarily imply social interaction

processes that show tendencies towards participation with or distinction from the social environment. The self is no thing, no mechanism or function, and it is not illusionary – it is an organizational property of human beings.

2) Precariousness means that the generation and sustainment of identity requires that both types of processes are connected such that they enable each other. *Self-constituting processes* are all actions and interactions that contribute to both the emergence of processes of distinction and participation as well as the negotiation of the tension that exists between them.

3) From this perspective, autonomous systems that are described at other levels of organization are now seen as contributing to the realizing structures of processes of distinction and participation. Their role in the organization is no longer defined solely in terms of their properties as individual autonomous systems, but also by the kinds of interrelations, i.e. processes of the new network that they now realize through their interactions.

Single neurons for instance can be described as autonomous systems (because they are living cells) and groups of individual neurons can form another level of autonomous system, a cell assembly that emerges for instance through so-called transient phase-locking during a particular cognitive activity (Varela, 1995). These cell assemblies in turn make up an even more complex level of autonomous system – the brain[85]. Shifting levels of autonomy we can observe a change in the particular domain as well as the organizational principle. For example, the particular mechanism constituting an individual neuronal cell is explained in terms of the metabolism and of particular biochemical processes. But the organizational principle of the cell assembly remains under-determined by considering only the properties of individual neurons (even though it can only be fully understood against their background). For an analysis of the neuronal cell assembly we rather transcend the metabolic level of description and describe its identity, e.g. in terms of dynamic interactions of neuronal processes.

The same applies to the organizational constitution of the brain as autonomous system in comparison to that of the autonomous body. We might describe the constitution of the brain in terms of several interacting neuronal networks, but when trying to assess the constitutive mechanism of a body in action we will refer to a description of autonomy in terms of neuronal networks only with regards to their role in realizing processes at a higher level of complexity, such as whole body sense-making interactions with the environment.

I suggest adopting a similar strategy in order to understand the relationship between the body and self. As I have tried to show in the discussion of LIS and BCI communication, we will not understand what it means to be a cognitive system that is embedded in a social world by only considering that it is a self-affective embodied individual. An organismic identity is based on "a self-isolation that accomplishes itself" (Barbaras, 2011, p. 93). But in order to do justice to the fact that we constantly have to negotiate our existence in a world of other subjects, a shift of perspective is required. The self from an enactive perspective is not accomplished by itself. The most proximate level of describing the realizing structures of the self as autonomous system should not be that of the individual body, but that of social interaction dynamics. Though it would clearly rely on bodily structures, the role of these structures is re-assessed: they contribute to the realization of socially enacted processes of distinction and participation. Our bodies are social.

4) From a phenomenological perspective these processes have been described as referring to self-experience as being someone distinct from others and someone who is engaged with others, i.e. that others can engage with.

It is not my intention to give an enactive account for human consciousness but I would like to suggest that the envisioned approach of the self would have some relevant implications for it. These could ultimately lead to a different assessment of consciousness in TLIS.

---

[85] Note that even the brain is not fully determined in terms of its material boundaries. Its functioning depends on it being connected to a variety of other networks that make up the living body, such as the systems regulating hormone and immune reactions (Thompson and Cosmelli, in press).

Let me briefly explain what I generally mean by 'consciousness' in the present context. In a recent article Thompson and Cosmelli argue that in explaining consciousness we have to distinguish between two notions, *state* consciousness and *creature* consciousness (Thompson and Cosmelli, in press). Whereas the former refers to particular states of phenomenally conscious experience, the latter refers to the creature that can have these experiences, viz. its capacity of having phenomenality at all. Studies about state consciousness involve thus *particular* experiences of the subject, whereas studies of creature consciousness focus on the "contrast between phenomenal consciousness and its absence (under anaesthesia or during coma) or between different global or background states of consciousness (such as wakefulness or dreaming)" (ibid., ms. p. 6). Thompson and Cosmelli then argue that in order to understand what role brain or body play for consciousness, cognitive science should not focus on state consciousness. This would miss that a subject already possesses phenomenal consciousness as a global background property. The focus should be rather on creature consciousness, and the explanation for "why the creature is conscious at all" (ibid., ms. p. 22).

In Chapter 2 we learned that enactive cognitive science is concerned with subjectivity as a general property of cognitive systems and not with particular phenomenal states. And according to Thompson, the task of enactive cognitive science is to understand how autonomy and subjectivity in that general sense are linked (2.3.1). When Thompson and Cosmelli shed light on the nature of creature consciousness then I take this to clearly contribute to this task. Therefore, when I here suggest reconsidering human consciousness in light of socially enacted autonomy I mean consciousness in the sense of a general property of human beings.

Consider now what creature consciousness according to Thompson and Cosmelli is based on:

> The enactive hypothesis is that *the minimal biological realizing system for creature consciousness is not the brain (or some neural subsystem) but an organism, understood as a self-sustaining system composed of some crucial set of dynamically entangled neuronal and extraneuronal subsystems* (Thompson and Cosmelli, in press, ms. p. 18, original emphasis).

The envisioned approach to human autonomy as socially enacted is directly linked to this hypothesis. I suggested that at the level of humans the processes required to generate and maintain the organizational autonomous network necessarily involve social processes. This would alter the view of consciousness as a general property at that level of autonomy and elaborate on Thompson and Cosmelli's perspective on creature consciousness: Human phenomenality arises not solely within the biological domain (of the brain in connection with the individual organism). It would also arise in the result of the individual's *social* engagement with other autonomous systems. As a consequence, the system's minimal supervenience basis for human consciousness would involve not only extra-neuronal, but also *extra-organismic*, social processes thus resisting an account strictly limited to the biological domain. In this view, consciousness is indeed a "life-regulation process", yet it would be a processes that is realized in a socially enacted dynamics, and not by the living body on its own, as Thompson and Cosmelli have it for the more general case of creature consciousness (ibid., ms. p. 25). As Zahavi says, "subjectivity and world are internally related, and since the structure of this world contains essential references to others, subjectivity cannot be understood except as inhabiting a world that it necessarily shares with others" (Zahavi, 2005, 167).

I believe that in these considerations we find reflections for yet another version of the classical body-mind problem. We recently learned that the problem is actually a body-body problem that concerns the relation between the body as object and the body as subjectively experienced (Section 2.3.1). We then saw that from an enactive (and phenomenological) perspective this problem is resolved by assuming that the material body and the body as experienced are integrated in world-directed existence (Hanna und Thompson, 2003, Thompson, 2007, Fuchs, 2011).

In the context of LIS and BCI communication we saw that the body-body problem is not fully solved by assuming that the material body and the subjective body are aspects integrated in the existence of an *individual body* directed at the world (see Sections 5.2.3, 6.5.2, 6.5.4). This is because from the perspective outlined in this thesis, the classical body-mind problem translates not merely to a body-body problem but rather to a body-body-*social* problem. We have to acknowledge that the existence of an individual living being is not independent from the existence of other subjects and how that affects the body.

The above phenomenological observations are therefore also expressions of how bodily self-experience in humans is shaped by being embedded in a social context. That I begin to feel nervous because I might be inappropriately dressed is based on my encounter with the stranger and the critical looks she gives me. That I experience fear and excitement before a conference talk is because I consider the recognition of others to be a confirmation of being a good academic and I anticipate their reaction to my presentation. That falling in love with somebody can involve feelings of lightness and joy is because I experience a resonance and harmony in interaction with the partner. None of these experiences can be fully understood on individualistic grounds – just like the case of BCI communication has dramatically shown, they imply a vital role of others.

For this reason we should expand the original question by asking how it is that being embedded in a social world shapes a conscious living human body.

One avenue towards addressing the body-body-social problem is to explore the linkage of phenomenal consciousness and the social to normativity, in the enactive sense. Again, I neither attempt to provide a full account of that linkage nor to defend a particular position on it. Here I would like to illustrate once again how an enactive conceptual linkage might change in view of a social perspective on autonomy.

We learned earlier that with the emergence of autonomy cognitive systems create a viewpoint on the world from which interactions with the environment acquire subjective meaning. With regards to understanding socially enacted autonomy, it would be therefore equally crucial that it always acts with respect to its own subjective experience and intrinsic norms, which are based on its ultimate goal to preserve itself. Based on its particular organizational structure the system not only strives to maintain that organizational structure but also adaptively generates further norms that are, though not immediately but ultimately, linked to its survival (see Sections 2.2.2 and 2.2.3).

But since human autonomy is considered in terms of an organizational network that combines constructive and behavioral processes of social interaction the system's basic concern reflects both our need to be distinguished from the social environment while at the same time striving to engage with it. The most basic goal of a socially enacted autonomous system would thus be to remain itself, while also being in relations with others.

Now if consciousness in an autonomous system is essentially connected to the perspective from which this system evaluates its encounters with the world, then the linkage to normativity, and through it, the link to sociality, is clear. Conscious experiences are a means to evaluate how autonomous systems fare with regards to their basic norms. Accordingly, consciousness at the level of human beings might at some point have to "tell us" something about how we fare with regards to our goal of being an individual that is at the same time a social individual. One aspect of human consciousness thus involves a two-fold structure, phenomenality of being a distinct social individual and phenomenality of being a social individual that is also a participant and connected to others.

Humans might have and, importantly, might also need to have a sense of identity and independence from others, but this independence must not also hold at the level of the minimal supervenience base of human consciousness. Being the conscious subject, it might feel to us as if our experience emerges solely within ourselves (within our heads, or even bodies, for that

matter).[86] But though it may seem less obvious for many people in Western culture, phenomenology of human self-experience is not merely individualistic (Stawarska, 2009, Schütz and Luckmann, 2003, Gurwitsch, 1964). Humans also have explicit self-awareness of the fact that they are participants in a social interaction and strive to engage with others. We have seen examples for this in the introduction to this chapter. It seems however that many people experience the participatory dimension of identity more clearly in contexts of a negative valence (such as rejection or social isolation) in which their identity as a social being is potentially endangered.

I do not suggest that all forms of human consciousness must be considered against the background of social interaction. The implication of the hypothesis outlined in this thesis is that the social dimension is an overlooked aspect of consciousness. If human autonomy is in part co-constructed then this should be considered for explaining at least some of our conscious bodily experience in hetero- and self-affection and their interrelation (see Section 6.1.2). This is reflected by Mead's perspective on the relation of body and consciousness. Without considering human consciousness against the background of *social experience* the "individual experiences his body…merely as an immediate part of his environment, not as his own, not in terms of self-consciousness" (Mead, 1934, p. 172). Even though it is possible at times to adopt the perspective of the living body itself to describe a felt sense of complete immanence, this does not mean that also at the organizational level the body is entirely independent of the world (see Section 6.1.2). Both dimensions are integrated in an engagement with the world that is socially informed. The body is both "an imprint of its natural and social environment" (Azevedo, 2006, p. 331). In this vein, the earlier discussed painful experience of rejection or lack of recognition would serve as an example of a phenomenal bodily experience by which we evaluate a particular (interactive) situation as a violation of our intrinsic goal to be able to engage with other human beings. Linking normativity to phenomenal consciousness based on a social perspective thus promises to also give an answer to *why* particular bodily experiences feel the way they do.

I consider the issue of consciousness in Section 7.5.4 in the context of the challenge raised by Kurthen et al. (1991) that there exists no epistemology to account for consciousness in complete LIS. I also come back to it in the next chapter, where I suggest a similar avenue of assessment for the phenomenon of bodily alienation in schizophrenia.

7.4 Related Approaches to the Self

The development of an enactive approach to the self obviously requires relating it to other approaches to the self that are similarly based on combining the individual and social dimensions. It is not my intention to provide a full survey of the existing literature that would support the enactive approach to the self. But I would like at least to briefly hint at some additional conceptual parallels and thus possible directions that such elaboration could take.

Firstly, we find the idea that the identity and cognitive capacities of the individual human being are shaped by social interaction processes reflected in the work of American philosopher and psychologist G. H. Mead who was at the forefront of the school of so-called "symbolic interactionism". According to Mead the human self never exists independently from the social interactions with others. As he writes in *Mind, Self and Society*:

> The self is not something that exists first and then enters into relationship with others, but it is, so to speak, an eddy in the social current and so still a part of the current. It is a process in which the individual is continually adjusting himself in advance to the situation to which he belongs, and reacting back on it (Mead, 1934, p. 182).

[86] In Hegel's dialectic this would refer to a pure self-consciousness that is not aware of its dependence on the self-consciousness that is directed towards the world. It would be a master that does not acknowledge that he is dependent on the slave in that through him he is related to the world (Hegel, 1807/1986, pp. 150–155).

To be a self consists in being able to become an object for oneself, i.e. being able to position oneself towards oneself as another person could do. This requires engaging with others (usually mediated by communication), to share their meaning and to internalize their attitudes. The self is co-constituted by the "social process of influencing others in a social act and then taking the attitude of the others aroused by the stimulus, and the reacting in turn to this response" (ibid, p. 172).

These considerations could be interpreted in enactive terms in that the self is both co-constructed by social interactions and at the same time standing out as an autonomous whole and engaging with the social environment continuously as that whole. Already in the work of Mead we might notice a particular type of closure: Human individual identity constructively relies on the very operations that are involved in its sense-making and interaction with others.

Another similarity is found in the work of Russian psychologist Lev Vygotsky. According to him, children develop their cognitive capacities in result of language-mediated interactions with primary caregivers. They learn the meaning of actions and objects through *shared* action. The language that mediates these shared interactions becomes then internalized and transformed into inner speech, i.e. actual thinking. This will then enable the child to regulate her own actions and interactions with the world:

> Thought development is determined by language, i.e. by the linguistic tools of thought and by the socialcultural experience of the child. Essentially, the development of inner speech depends on outside factors; the development of logic in the child…is a direct function of his socialized speech (Vygotsky, 1986, p. 94)

Similar to Mead's criticism that human cognitive capacities cannot be fully grasped from the viewpoint of physiology, Vygotsky encourages us to transcend "the limits of natural science" and consider cognition from a *socio-psychological* perspective (Vygotsky, 1986, p. 95). This would reflect the basic upshot of the present proposal: the human mind does not arise in isolated beings, but in embodied and social beings that interact and relate to each other.

Vygotsky's idea that cognition is based on the internalization of linguistic processes mediated in social interaction also resonates with the recent "dialogical self approach" put forward by Hermans and colleagues (1992). According to them we should accommodate the self from a constructionist rather than an individualistic and rationalistic perspective. The self, in this view, is a "diaological narrator", which is on the one hand organized spatially, in terms of the organismic embodiment, and on the other socially organized via a variety of several "I positions", i.e. internalized voices of others. For Hermans the dialogical self is social "not in the sense that self-contained individual enters into social interactions with other people outside, but in the sense that other people occupy positions in the multivoiced self" (Hermans et al., 1992, p. 29).

What could be especially interesting regarding the context of LIS is that according to Hermans et al., a crucial component of our understanding of the world consists in having imaginary dialogues in which "we find ourselves communicating with" others. Similarly to Vygotsky, extra-individual social interactions are internalized and form the basis of the subject's perspective on the world. The self relies on an "imaginary social world rather than a purely inner world" (Hermans et al., 1992, p. 28). This supports the idea that even when we are not actually engaging in social interactions with others we still 'engage' with them in our imagination.

In the following and last section I return to the challenges from Chapters 3, 5 and 6 and indicate how the conception of human autonomy as socially enacted could help to address them.

7.5 An Enactive Approach to the Self and LIS

Recall the several claims made about the relation between body, self and the social (see Section 3.2.4). On the one hand, it was argued that body and self are intimately related, if not identical.

The self was also described as embodied. On this view, bodily impairment would directly lead to an effect on the processes that sustain the self. On the other hand bodily capacities were seen as bearing a social purpose and it was argued that bodily impairment would directly affect the social capacities and thus the subject's social skills. In this section I show that and how both variants of interrelating self, body and social are linked.

7.5.1 Body and Self are not Identical

A first consequence of the above proposal is that the human body and self would no longer be conceived of as identical. They rather refer to two different types of autonomy and they interrelate in that the body (including the brain) is assessed in virtue of its role for realizing the constructive mechanism of the self as a socially enacted type of autonomous system.

Based on this we can provide a way to address the refined challenge from Chapters 5 and 6: "How can we account for the role of embodiment in interactions with the social environment?". The body can play a role for realizing social interaction processes with both tendencies – towards participation and distinction from the world. From this perspective the body is not identified with the self because the self is co-constructed in a dynamic interaction process with other subjects. The body is the means to enact this social co-construction of the self. An implication of that is that bodily impairment like LIS is not necessarily an impairment of the self. It affects the self insofar as it restricts the person's means to engage in those social interaction dynamics from which a self must emerge.

7.5.2 The Role of BCIs in Light of the Enactive Self

Let me now return to the context of LIS and re-examine in terms of socially enacted autonomy, the status of BCI technology for the patient. If the body in part realizes processes that constitute the human being's autonomy, then a global bodily impairment such as LIS will affect that autonomy to the extent that the realization of the interwoven network of co-constitutive social processes is affected. Because the patient cannot or can hardly communicate, LIS makes it difficult to impossible for the patient to engage in the constitutive processes for the emergence of her socially enacted autonomy. She is close to other human beings but at the same time as distant as it gets, because her severe disability eliminates or radically reduces the possibilities to interact with them. This is a terrifying situation, because the patient is still embedded and directed to a socially structured environment and strives, according to the present proposal, to engage with it.

Herein lies the important role of Brain Computer Interfaces. BCIs do not extend the self as Fenton and Alpert have suggested (Fenton and Alpert, 2008), if "extended" only refers to the fact that the constitutive mechanism of the self is not found within brain or body alone. In view of the above proposal the self rather is – even prior to LIS – seen as already and necessarily directed to and based on social engagement with others. It *always* involves non-individual, non-organismic processes and thus is not suddenly extended by the use of a technology like BCI. From the present perspective the status of BCIs is explained in terms of their role as a substitute for severely reduced bodily action. By acting as a means for communication, BCIs can sustain a part of the constitutive dynamics of the self that has been impaired. Because it re-connects the patient with the social environment, the technology essentially secures the maintenance of its identity as a social being.

Using a tool such as BCI thus differs fundamentally from other instances of tool use, say for example of a hammer or a saw. While we use the latter to build something external and thus rely on them to shape our *material* environment, BCIs are used to sustain our (social) existence as such. They therefore contribute to maintaining the self as what we might call a social centre of activity only based on which we can at all engage with the social world.

I previously suggested that the BCI becomes incorporated because of the patient's experienced habitual acquaintance (see Section 5.2.3). Now we can elaborate this idea by saying that the development of a bodily habit relies on a dynamic social interaction process, it is not merely a matter of continuous bodily engagement of the individual.[87]

7.5.3 Locked-in Syndrome as a Social Injury

Coming back to Dudziniski's phenomenological inquiries on LIS's impact on the life of Bauby, I would argue that for similar reasons, and contrary to what Dudzinski has suggested, LIS does not affect the self directly (see Section 3.2.2). The condition rather impacts on bodily and socially mediated action, which are both necessary for the maintenance of the self. In the present view, a bodily impairment would be a reduction of the self to the extent that the patient's abilities to realize the co-construction of its identity are affected.

This is clearly reflected by approaches that consider bodily impairment as a "double injury", not only an impairment of *physiological* but also of *social* abilities (Gillett and Chisholm, 2007, p. 5, see Section 3.2.1). Since we are social beings and as such embedded in a world of others, we strive to show ourselves to others. Being able to regulate how this "showing to others" is realized is thereby of crucial importance. The expressive limitations of patients with Möbius syndrome and LIS are not merely bodily, but social injuries, with different degrees of severity. For this reason, bodily capacities might be said to have a social purpose or to have been "evolved for intersubjectivity" (Cole, p. 344).

Nevertheless, to determine bodily abilities with regards to their role in driving social interaction processes would only give us half of the picture. According to the above suggestions the dynamics of distinction and participation involve other people. LIS and other social injuries should thus be accounted for by accommodating the role of others in maintaining these processes and negotiating the tension between them.

Based on this we can finally begin to make sense of the third group of claims focusing on the question how bodily impairment affects how others perceive a patient and hence alter their particular behavior and the reaction towards her (see Section 3.2.3). If a person changes her attitude towards the patient she thereby changes her participatory sense-making engagements with her. There is a chance that this affects the patient's self because shared sense-making contributes vitally to the maintenance of the patient's identity as a social being. Bodily impairment is thus a social injury not only with regards to the patient's capacities to engage socially, but it is also the potential cause of alterations in the attitude of the other and thus in the dynamics of the social interaction processes vital for the patient's self.

I want to make clear again that the present proposal is not merely to state that the social world is important for patients because, e.g., the patient still cares about other people or others could help her with problems or realization of actions. It is not just to say that the existence of other people has some effect on the patient's well-being. My claim is much stronger: the social world is essentially important for the human self because there are certain social interaction processes that are vital for the existence of a human self *as such* and not merely for some of its dimensions or particular experienced qualities of being that human self.

7.5.4 An Enactive Approach to Consciousness in TLIS

In Section 3.3 I discussed Kurthen et al.'s conviction that there is no current epistemology to account for consciousness in patients with TLIS and if there was one that could do this job, then

[87] Havelange has recently made a similar suggestion for the use of TACTOS, a device that helps blind people to communicate via a tactile interface (Gapenne et al., 2003). According to Havelange, the technology becomes incorporated into the lived body as a social body "to the very extent that it contributes to the constitution of a common world" (Havelange, 2011, p. 354)

it would have to involve a complete neuroscientific image of man. As this is utopian, out of pragmatic reasoning, one should treat patients with TLIS as if they were conscious.

In line with Thompson and Cosmelli, this problem can actually be rephrased as a question concerning the "contrast between phenomenal consciousness and its absence (under anaesthesia or during coma)" (Thompson and Cosmelli, in press, ms. p. 6). If, according to the above proposal the basis for creature consciousness in humans is neither to be found in the brain, nor in the individual organism alone, but also in the social realm, then assessment of LIS patients would consequently not be restricted to neurophysiological constraints. It would also acknowledge the restrictions of embodiment in light of their role for the intersubjective engagement of the patient.

This will of course not readily permit a decisive answer to the question of whether or not a TLIS patient is or remains conscious. But contrary to Kurthen et al.'s consideration, it would at least offer the beginnings of theory about what could happen to consciousness in patients with that condition. Because we are not only embodied, but also primarily social, the TLIS patient's identity would be affected at its the core: with regards to individual abilities (in terms of embodiment) as well as the possible social interaction processes she engages in with others and the possibilities of others to engage with her. Both social processes of distinction and of participation are in principle required for the emergence of her autonomy and thus for the patient's self-awareness as a social being. If the patient's continuous engagement is not only temporarily but also completely interrupted we might hypothesize that her phenomenal consciousness (as a global capacity) and thus the awareness of herself and of the environment would be deeply affected.

7.5.5 LIS and the Quality of Life

Let me now provide an example from empirical research on self-experience in LIS that illustrates that the present proposal can have direct practical consequences. In a recent paper, Nizzi and colleagues claim to make an advance against intuitive armchair philosophy by approaching real patients in the wheelchair and asking them for their subjective experience with LIS. They have conducted interviews with LIS patients in the classical state to assess how they have integrated bodily changes into their identity (Nizzi et al., 2012). More specifically they wanted to know 1) whether the patients feel the same as before their accident, 2) whether they recognize the altered body as their own and 3) if they positively respond to 1), that is, if they in fact experience their identity as continuous, what it is that ensures this kind of experience. The result was that with regards to question 1), CLIS patients mostly felt that they have "well-integrated" their paralysis and that they "feel the same as before the accident". According to Nizzi and colleagues this indicates that "such an *objective change*, in real life can be integrated in one's experienced identity" (ibid., p. 435). With respect to question 2), the authors found that patients mostly continued to recognize their altered body as their own. The authors assume the experience of feeling the same and the positive judgment about their bodily identity to strongly correlate – the fact that patients experience a feeling of continuous identity relies on the patient's acceptance of the massive bodily change that the accident has brought about. In reply to question 3) patients report that they experience their identity as continuous when their "life, no matter how *objectively different*, kept its subjective meaningfulness" (ibid. 436).

What is most interesting with regards to the present proposal is the authors' explanation and interpretation of the responses to 2) and 3). That patients report to experience their bodies as their own is explained as follows:

> This finding suggests that even if the *objective body* can undergo massive changes, the *body representation is the component of the experienced identity that matters* from the wheelchair (Nizzi et al., 2012, p. 435).

The assumed reason why the patients adjust to the paralysis is because they still experience their life as meaningful. From the authors' point of view this suggest that personal identity is not fixed, but "plastic" and that as long as a patient subjectively adjusts to an objective bodily change, she will still experience herself as being the same person. A positive bodily representation thus lives on, despite objective change, when subjects accept change:

> If real patients can adjust to massive bodily changes because of a psychological meaningfulness in their life, then personal identity might be as plastic as brain plasticity: no matter how important the objective bodily change is as long as the patient acknowledges it as meaningful (Nizzi et al., 2012, p. 436).

Even though the authors emphasize that philosophers should not merely speculate about what it is like to be locked-in, but ask the patients themselves, I think that problems lie exactly in the way that Nizzi and colleagues theorize about the answers they got from the patients. Their interpretation seems to imply a view of the relation of body and subjective experience according to which the body is a material object, a physical envelope inhabited by the patient. Whether or not this body is subject to severe objective change, as the authors have it, plays no role for the patient as long as she consciously *decides* that it does not play a role because she has a meaningful life. It also seems to suggest that physiological change has absolutely nothing to do with subjective experience. In addition, the authors' approach could be interpreted as saying that a patient's life and her identity are entirely a matter of her own individual existence. In consequence of this, Nizzi et al.'s considerations entail not only the bemoaned split between objective and subjective body (see Section 5.1.1), but also run the risk of maintaining a split separating the individual system from the sphere of others.

Patients, who are crucially able to live at home with their families, say that their life is meaningful, but what the authors did not ask them is *why* this is the case. Is their life really meaningful just because they have *decided to accept* that their bodies are different now? In light of the present proposal this seems not only highly unlikely but also contradicts what other reports of the quality of life arguably indicate (see Section 3.1), namely that patients feel well-adjusted and not significantly different from normal subjects because of a "subjective feeling of control over one's life" as well as the *social support* of others (Lulé et al., 2009, p. 344) and that "quality of life often equates with *social rather than physical* interaction" (Gosseries et al. 2009, p.199, my italics).

My interpretation of the study conducted by Nizzi et al. resonates the tenor of the present thesis that phenomenological inquiries are in need of amendment by an integrated organizational perspective. I do not say that the patients' positive evaluation of the quality of their life is wrong. But their positive judgment alone is not yet helping us to understand human identity or the role of the body in it. It is neither a justification for assuming that what is at stake in cases of global impairment is primarily the patient's self as a *bodily* identity.

If the physiological body is considered as the only relevant factor in the constitution of the patients' experienced identity we might miss that it may be because others are still engaging and communicating with them that the patients are actually able to adapt to the new situation. Following the envisioned approach to the enactive self, we could explain why even an extreme change in physiology is not experienced as a rupture from their identity. As long as patients are able to engage in social interactions (communication for instance) they ensure the maintenance and negotiation of the self-constituting processes of distinction and participation. Because of this and regardless of their bodily impairment they experience their identity as unaffected, which essentially equates to experiencing themselves as somebody that is both recognized as a distinct individual that at the same time can be addressed and engaged with. I believe it would be an important task to investigate this issue in future research on the role of the social for embodiment and cognition.

7.6 Concluding Remarks

In this chapter I made a proposal for an enactive account of the human mind in terms of an autonomous system that can be defined in terms of a precarious autonomous network in which behavioral and constructive processes overlap. More specifically these processes are defined as socially enacted processes that show tendencies towards a distinction from and a participation with the social environment. I proposed that precariousness at that level consists in the fact that these processes are organized in a co-dependent way and that both the maintenance of this organizational interrelation and its negotiation involve extra-individual social means.

I have then outlined how this envisioned elaboration of the concept of autonomy can be a starting point for developing an enactive perspective on the self. I have supported this exploration in the final assessment of the challenges derived from the previous discussion of LIS and by illustrating how conceiving of the self as a socially enacted autonomous system would help to approach the pressing question of how body, self and sociality are interrelated. The general key to resolving body-body problem is to integrate the material and living body in a world-directed existence (see Section 2.3.1). The body-body-social problem can be resolved on assuming firstly, that the body entails not only a subjective dimension referring to the external world of objects. It also requires a pre-reflective reference to itself, viz. a "non-object-directed awareness of oneself as the *perspective-holder*" and thus a bodily consciousness of being an individual subject separated from the world (Thompson and Henry, 2011, p. 247, my italics).

However, to integrate the material body with the two dimensions of the living body with sociality requires also considering that holding a perspective on the world in case of humans is co-constructed in social interaction. The human cognitive system as a whole should thus not be equated with bodily identity but considered as a socially enacted autonomy (the self). As a consequence the integration takes place not in the embodied interaction of the individual but in an intersubjectively enacted dynamics for which the body takes a mediating role. The two dimensions of bodily subjectivity (hetero- and self-affection) are therefore fundamentally affected by the fact that being human is being a social individual. Bodily experiences let us evaluate not only engagements with the world of objects but also with other subjects.

The present considerations do not constitute a fully defended position. They are a first proposal providing avenues for exploring how the perspective on human autonomy can change on acknowledging the vital role of social interactions. There are many issues in this proposal that need elaboration and that will be research topics for the immediate future. One general task is to spell out more concretely how social interaction processes figure in both the development as well as the sustainment of individual autonomy throughout life. This will imply elaborating and explicating the two basic kinds of processes, i.e. of distinction and participation at multiple levels of descriptions and both from an organizational and an experiential perspective. It would be of particular interest to work out how autonomy develops at the embodied level, by which I mean bodies in intersubjective engagement. There are several (social) contexts in which this could be done. One could be the initial development of socially enacted autonomy, i.e. how newborns and young infants first develop a socially enacted autonomy in interaction with their primary care givers. Focusing on the first form of interaction that humans encounter in early childhood may help to identify processes that are at play in human autonomy through life. Another could be the neuromedical context. One might conduct empirical research with LIS patients and deepen the attempts to investigate the patients self-experience in the direction of the present proposal: In contrast to the research by Nizzi and colleagues this would involve not only a description of the patient's self-experience but also an attempt to explain *why* the patients experience their identity as continuous despite global bodily impairment. For that matter one could take seriously a social perspective on cognition and ask patients questions concerning their existence as a social individual. One could then investigate whether or not the experienced continuity of identity is actually related to the fact that humans are socially embedded, not merely embodied. This could

be relevant for practical and ethical purpose of improving the quality of life in patients with paralysis.

How exactly it is that humans become autonomous and are affected by interaction with others requires a detailed elaboration on the interplay of the constitutive processes of distinction and participation. It probably needs to be assessed on a case-by-case basis and against a particular social-cultural background

In the next and final chapter I focus on socially enacted autonomy at later stages of development. To support plausibility of the present proposal, I show what implications it could have for our understanding of mental disorders in general, and schizophrenia in particular.

Chapter 8.

Disorders of the Self In the Light of the Enactivism – A New Look at Schizophrenia

> [E]mpathy and social understanding are the precondition for any science of mind and brain. (Fuchs, 2011)
>
> The patient is ill, that means, his world is ill. (van den Berg, 1972)

In the last chapter I have presented a proposal for an enactive approach to the self. I suggested that the self might coincide with an approach to human autonomy as partially constituted by social interaction processes. This is a proposition for a new look at human beings as not only embodied but also primarily social. We have seen how this proposition can shed light on our understanding of the impact of bodily impairment and the role of BCI in LIS. In this chapter I would like to provide further support for this proposal by applying it to the field of psychiatry and exploring enactive implications for psychopathology and in particular for schizophrenia.

The way in which we think about human cognition clearly affects how we approach mental illness, i.e. cases of cognition "gone wrong". Zoe Drayson (2009) has recently explored this linkage suggesting implications that a shift from orthodox to embodied cognitive science entails for our understanding of psychopathology. We have learned that orthodox cognitive science conceived of the mind as a computer and that the brain is the hardware on which the cognitive software runs. As a consequence, disorders of the mind have been associated with neuronal dysfunctions – they are "bugs in the software", as Drayson puts it (Drayson, 2009, p. 337). The benefit of conceiving of mental illness in that way was that psychiatry could explain mental illness not only in terms of damaged brain areas but also account for the relation of the brain to psychological and social processes. These interrelations have been conceived of as input-output relations gone wrong due to neurological dysfunctions. According to Drayson embodied cognitive science can inform this perspective on mental illness because it additionally "can point to a neglect of the phenomenology of the disorder in question, or its biological basis, or to the temporal aspects which a computational account cannot capture" (ibid., p. 338). An example for this, Drayson suggests, is Gallagher's embodied approach to autism (Gallagher, 2006). Instead of assuming autism to involve a neurological dysfunction of the ability to theorize about other people's mental states Gallagher suggests that it might be explained in terms of disturbances located at the sensorimotor level of cognition. In Drayson's opinion this provides an innovation for our understanding of mental illness: "If sensory and motor processes are basic to all other cognition, as much research in embodied cognitive science posits, then disorders that have traditionally been viewed as dysfunctions of higher cognitive processes could in fact be explained by lower level sensorimotor processes (Drayson, 2009, p. 338)." In this vein she suggests replacing the orthodox model of psychopathology by a new embodied model, which "conceived of mental disorders as disorders of embodied brains embedded in their natural and social environments" (ibid.). Gallagher's approach to autism is an example how embodied cognitive science can elaborate on the biological basis of mental illness.

Recently we also find approaches to psychopathology that consider the phenomenological dimension of embodiment (Fuchs, 2005, Gallagher, 2006, Blanke and Metzinger, 2010, Parnas and Sass, 2010). Parnas and Sass for example propose an account of schizophrenia in terms of a *phenomenological* concept of "minimal self" as defended by Zahavi

(2005, 2010). According to these authors, schizophrenia should be conceived of as a disorder of "substantial sense of embodied subjectivity" (Parnas and Sass, p. 229). It can be associated with disturbances of the structures that constitute the normal minimal self: self-presence, first-person perspective and phenomenality.

In this chapter I would like to focus on the proposal by Parnas and Sass' because their approach illustrates deeper and more general issues that arise when adopting a merely embodied approach to psychopathology. Existing embodied explanations of psychopathology are not fully satisfactory for three interconnected reasons. First, they remain individualistic and do not do justice to the fact that we are not isolated subjects, but deeply embedded in a social environment, constantly having to negotiate our existence in a world of other subjects. Second, they seem to be limited to the material structures underlying mental illness, which in itself does not yet explain what the disorder is at its roots. Third, even though they consider the phenomenological dimension of embodiment in acknowledging the patient's bodily experience of mental illness, this is not sufficient to explain why bodily experiences matter for mental sanity in general or why pathological experiences of embodied subjectivity are problematic and cause suffering. My suggestion is that these issues could be resolved by adopting an alternative perspective on mental illness that is informed by Merleau-Ponty and specified in terms of the enactive approach to social autonomy that I have proposed in this thesis.

The chapter is structured as follows. First, I briefly introduce the concept of minimal self by Zahavi. I then summarize how Parnas and Sass explain schizophrenia as a disorder of the minimal self. In the second part I outline the explanatory shortcomings that come with that approach and why they are of general concern for embodied cognitive science. I will ground this criticism in Merleau-Ponty's account of illness as an expression of an underlying structure we already find in the healthy bodily subject as a whole. This approach involves a normative perspective missing in both orthodox and some embodied approaches to cognitive science but not in the enactive perspective.

For this reason, I suggest adopting an alternative approach to schizophrenia, which grounds its examination on the notion of self as socially enacted. From an enactive perspective the self is intrinsically goal-driven and fundamentally concerned with its own maintenance. The approach does not take for granted that schizophrenia reflects a 'malfunctioning' of the self. The basic idea is rather that mental illness is an expression of an extreme struggle to maintain the self but that the struggle essentially already exists in non-pathological cases. In terms of the concepts that I have proposed in the previous chapter, it is because we strive to keep a balance between the social processes of distinction and participation, and that this is not achieved independently from others, that we can become mentally ill when this balance goes wrong. In that vein, I suggest considering schizophrenia in terms of a dramatic and fundamental imbalance of distinction and participation. Based on this idea I offer in the final part of this chapter a different interpretation for a selection of symptoms in schizophrenia also considered by Parnas and Sass.

8.1 Disorders of the "Minimal Self" as the Basis of Schizophrenia

In this section I briefly summarize Zahavi's approach to the self and the three dimensions he describes as characteristic for the minimal self. I then show how Parnas and Sass examine schizophrenia as a disorder of the minimal self.

8.1.1 The Structure of the "Minimal Self"

According to Zahavi, in order to account for the complexity of the phenomena of the self its "various complementary accounts must be integrated" (Zahavi, 2010, p. 3). He has suggested differentiating between two basic notions. Firstly, the *minimal self* that refers to an embodied experience of a "for-me-ness" and is present in every other conscious experience. Secondly, the

narrative self that refers to the more sophisticated dimensions of the human self (Zahavi, 2005). More recently he has argued that these concepts are insufficient to offer an integrative account of the self. Such an account would require "even more notions of self" (Zahavi, 2010, p. 6). Zahavi begins to sketch a supplementation to minimal and narrative selves by proposing what he calls the "*interpersonal self*". The three concepts, minimal, narrative and interpersonal self are complementary dimensions of *the* self, "different aspects or facets of selfhood" (ibid., p. 6). The concepts of the *minimal* and *narrative self* are "placed at each end of the scale" (ibid.). They capture the embodied first-person perspective of selfhood on the one hand and the dimension of the self as situated in "culture and history" (ibid.) on the other.

Though Parnas and Sass acknowledge the interrelation of minimal and narrative self (Parnas and Sass, 2010, p. 230), in their analysis they solely rely on the former. They deliberately disregard "important developments pertaining (for example) to *intersubjectivity* and temporality" (ibid., p. 232, my italics). We will see that this decision seriously limits the explanatory power of their strategy.

Let me now repeat the three dimensions of the minimal self. *1) self-consciousness and self-presence,* which refer to the experience of "for-me-ness" accompanying all experiences. They are the basis of the ability to have experiences of objects and to have them as my own. This dimension of self-hood has also been referred to by the concept of *perspectival ownership* (Albahiri, 2007, pp. 53–54) and the concept of bodily *self-affection* by Henry (introduced in Section 5.3 and discussed in light of the enactive epistemology in Section 6.1.2, Zahavi, 2005, p. 106) as well as *core consciousness* by Damasio (2006). A basic sense of for-me-ness is also part of the second part of the minimal self, *2)* the *first-person perspective* in perception and thinking. In perception, the first-person perspective is a "function of the lived body, moving in space". It is the basis for a "*sense of self,* of being a substantial (spatial) subject" (Parnas and Sass, 2010, p. 234). In thinking, the first-person perspective does not imply such "*distance* between myself and my thoughts". Here "for-me-ness" is clearly experienced as part of experience itself. The third dimension of the minimal self is *3) Phenomenality*. It refers to consciousness of the self as an "ontological domain" and thus to a basic subjectivity (see Section 2.3.3). Thompson and Cosmelli have recently described this basic subjectivity as "creature consciousness", the basic capacity of being conscious at all. (Thompson and Cosmelli, in press, see Section 7.4.4). It is the "medium" and the "condition for the emergence of any phenomenal *structure* whatsoever" (Parnas and Sass, 2010, p. 237). Note that the minimal self is thus seen as a property of the "living organism" (ibid., p. 230). It is arguably present from birth (Krueger, 2011, p. 41) and also in animal consciousness (Parnas and Sass, 2010, p. 230).

8.1.2 Schizophrenia as a Disorder of the Minimal Self

Parnas and Sass propose that schizophrenia should be "associated with disorders *that affect the articulation and functioning of minimal or core self*" (ibid., p. 230). In schizophrenia the "normally smooth pre-reflective sense of self (and world) loses its automaticity and transparency" (ibid., p. 231). This means considering symptoms and subjective reports of the persons as referring to structural disturbances of these "*three axes* of normal self-experience", 1) self-presence, 2) first-person perspective and 3) phenomenality (ibid.). Schizophrenia is thus not *a* disorder of the self but rather a disturbance of the *several* "orders" that make up the minimal self. These disorders should not be considered as "independent, object-like features", or "independent symptoms", but rather "*aspects* of a single whole, a phenomenon, considered in its different facets" (ibid., p. 231).

As a consequence, the authors interpret e.g., the commonly reported experience of alienation, "a strong and disturbing feeling of being different from others" as a deviation from one's normal sense of self-presence or one's "existence as subjectivity" (ibid., p. 232). They then argue that the sense of alienation leads to social isolation, which is another acknowledged symptom of the schizophrenia spectrum.

Consider another example: so-called hyperreflexivity. Persons with hyperreflexivity often report the experience of a distance between themselves and their thoughts. They are "*having* thoughts" rather than "being" them (ibid., p. 236)[88]. Parnas and Sass explain this as an abnormality in the second dimension of the minimal self, the first-person perspective. Normally, the first-person perspective in thinking processes involves an immediate self-acquaintance or self-affection that is required (in line with Henry) by every other experience. This renders the several experiential contents "co-conscious" and integrates them into an "experiential whole" (ibid., p. 235). In schizophrenia, so the authors claim, this experiential unity of consciousness is experienced as ruptured. Thoughts are no longer grounded and "appear increasingly objectified and spatialized" (ibid., p. 236).

According to Parnas and Sass, hyperreflexivity can also be associated with disorders of the third dimension of the self, i.e. phenomenality. Since phenomenality underlies all other phenomenal manifestations a disorder of phenomenality manifests itself at more general level, not solely with regards to a particular experience. Usually we are "absorbed in the world" and this immediate acquaintance with the world is only "interrupted during occasional moments of introspective self-reflection". In schizophrenia, however, phenomenality loses this directedness to the world. It leads to a dramatically increased self-awareness and "*intensified self-consciousness*" (ibid., p. 239). The persons suffer from a "constant, reflective self-scrutiny" which make it impossible for her "to reach out and become immersed" in the world (ibid., p. 238). Experiences become generally more intense, including increased awareness of "bodily sensations, inner speech or the presuppositions of thinking". Aspects of selfhood that usually remain tacit and in the background, become "reified, spatial-like entities" (ibid., p. 239). Interestingly, persons with increased self-awareness also report feeling as if they have no or only "fragmented" self-consciousness (ibid., p. 240). Hyperreflexivity thus does not rule out that the person could experience a loss in the sense of self-presence, rather

> there is a sense in which the person with schizophrenia has *both* too little awareness of self (diminished self-presence or self-affection) *and* also too much self-consciousness (hyperreflexivity). This means that a certain heightening of phenomenality of self-consciousness (increased focal awareness of the "inner") is intimately bound up with a failure to experience the normally implicit foundations of self-presence, namely self-affection (that is, to *inhabit* one's kinaesthetic sensations or inner speech) (ibid., p. 239)

This arguably shows that hyperreflexivity and a disorder of self-presence are linked as "two aspects of the *same* process":

> Whereas the notion of hyperreflexivity emphasizes the way in which something normally tacit becomes focal and explicit, the notion of diminished self-affection emphasises…the fact that what once *was* tacit is no longer being inhabited as a medium of taken-for-granted selfhood. (ibid., p. 239)

One might say that in hyperreflexivity the person's ability to just exist is affected by a constant reflection on the conditions of her existence. To 'just exist' means to be able to more or less directly engage with the world. It does not involve a particular awareness of this engagement or what it is based on. But in hyperreflexivity the usually only tacitly experienced self-awareness becomes explicit and is thus "getting in the way". Yet because having a sense of continuous self-affection is also a prerequisite of our basic awareness of our self, the person may additionally report reduced experiences of self-consciousness.

[88] I agree with Thomas Fuchs when he says that this appears to be an odd observation as no one would probably claim to "be" his or her thought (personal communication). However, I interpret the authors as really only emphasizing the losing of the experience of being the owner of one's thoughts.

It is important, with regards to the present context, that Parnas and Sass consider the above disorders of the minimal self as "primary – in a causal/pathogenetic sense". They are the so-called "*trouble générateur*" of schizophrenia, which is

> not a consequence of other psychical disturbances, but an essential point [or state] *from which spring,* or at least *from which it is possible to view in a uniform way* all the cardinal symptoms (Minkowski in Parnas and Sass, 2010, p. 240, emphasis added)

To understand the disorders of the minimal self in this way would help to explain the symptoms in schizophrenia in two senses of explanation. Firstly, in that it shows how symptoms are different but "interdependent aspects of the same experiential whole involving a disorder of minimal selfhood" (ibid., p. 240). Secondly, it helps explaining the transition from one experience to another in that "certain symptoms of schizophrenia may be viewed as consequences, whereas others as compensatory, coping attempts" (ibid., p. 241)

In the following section I show with Merleau-Ponty that this strategy describes ways in which different symptoms are interrelated, but does not suffice to explain *why* they should interrelate and how they came about in the first place.

8.2 Schizophrenia in the Light of Merleau-Ponty's Account of Illness

I would like to preface my worries about Parnas and Sass' strategy with a consideration of Merleau-Ponty's account of 'psychic blindness' (Merleau-Ponty, 2002/1945, pp. 118–150). He considers the famous case of Schneider, an apraxia patient of neurologist Kurt Goldstein. Schneider displayed a 'dissociation' of two kinds of bodily movements. He was unable to perform *abstract* movements that are not directed at a particular situation, such as pointing at his nose, or just moving his limbs. He was however able to perform *concrete* actions, such as grasping a handkerchief or switching on the light.

8.2.1 Merleau-Ponty's Alternative Approach to Illness

According to Merleau-Ponty, this dissociation between grasping and pointing cannot be explained by saying that it indicates a disturbance of a singular kind of action, say of directed movement, for which we then have to find an explanation in terms of a physiological or psychological deficiency. Instead the dissociation must be approached from a "behavioural dimension" (Merleau-Ponty, 2002/1945, p. 143), from which the physiological and psychological perspective are merged in existence. From this perspective, a pointing movement appears not as a part of grasping movement, but as an entirely different kind of action. Abstract and concrete movements are "two ways of relating to the object and two types of being in the world" (ibid., p. 141). The inability to perform abstract movement, i.e. pointing, displays the lack of a *function*, it does not indicate either a psychological deficiency in terms of conscious awareness or a physiological problem in terms of movement. It is a deficiency of something in between, namely of a kind of anticipation of future motor behavior, a "motor intentionality". Merleau-Ponty says:

> Illness, like childhood and 'primitive' mentality, is a *complete form of existence* and the procedures which it employs to replace normal functions which have been destroyed are equally pathological phenomena. It is impossible to deduce the normal from the pathological, deficiencies from the substitute functions, by a mere change of the sign. We must take substitutions as substitutions, as *allusions to some fundamental function that they are striving to make good*, and the direct image of which they fail to furnish. The genuine inductive method is not a 'differential method'; it consists in *correctly reading phenomena*, in grasping their meaning, that is, in treating them as *modalities and variations of the subject's total being.* (Merleau-Ponty, 2002/1945, pp. 123–124, my emphasis)

Furthermore:

> Induction succeeds only provided that it is not restricted to noting things as present or absent, with concomitant variations, and that it conceives and *comprehends facts as subsumed under ideas not contained in them*. It is not a matter of *choosing between a description of the disorder which furnishes the meaning and an explanation which provides the cause*. There are, moreover, *no explanations without comprehension* (ibid. pp. 132–133, my emphasis)

In other words, Merleau-Ponty does not explain illness by either describing its symptoms or by assuming that the illness is caused by disturbances of particular physiological processes. An illness is also not simply the opposite of being healthy. This would involve an extrinsic normative perspective according to which we evaluate what is going wrong only from the outside. This way of approaching illness rather departs from the assumption that a patient is a complete being and that the illness is a particular variation or expression of that complete being. Such a perspective crucially involves an *intrinsic* normative stance. Understanding symptoms means interpreting them as indicators for something more fundamental that has to be comprehended from within the subject, not by merely looking at the symptoms from an external perspective. By saying that pathological behavior strives to make good a fundamental function Merleau-Ponty implies that the subject already has an intrinsic bodily understanding of that fundamental function. She already has grasped that something has gone wrong and tries to substitute it with an alternative strategy. This alternative strategy then shows up as a symptom or visible pathological behavior.

8.2.2 Worries about Parnas and Sass' Account of Schizophrenia

I suggest that Parnas and Sass' account of schizophrenia can be seen as an example of the strategy that Merleau-Ponty criticizes. They account for symptoms in schizophrenia either in terms of a description or of a causal explanation that remains with an extrinsic account of symptoms. They do not treat pathological experiences as expressions of a "complete form of existence" but assume they are the opposite of what is considered healthy (good) in an external normative sense.

The first issue arising in Parnas and Sass' strategy is that it indeed accounts for symptoms by what Merleau-Ponty above calls "a mere change of the sign". They explicate the pathological experience by directly deducing it from the non-pathological. For example, a feeling of alienation (the pathological) is said to be an "altered sense of self-presence" (Parnas and Sass, 2010, p. 232). This implies an extrinsic perspective on normativity. Normal self-presence involves that one simply *is* a subject, directed at and immersed into the world – this is the way it should be. But this is not the case in alienation where there is no such direct acquaintance with the world. Alienation is evaluated as the absence of an experience that the patient should normally have. Alienation must therefore be a symptom of a disorder of self-presence. Similarly, to claim that a person's experience of distance between herself and her thoughts hints at a disorder of first-person perspective is to deduce a deviation from what is taken to be normal of the first-person perspective in the minimal self, namely a direct for-me-ness of my thoughts. The for-me-ness in first-person perspective is absent, so the experience of distance is considered "unhealthy" and treated as a symptom of a disorder of first-person perspective.

This strategy is exactly what Merleau-Ponty rejects as a "description of the disorder which furnishes the meaning" (Merleau-Ponty, 2002/1945, p. 132). A report of a particular experience associated with schizophrenia simply *means* an absence of a proper sense of self-presence. Crucially however, this does not consider what the disorder means from the perspective of the patient herself.

One might object to these considerations that Parnas and Sass actually attempt to provide an explanation of symptoms in schizophrenia that not merely involves considering them as a deviation of what should be the normal case. They suggest for example that social isolation can

be explained by linking it causally to the symptom of self-alienation. It is *because* the patient feels alienated that she feels "*ontologically different* from others, and therefore alone" (ibid., p. 233, original emphasis). Social isolation is "an outgrowth *from within*, from this inner sense of profound ontological solitude" (ibid.). According to Merleau-Ponty however, what causes "a 'psychic fact' is never another 'psychic fact'" (Merleau-Ponty, 2002/1945, ibid., p. 132).[89] A second issue with Parnas and Sass' approach to schizophrenia is therefore that while a particular symptom is explained by an underlying cause, this cause is yet another experience, namely alienation. Explaining a symptom in terms of another symptom does not consider the deeper function *for* the patient that the symptom is striving to make good. The worry thus remains the same. To say that the patient feels alone because she is alienated does not yet capture what being alienated or feeling alone means for the patient herself. Understanding this meaning is required in order to see *why* the patient is socially isolated. The inner linkage between the symptoms is not graspable from the outside but only from the perspective of the patient's striving to accommodate her situation.

Comprehension of illness thus needs to go beyond enumerating what is present or absent, it consists in treating the symptoms or reports as referring to something that is "not contained in them", namely that which is the "subject's total being" or its "complete form of existence" (Merleau-Ponty, 2002/1945, p. 124, p. 132). This kind of comprehension is missing in Parnas and Sass' account of schizophrenia. Their analysis does not allow understanding symptoms as "allusions to some fundamental function that they are striving to make good" (ibid., p. 123) because it evaluates symptoms and their interrelations from an external perspective. It does not consider that the patient has her own intrinsic norms according to which she evaluates her experiences and interacts with the world.

8.3 Intrinsic Normativity and Sociality – An Alternative Perspective on Mental Illness

Consider what an alternative explanation of mental illness inspired by Merleau-Ponty's criticism on the descriptive-causal approach to illness could generally look like. Recall how Merleau-Ponty accounts for apraxia in the case of Schneider: rather than considering pointing as part of grasping movement, it should be acknowledged as an action in *its own right*. Both grasping and pointing are two *types of being in the world*, and as seen in Chapter 5, this being in the world is an *entire* embodied existence that is genuinely directed at the world (see Section 5.1.1).

In order to explain a mental disorder we thus have to understand what function or reasons the symptoms and "abnormal" experiences have with regards to the person's entire existence. A disorder of the self is not simply the absence of what is considered normal for an outside perspective. Symptoms and experiences still display a way to enact the self even if it is a way that may dramatically differ from previous or other "ordinary" ones. They are reflections of an intrinsic striving of the person as a whole to negotiate her existence and interaction with the world.

Before considering what mental illness is we should thus determine what the fundamental properties of human existence are that a mental disorder arguably substitutes and strives to "make good" (Merleau-Ponty, 2002/1945, p. 123). In which ways do humans generally adapt to the different contexts of existence? What is this "human consciousness which is wholly present in every one of its manifestations" (ibid., p. 138).

In the following section I show why these questions cannot be addressed when the conception of a healthy human existence is equated with the minimal self. I illustrate that this

[89] Thomas Fuchs (personal communication) worries hereby that to give a systemic account of the disorder is precisely the job of phenomenology and psychopathology, and to consider the genesis or unfolding of a disorder may be crucially dependent on such a prior account. I agree with this. I do not mean to say that a phenomenological perspective is wrong per se, but that in order to *better* understand the phenomenological perspective we need to have an idea of what gives structure to this phenomenology; this could involve for example an intrinsic normative stance.

worry is actually due to a deeper issue arising in embodied approaches to psychopathology in general.

8.3.1 The Individualistic Perspective in Embodied Cognitive Science

Merleau-Ponty's account of apraxia suggests that interactions with the world are bodily and that the properties of our existence are properties of the lived body (see Chapter 5). Grasping and pointing are seen as variations of a more general "motor intentionality". Being directed at the world means to be bodily directed at the world. From this perspective the patient evaluates her existence and interactions with the world according to intrinsic norms that are concerned with her embodiment. Recall from Section 8.1.1 that the minimal self is primarily associated with the pre-reflective and self-affective organism, i.e. with the body: the constituting dimensions of the minimal self are self-presence, first-person perspective and phenomenality and they manifest in different kinds of experience (for-me-ness, directedness, tacitness etc.). Existence from this perspective is an individual bodily existence directed at the world. The dimensions of minimal self realize this embodied and embedded existence to ensure, as Parnas and Sass' explicitly suggest, the "ongoing self-affection of a living organism" (Parnas and Sass, 2010, p. 230). This means that the minimal self cannot be equated with the complete subject's being because it is after all only a part of the "workings of subjectivity" (ibid., p. 230). The self is not only the living body or organism but also implies other dimensions, e.g. the intersubjective or narrative self, which Parnas and Sass have however chosen to not consider in their assessment (ibid., pp. 230, 232, see Section 8.1.1).

This has an important consequence. Even though Parnas and Sass' analysis does justice to the patient's subjective perspective by considering her reports of pathological experience, their analysis rests upon a phenomenology that is as it were primarily embodied in terms of the organism and individualistic. The minimal self is explicitly not concerned with intersubjectivity, but only with the phenomenology of individual embodiment.[90]

This mirrors a view that has recently become acceptable in cognitive science. Cognitive systems are not brain-bound, but embodied cognitive systems (see Chapters 1, 4 and 5). I have suggested that even with recent elaborations in terms of phenomenology the embodied perspective on cognition remains individualistic and I have shown in the context of LIS and BCI that it is unable to account for the complete nature of human existence, which is crucially social (chs. 4–7). The focus on individual embodiment is also reflected in recent accounts in psychopathology (e.g. Fuchs, 2005, Fuchs and Schlimme, 2009, Gallagher, 2006, Drayson, 2009, Blanke and Metzinger, 2010). One begins to acknowledge that since human beings are embodied we might have to consider that psychopathologies can be explained in terms of disturbed bodily structures. Accordingly, new forms of body-centered psychotherapy are put forward that shall, e.g. re-establish a patient's disturbed sense of bodily self-presence (Röhricht, 2009, Koch et al., 2007).

And yet, as seen in the previous chapter, human existence is not exhaustively accounted for from the perspective of the individual body. We are not only subjective bodies embedded in a world of objects, but we are also surrounded by others. It appears that the social plays a vital role in accounts of the nature and phenomenology of complete human existence. For this reason, it may also be of crucial relevance for the way we make sense of mental disorders.

Interestingly, Parnas and Sass' proposal already acknowledges this implicitly. After all, their assessment of schizophrenia is not limited to consideration of bodily or mere individual experiences, as they seek to explain e.g., *social isolation*. They argue that social isolation is caused by

[90] Beata Stawarska (2009) has recently argued that classical phenomenology does not acknowledge the dialogical character of experience. She proposed a polycentric approach on the structure of experience in which the individual's experience is "situated *and* intrinsically interrelated with the situated perspective of the other" (Stawarska, 2009, p. 16).

alienation. However, alienation arguably indicates a disturbance of the minimal self, i.e. the self-affective organismic individual. What is worrying is that the self-affective organism is concerned with "obstacles, tools, objects of the desires" (Parnas and Sass, 2010, p. 236). But obviously, the patient is not alienated merely from a world of objects, but also from a world of other subjects, because something is not quite right in the relations to them. Alienation might be a socially informed phenomenon. Parnas and Sass seem to presuppose this, yet nothing in their explanation of *social* isolation does justice to the fact that alienation and other particular kinds of suffering we find in schizophrenia, e.g. of "being alone", feeling "isolated" and "extra-terrestrial" are actually of a *social nature* (Parnas and Sass, 2010, p. 210). This is because nothing in their account of self-affection (minimal self) is related to intersubjectivity or what it means for the patient to exist as a social being. The origins of a social phenomenon, on this account, are by default and without argumentation assumed to lie solely within the body (or brain) of a single individual.

To conclude, if the minimal self would serve as the basis for an explanation of mental disorders, then the realm of these disorders should perhaps at the most be concerned with disturbances at the level of object perception, say, as in hallucinations. It is a risky strategy to attempt to explain symptoms that clearly indicate disturbances with regards to our social existence on that same basis. This applies similarly to every other approach to mental illness that is based on an embodied perspective on cognition. When the underlying model of the self is essentially individualistic, an account of schizophrenia, as well as of any other disorder of the self, must necessarily remain impoverished.

8.3.2 Extrinsic Normativity Downplays the Role of Phenomenology

Another general issue brought to attention in the discussion of Parnas and Sass' account of schizophrenia is that the minimal self and considerations of disorders of the minimal self lack an account of intrinsic normativity. Parnas and Sass' make an important contribution to our understanding of mental illness because they illustrate for schizophrenia how psychiatry could accommodate the phenomenological dimension of psychopathology. Yet it seems that they also underestimate the relevance of phenomenology because they adopt at the same time an extrinsic normative perspective. This illustrates another more general issue for embodied approaches to psychopathology. The worry is that if subjective reports of pathological experience are mainly evaluated against external standards of what counts as right they risk being reduced to mere affirmations for an external assessment that in mental illness something is wrong. To say for instance that a sense of alienation means that the person has a disturbed sense of self-presence does not explain alienation but basically underlines that it means the patient is not healthy. To say that a symptom indicates that something is not normal is reducing the subject's being in schizophrenia to being a pathological being. This is odd since by looking at symptoms we also try to understand the pathology, not just confirm that it is, indeed, a pathology.

Again, it seems that this issue in Parnas and Sass' is inherited by a general shortcoming of recent embodied cognitive science. Shifting from a classical to an embodied perspective on cognition did not change that cognitive science still does not acknowledge that cognitive systems have their own perspective on the world from which they evaluate their interactions and thus also their experience of interactions. Yet without this perspective we cannot arrive at a full understanding of why a bodily structures and bodily subjectivity matter for cognition. As a consequence, if cognitive science were to account for psychopathology in terms of phenomenology then it has to first show that phenomenology matters for cognition in a deeper sense than merely adding the subjective view to what is already known from an objective perspective. Only by adopting an intrinsic normative perspective on cognition can we begin to consider what human existence amounts to and what a pathological variation of that existence and its phenomenology in its own right consists of. The only framework that considers cognition and normativity in that vein is the enactive approach to the mind.

For this reason in the last three sections of this chapter I explore some implications of my proposal of socially enacted autonomy for psychopathology in general and schizophrenia in particular.

8.4 Approaching Psychopathology from the Perspective of the Socially Enacted Self

The question to be asked in order to explain mental illness should be: what do the symptoms tell us about the way in which the whole subject currently exists and how she evaluates her situation? Of course this depends on what we think that being a whole human subjectivity essentially and generally amounts to.

In this section I apply the proposal for a socially enacted self outlined in the previous chapter and show how it might serve as an alternative starting point for understanding mental illness. Please note that this is a proposal in development and not a fully defended position. It raises issues that will require wide and thorough elaborations in future research, but it also suggests new avenues for pursuing them.

It should be clear by now that the problem with the minimal self is not that it is based on a phenomenological account but that it is based on an acknowledged reduced version of phenomenology. It is the phenomenology of a self that is basically individual. The core dimensions of the minimal self (self-presence, first-person and phenomenality) all imply aspects of experiences about being someone distinct. However, as seen in the previous chapter, these experiences hardly capture the essence of *human* self-experience. We experience ourselves not primarily as distinct bodies that interact with a world of objects. Rather, we experience ourselves as social beings. The fact that persons with schizophrenia but also with other mental illnesses, such as borderline personality disorder appear to show disturbances of intersubjectivity (Fuchs, 2005, 2007) requires an explanation that goes beyond phenomenological considerations of subjectivity as an embodied, yet isolated center of existence. We need an approach to individual subjectivity that acknowledges, what is expressed in symptoms of psychopathology, but obviously not entailed in recent conceptions of the embodied individual: the role of the other. An exception to this is Fuchs (2011). He acknowledges that the body is a "centre of subjective and intersubjective experience" (p. 200) but does not elaborate what being that centre is.

In the last chapter, I suggested that human self-experience has a basic two-fold structure. On the one hand we experience ourselves as being distinct from other people. But on the other, even though this seems to be widely neglected, not only in cognitive science, but also in Western culture and society, we also experience ourselves as directed at others and as striving towards engagements with them.

If we want to explain the phenomenology of disorders, i.e.. the symptoms and experiences of schizophrenia, then this general characterization of human experience as social should be accommodated in the model of the self underlying the explanation of mental illness.

We have learned that enactivism seeks to account not only for a part of cognition or a particular cognitive capacity but rather for the sense-making system as a whole, in its general organizational structure and interaction with the world. With this integrating attitude enactivism is clearly on a par with Merleau-Ponty who asks to consider the "subject's total being", i.e. its complete existence. In the previous chapter I have proposed steps towards an enactive model of the self that accommodates the nature of human minds as genuinely social. Our being in the world is intentional, but adding to Merleau-Ponty, we do not characterize it only in terms of bodily intentionality, but primarily in terms of a *social intentionality*.

I have suggested defining the enactive self in terms of a precarious operationally closed network in which behavioral and constructive processes overlap. These processes are socially enacted processes that realize a distinction from and participation in the social environment. I proposed that precariousness at this level consists in the fact that humans strive to maintain both

distinction and participation and that the maintenance and negotiation of distinction and participation is not achieved solely by individual means but requires interactions with others.

In contrast to Zahavi (2010), the social dimension is therefore not an add-on to a bodily core self, it is rather fundamental to the organizational structure of human existence itself. The idea is that the human self as socially enacted autonomy is ontologically and epistemologically speaking an intersubjectively constituted system. Yet this does not mean that it exists as soon as we are born. Even though I would certainly assume that young infants already possess the capacities for the development of the human self, it nevertheless requires an initial phase of genuine social interaction.

As a consequence, human self-consciousness arises not solely within the biological domain of the brain in connection with the individual body but in result of *social* engagement with others. Human beings do possess a pre-social form of self-awareness, but I suggest that this self-awareness is only a self-awareness of being an embodied identity (as captured by the minimal self). It is however not yet that of being a self in relation to others.

Let me also clarify at this point that the present suggestions for an alternative look at psychopathology do not imply that biological or neurobiological explanatory factors are any less relevant for understanding psychopathology. The proposal for a socially enacted self does not deny that neurological and physiological structures matter for cognition. The point, quite to the contrary, is that in order to fully understand their role for cognition they have to be integrated in an encompassing social perspective on human existence. The social perspective helps to differentiate human beings from other animals in terms of new organizational structures and sense-making strategies. It is however not meant to rule out that humans can also be described in terms of their biology or neurobiology. The social perspective rather *informs* other levels of description. Understanding the biological basis of mental illness is therefore doubtlessly important but this biological understanding should be approached in a way that accommodates the social nature of human beings.

For Parnas and Sass to say that schizophrenia is a disorder of the self does not mean that a person loses the minimal self, it "indicates, rather that 'minimal self' is fragile, constantly threatened and *unstable*" (Parnas and Sass, 2010, p. 230). Thus the authors correctly observe *that* the minimal self is fragile. Yet because the notion of minimal self lacks a normative account that is intrinsic, the question of *why* this is the case remains open. An enactive perspective on the self can offer a way of addressing this question.

Parnas and Sass have identified disorders of the self as the *trouble générateur*, i.e. the viewpoint from which we can understand symptoms of schizophrenia. But in light of the precariousness of human existence, the *trouble générateur* has to be identified at a much more fundamental level. We do not only find it in examinations of pathological forms of existence, but in fact also as part of human existence in general. Human beings strive to remain individuals that separated from others but also required to engage with them and in order to achieve this they rely on each other. My suggestion is that here lies the reason why we can experience problems in our existence as social beings, i.e. why the human self is "fragile" and "constantly threatened". The human self is never a given and being a self always comes at a price: it is genuinely and constantly *concerned* with its continuation and maintenance. To ensure this basic goal requires the engagement and negotiation with the social world, the outcome of which is never certain or guaranteed. Vulnerability and dependence are simply part of the description of the "conditio humana". This is where we find the origins of mental illness.

In this vein, I would like to interpret Merleau-Ponty when he says that understanding an illness requires "correctly reading phenomena, in grasping their meaning, that is, in treating them as modalities and variations of the subject's total being" (Merleau-Ponty, 1945/2002, pp. 123–124). A particular disorder is not an absence or dysfunction of the mechanism, which constitutes the normal self as Parnas and Sass' interpretation suggests. A disordered self never stops being or becoming a self – it remains a self that strives to maintain itself. The disorder is an expression of this striving, it hints at our existence as being genuinely precarious and *concernful*, and the

symptoms and "abnormal" experiences should be read as particular manifestations of the struggle to maintain this human existence. We find this proposal nicely reflected in a recent account of schizophrenia as a disorder of the person by Fuchs:

> Schizophrenia is the illness that can only emerge because humans become conscious of themselves and thereby acknowledge themselves as a person amongst other persons. In schizophrenia this ability of human beings to perceive the other person as such is reversed against the patient herself and it threatens her with the demise of her person. Nevertheless, because the schizophrenic is suffering, feeling anxious, threatened, alienated and overwhelmed, she attests that she remains a person. Even in the extreme case of depersonalization we still recognize the person in the sick person. (Fuchs, 2002, p. 245, my translation)

If the human self is genuinely social then an account of a disorder can start but cannot remain with a description of the person's momentary pathological experiences. We would have to understand that the pathology is a particular variation of a subject's total being but that the patient still strives to maintain her own human self. The patient is not all of a sudden an isolated creature existence. For this reason it is important not to consider reports or experiences independently from the social realm. And they should be understood not merely from an external perspective, but based on the patient's own evaluation.

In the next section I elaborate this proposal by showing how the socially enacted self can help to account for a particular pathological variation of the self. I suggest considering schizophrenia as as the manifestation of a fundamental imbalance of the two basic dimensions of the socially enacted self – distinction and participation.

8.5 Schizophrenia as an Imbalance of Distinction and Participation

Whereas Parnas and Sass associate schizophrenia with the disorders of the minimal self, I propose that it expresses a problem or a complex of problems of the socially enacted self. The autonomous self emerges from a dynamic interrelation of social processes that show a tendency towards distinction and/or participation. My suggestion is therefore that schizophrenia indicates a severe imbalance in the dynamics of these processes of distinction and participation. It expresses a struggle to balance distinction and participation and thus to negotiate the two-fold striving for being someone on her own and being someone with others. The quality of this striving manifests itself as pathological in part because it does not meet the standard and expectations of the present context in which the person is embedded. But in part also because the patient loses basic capacities for maintaining her human social existence. The symptoms express the patient's suffering from permanently failing to be someone who can be recognized by others or who can successfully engage in social interaction dynamics.

This is nicely mirrored by two observations that persons with schizophrenia make about themselves. They report to have a "weakened sense of being a self-present subject" and a "lacking immersion in the world" (Parnas and Sass, 2010, p. 228). The first observation can be related to processes of distinction; it indicates that the individual self strives to be able to experience itself as an entity distinct from other people, as the source of her own action, as the initiator and enabler of her own goals and intentions. The second observation concerns what we have described as the intrinsic need of the human self to be recognized as a participant in social interactions.

According to the approach outlined in Chapter 7, both processes are intimately related. In order to be a distinct individual I also need to be able to engage with others. On the other hand, I can only engage with others when I have also some sense of being an individual. The general idea would be that an emphasis of processes that constitute the self as distinct could lead to a reduction of the self as participant and vice versa (see Section 7.2.3).

For example, if I constantly experience a mismatch in social interactions between what I consider to be expressing my own intentions and how other people respond to it, or if social interactions repeatedly fail to provide recognition for who I think I am then this might increase my sense of being distinct from the people I interact with. If this experience of myself as distinct from others is extreme then my self-experience as an individual could become distorted. It might weaken my general sense of self-presence which at the embodied level expresses it self in terms of a disturbed sense of agency and/or bodily ownership (Gallagher, 2000, Jeannerod, 2009). A disturbed self-presence can negatively affect further encounters with others eventually leading to feelings of alienation, distance to others and experiences of isolation or dis-integration. On the other hand, if my sense of being a distinct individual is too weak because I am for example constantly losing myself in the dynamics of social interaction this can lead to another disturbance of my general sense of self. Here the participatory sense of self-experience is increased. This might manifest in a variety of phenomena, which do not even have to be considered pathological: shyness, feelings of dependency, feelings of being overruled, and flooded by the other, or even utter powerlessness.

The maintenance of the self thus requires not only variations of both types of processes (distinction and participation), but also that they are balanced with each other or that adaptive resources are available to cope with states of imbalance. To be able to ensure being someone who is somebody in its own right and at the same time somebody who can be with others is generally difficult. But this struggle is not exclusive to severe disorders; it is – to different degrees – part of human life.

As a consequence, the phenomenology of schizophrenia is not the opposite of the phenomenology of normal existence, but rather follows its basic logic (granted that this logic is intrinsically normative). Schizophrenia is a severe variation of the phenomenology of human existence. It is an existence that struggles with the maintenance of itself under the most extreme conditions. Yet because a patient still is a social human existence it is also a *comparatively* extreme way. We understand better why schizophrenia is such a serious condition when we read the symptoms as expressions of a dramatic imbalance between the two dimensions of distinction and participation. It may be so dramatic that one dimension may almost entirely inhibit the other one.

A genuine worry at this point is that this account would downplay the seriousness of the condition in that schizophrenia now appears as only a special existential condition and not an illness anymore. But this is precisely what I am suggesting in line with Merleau-Ponty: we *should* consider illness as a whole existence. To say that illness is a particular existential condition does not have to mean that we should not take this condition seriously or assume that it is any easier to deal with or that it would not require careful treatment. I do not want to suggest that an extreme imbalance in distinction and participation is not a severe disturbance of the self. The point is that in light of the present proposal in this disturbance we find expressed a particular striving to overcome this disturbance, a striving that already belongs to the human condition in general and that in this respect is thus "pretty normal". Crucially, from this perspective we could explain why schizophrenia is such a serious condition. It is so severe because the self is threatened in two basic dimensions that condition its very existence.

In the following section I apply these considerations to the experiences and the symptoms in schizophrenia that Parnas and Sass explicated as disorders of the minimal self in Zahavi's sense and offer a different interpretation.

8.6 An Alternative Look at Symptoms in Schizophrenia

To conceive of schizophrenia as a severe imbalance in distinction and participation may offer a more complex and dynamical interpretation than individualist accounts. The subjective experiences are still considered as "interdependent *aspects* of a single whole phenomenon" (Parnas and Sass, 2010, p. 231), but the underlying model of the self provides an idea of the kind of

interrelations that structure this phenomenon. Please note, again, that this is a proposal in development, I do not attempt to provide a full new model of explanation of symptoms in schizophrenia.

I now come back to the selection of experiences and symptoms that we have discussed and offer an alternative interpretation in terms of a phenomenological account that also entails an intrinsic normative perspective of the patient herself. There is firstly, the experience of alienation and the phenomenon of social isolation. Secondly, I would like to address experiences of hyperreflexivity.

8.6.1 Reconsidering "The Feeling of Alienation" and "Social Isolation"

Recall from Section 8.1.2 Parnas and Sass' explanation of the feeling of alienation frequently reported by persons with schizophrenia. In their view it can be explained as a "diminished or altered sense" of self-presence and of "one's very existence as subjectivity" (Parnas and Sass, 2010, p. 232). In the light of the socially enactive self, Parnas and Sass' observation is partly right: statements such as "I don't feel myself", "I am not myself", "I am losing contact with myself", "I have a strange ghostly feeling as if I was from another planet", "I am almost nonexistent", "I am becoming a monster" or "I am turning inhuman" (ibid., p. 233) are all testimonies of something that concerns one's existence as a subject. However, as suggested above, to claim that these reports are just the opposite of the normal sense of self-presence does not sufficiently explain them. From an enactive perspective understanding the disorder involves adopting an intrinsic normative stance and consider the social interaction dynamics that co-determines the manifestation of symptoms. Subjective human existence is existence in distinction *and* in participation. A feeling of alienation from the present world makes sense if one presupposes that subjects actually strive towards some kind of balanced engagement with the world.

That I feel like an alien already implies that I could actually feel different and that I might possibly also feel more like the others. Even if a feeling of alienation certainly concerns an individual's identity as someone distinct, at the same time it concerns its identity as someone amongst others. It is primarily against other people that my experience of alienation manifests itself. Similarly, to experience that one is becoming a monster hints at the possibility that either I could have or might have had a different identity or that presently everyone else but me seems to have a non-monster identity. It is central to this experience that others already see me or might see me as a monster. There is thus an implicit discrepancy between what I am as a distinct human identity and what I should be in order to also be a participant in social interactions, a fundamental difference between the other and me. This is problematic because from an enactive stance in order to be my own self I also need others. In alienation the patient's experiences of identity in distinction could thus overweigh those of participation. This matters for the patient herself because it affects her ability to engage with others and how others respond to her, which in turn is required in order to be a social individual.

Consider now Parnas and Sass' explanation of social isolation: they have argued that it results from the sense of alienation. In contrast, in the light of the present proposal we can say that alienation and social isolation actually refer to the same problem, namely that of a social self struggling to maintain a balance between the processes that constitute its identity and failing to sustain fluid participation.

Note that I do not mean to say that alienation always needs to be accompanied by "actual" social isolation. A person may be visited by friends and family and still feel alienated. I would thus differentiate between social isolation as the observation of the fact that someone is not (or only rarely) having actual social interactions, and social isolation as experienced from the patient's perspective, a required result of a sense of alienation, which can but does not have to lead to an actual diminishment of social interaction.

In order to link social isolation to alienation we need to consider how both have come about. The questions thus are "Why should an experience of alienation lead to social isolation?"

or "Why do we often find that schizophrenic persons are socially isolated?". Why does a person feel "*ontologically different* from others", or desires to be a "time-traveller or an extra-terrestrial"? Why is she feeling "alone"? (Parnas and Sass, 2010, p. 233).

An answer in light of the present approach requires a consideration of the subject's striving for the opposite of isolation, namely to be involved with others and the risks that this entails. To feel ontologically different from other people implies the possibility that one could or should not feel different from them and would actually want to be more similar to them. Likewise, feeling alone implies a suffering that hints to the contrary, namely the wish to feel connected with others. That somebody desires to be a time-traveler or extra-terrestrial could indicate that she might either want to be different or that she feels different from the other people surrounding her now.

The basic idea would thus be to understand the above reports as indicating that the person still strives to be a distinct and participating self, but that she is not or has not been able to fulfill this goal given the circumstances of her present or past social embeddedness. If the person feels like an alien this can negatively affect interaction with others. Others might get the impression that the person is acting oddly and let this change their reaction to the patient, say, by being puzzled or protective. This alters the social interaction dynamics so that she patient might experience repeated failure to connect to others. The quality of the social interaction can then even increase the patient's sense of alienation.

As a consequence a person may actually have to escape from the present moment to avoid others that she feels are different from her and the risk that to engage with them makes her feel even more like an alien. But this does not mean that the patient would not want to be engaged with other people. The apparently odd wish (or even conviction) to be a time-traveler or an alien can be interpreted in a positive way, namely as an indicator that the patient wants to ensure that the discrepancy between her and another person is not as intense or even absent. From her own perspective this would allow her to re-engage with the other and thus conform to her two-fold goal of being a social individual.

We can specify this interpretation by considering the symptom with regards to either of two basic goals. In this vein we can interpret a sense of alienation as an expression of the patient's striving to be a recognized individual amongst others and the difficulties to achieve this in the social interaction. Wanting to be an alien or a time-traveler could indicate that the discrepancies in current social interactions are experienced so dramatically that the person rather "prefers" to not be human or to at least live in another time. From the outside this sounds wrong, crazy or maybe just incomprehensible. Yet from the perspective of a self that intrinsically strives to ensure that she is able to engage with others we can understand these symptoms as alternative strategies to maintain that goal: Being an alien is a way to remain somebody who could be engaged with (not by humans, but maybe by aliens) and time-traveling counts as a strategy for finding an alternative social world that one could be engaged with (not in this century, but maybe in the past or future).

As a consequence, social isolation is understood not as a hardship ascribed from the outside; it corresponds to a person's way of protecting her identity from dissipation. It could be a needed escape and a survival strategy: if others are different from me then in order to interact with them I need to be more like them. Yet if I am more like them, then I stop being who *I* am, that is, somebody who is different. I might thus choose to rather not engage with an environment so different because it is threatening my self. For the patient being an alien and avoiding social interaction could be a strategy to ensure that her identity is protected against a world of others that seem to neglect it.

It is thus possible to read the report in both senses, with an emphasis on the individual as identity or with an emphasis on its participatory nature. Both need to be understood against the dynamics of the social context in which the patient attempts to realize her goals. The patient still strives for participation, yet she does not want to (and probably cannot) engage with people that she feels are fundamentally different from her and that continue to make her feel that way. The

person exhibits what seems characteristic of every human self. However, her desire to be oneself and also not alone creates a terrible dilemma that is deepened due to a permanent failure to achieve this goal. Repeated odd or altered social interaction cannot only cause but also reinforce serious existential suffering and pathological symptoms.

8.6.2 Reconsidering "Hyperreflexivity"

Recall from Section 8.1.2 that for Parnas and Sass, hyperreflexivity indicates a disorder of the first-person perspective and phenomenality. Both require an ongoing self-affection, a "function of the lived body, moving in space" (Parnas and Sass, 2010, p. 234) to constitute "the *sense of self*, of being a substantial (spatial) subject" (ibid.). Self-affection is implicitly involved not only in perception, but also in thinking processes and consciousness as such.

I have suggested that schizophrenia displays a fundamental difficulty in negotiating the tensions of the processes of distinction and participation. Therefore, even though I agree with Parnas and Sass that self-affection grounds our basic sense of being a subject, I do not think that this sense of self is described sufficiently in terms of a bodily or organismic functioning. If bodily self-affection is part of human subjectivity and human subjectivity is considered as genuinely social, then we should consider that aspects of human self-affection would have a social structure as well.

Self-affection involves not only "experience of appetite, vital energy, and point of orientation". It not only "organizes our experiential world" (Parnas and Sass, 2010, p.236) *of objects*. But from an enactive perspective, self-affection also endows us with an implicit for-me-ness underlying our orientation in a *social world*. It should give one a sense of vitality as a social human being, as someone distinct from and involved with directed at others.[91]

This is why hyperreflexivity should not be understood on the grounds of a malfunction of the body. That the person experiences her body as alienating and uncanny in schizophrenia does not primarily "illustrate the essentially embodied nature of the self" (Fuchs, 2005, p.105). This claim would imply the same "change-of-sign"-strategy we find in Parnas and Sass. It is rather an expression of a fundamental struggle to maintain and negotiate the tension between processes of distinction and participation that had become extreme. The body expresses this struggle. The "disembodiment" in hyperreflexivity does not refer to a "disturbed relationship to objects", but should be considered as an indicator of an extreme difficulty in engaging in social interaction dynamics required for maintaining relationships to other subjects (Parnas and Sass, 2010, p. 236). To experience a distance between oneself and one's own thoughts can be seen as a problem because this distance prevents the patient from just engaging with others and it prevents them from engaging with her. When the patient remains preoccupied with the observation of overly intense bodily sensations, she is mainly concerned with herself as being distinct. In the interaction she might find it difficult to coordinate or tune in, the encounters with others involve an extreme care. Instead of being able to just experience, the patient performs a parallel analysis of what is currently experienced. Instead of letting herself be affected by the other, she must understand what others may think. Instead of just replying or reacting, she reasons about what should be said or done. Obviously, this increased awareness on the conditions of the engagement prevents the person from just engaging with others. But the permanent absence of automaticity and transparency in social interaction even increases the patient's self-awareness. The patient feels much aware of herself yet she repeatedly fails to comply to the goal to be somebody who is also open and engaged with others. As a consequence, in hyperreflexivity experiences of being distinct critically overshadow experiences of being a participant.

[91] In line with the proposal I have made in the previous chapter I would like to emphasize again, that this does not mean that we cannot assess how other kinds of autonomous identities may play a role in an account of human beings. It is possible to speak of a systemic identity already at the bodily level, but this would be always an identity with regards to the body, not the human being as a complete existence.

Recall from Section 8.1.2 that Parnas and Sass point out that increased self-awareness (in hyperreflexivity), goes together with a *de*creased self-presence (the sense of being someone). From a social and intrinsic normative perspective self-affection ensures that a person experiences herself as both participant in interactions and yet distinct from others. In case of a permanent failure to *just* engage with others self-affection becomes reduced to only one dimension of self-presence, namely distinction. It lacks a clear sense of the second dimension, namely participation. This could explain why hyperreflexivity causes such great suffering: the patient feels mainly distinct but not connected anymore. Her general sense of being someone is therefore decreased.

As a consequence we can offer an alternative interpretation of experiences of increased bodily awareness. To experience my body as a "mere thing", to "feel amazement *to be a body*" or to observe an "extreme separation between mind and body" (Parnas and Sass, 2010, p. 233) are expressions of the person's difficulty to comply with her basic goal to remain related to others. The person with schizophrenia experiences her body as instrumental or a "mere thing" because it is delivered of its role as a *mediator for her social existence*.[92] From the perspective outlined in this thesis, bodily identity on its own does not provide me with a feeling of being someone, i.e. a person or a subject. I also need to engage with others. I will feel like an object or "refrigerator" when I am no longer a subject, viz. no longer some*one* who is involved with other subjects or who other subjects are involved with. On a more positive interpretation, the fact that these feelings cause suffering and amazement could be an indicator that the patient also continues to strive to remain a social subject.[93]

8.7 Conclusion

In this chapter I have explored further empirical implications of my proposal for a socially enacted self and the role of the body. I discussed an embodied and phenomenological account of schizophrenia by Parnas and Sass. I have argued that their explanation of schizophrenia in terms of disorders of the minimal self is not fully satisfying as it remains with a description of symptoms or assumes the pathology to be the exact contrary of what is considered as normal. I have suggested that their approach illustrates a more fundamental worry arising for any merely embodied approach to cognitive science. Much of embodied cognitive science implies an individualistic perspective on cognitive systems and it adopts an external perspective on normativity. For this reason it is a useful framework for understanding psychopathology and role played by the body therein.

Inspired by Merleau-Ponty's approach to apraxia I have suggested that understanding pathologies requires to not merely interpret symptoms based on externally established norms but to consider that a patient is still is a complete subject evaluating her interactions with the world according to her own intrinsic norms.

Enactivism is currently the only framework in cognitive science that adopts an integrative perspective on cognitive science which it also does justice to the fact that humans have intrinsic

---

[92] I have to admit that I am currently not sure how strongly the claim that the body is the mediator of our social existence should be interpreted. While the body plays a crucial role in realizing socially defined actions, it obviously also mediates other forms of interactions (as seen in chapter 6). If the mediation of these non-social actions is disturbed there might be an experience of alienation that is not socially structured. On a radical reading of the claim, *every* bodily alienation is ultimately rooted in problems in intersubjective engagement. I am sympathetic with this, but do not present any arguments for it here. Support for this claim might come from research on the role of social mediation in early infant development (see De Jaegher et al., 2010 for a more detailed elaboration on this question). But to elaborate on the consequences of the present proposal requires extensive future work.

[93] Note that this would also explain why not only people with mental illness, but also those with partial or global bodily paralysis as in LIS experience a sense of alienation (see section 3.2). Their body feels uncanny not because it no longer engages in interactions with the environment but because it ceases to ensure an ongoing connectedness with other people.

goals based on which they maintain their existence. For that reason I proposed that enactivism might be a better starting point for understanding psychopathology in a Merleau-Pontian sense. Humans are social beings and they have an intrinsic striving to negotiate their existence as being distinguished from other subjects while also being able to engage with them (distinction and participation). The maintenance of this social existence is not achieved solely by individual means but relies in part on social interaction processes. This, I suggested, provides the basis for developing a new approach to psychopathology. To understand symptoms is to assume that the basic striving to maintain the self is still present and that "abnormal" behavior and experiences of the person can be interpreted as attempts or alternative strategies to accommodate this intrinsic goal. In this vein, I have proposed to consider schizophrenia not as a disturbance of bodily self-affection, but as indicating a most severe imbalance between processes of distinction and participation. Symptoms of schizophrenia, such as bodily alienation, social isolation and hyperreflexivity are then seen as reflections of the individual's struggle to be somebody who is both an individual and a participant. They can be read as strategies for negotiating a social existence. Crucial part of such an assessment of psychopathology is to consider this struggle against different social interaction dynamics that the patient is involved in. We should ask how the particular quality of social interaction can affect the patient but also how the patient's symptoms can contribute to mis-coordination in the interaction leading to an increase of suffering and symptoms.

The enactive perspective inspires future research on psychopathology to adopt a multiperspectival and dynamical perspective involving a back and forth between the patient's intrinsic perspective, an external evaluation of her symptoms and reports (including neurobiological and other physiological assessment) and the dynamics of the social interaction she is involved in.[94]

Human existence is that which always lives and has lived through others. An important implication of this is that pathological experiences are not merely testimonials of an individual's illness. A failure to be a self, apart from its biological and individual underpinnings, always mirrors a failure in part of the social context in which the person has been or is presently embedded in. As a consequence, the understanding of psychopathology should also involve considering the role of others, say family members or friends but also, in the context of treatment and therapy, the actual interaction dynamics between therapist, doctor or caretaker and the patient. An example indicative of this more interactional perspective on psychotherapy could be the "open-dialogue" approach (Seikkula et al., 2008, Seikkula and Olson, 2003). It is based on the assumption that therapy is "created jointly" in a social dialogue and thus involves treatment meetings bringing the patient together with a team of professionals as well as the patient's relatives and friends (Seikkula and Olson, 2003, p. 408).

[94] Clearly, this perspective has implications not only for the context of mental illness but also for other cases of illness. It should be of relevant concern for treatment and diagnosis in general to consider the dynamics and structures underlying various self-other relations (e..g. patient-doctor, patient-caretakers, patient-relatives) that occur in the medical context (for recent phenomenological elaborations on this issue see, e.g. Toombs, 1988, Zaner, 2003)

Conclusion

Let me now provide a general conclusion to this thesis, summarize its results and consequences for philosophy of cognitive science, and consider questions and issues that can be relevant for future research.

In this thesis I have looked at a basic question in the philosophy of cognitive science: what is the nature of human beings? I have taken a fresh look at this question by grounding my inquiries in the neuromedical context and confronting approaches to embodied cognition with cases of Locked-in Syndrome (LIS) and the use of Brain Computer Interfaces (BCI). I have proposed that these cases present a basic challenge for embodied cognitive science, namely to spell out the concepts used to describe human cognition and also the background epistemology that allows us to relate them in an intelligible way. Throughout the thesis I have refined this basic challenge in three basic steps.

First, I have discussed in the context of LIS and BCI-use three recent approaches to embodied cognitive science – the sensorimotor, functionalist and phenomenological approach. I explored whether their respective conceptions of body and bodily action can satisfactorily account for the fact that patients with global bodily impairment seem to remain cognitively intact. Throughout the discussion I have progressively refined this challenge and in doing so I clarified what we mean by body and bodily action. The discussion of the sensorimotor approach revealed three important issues: a body should not be restricted to biological structures, bodily action is not reducible to movement, and current embodied cognitive science runs the risk of limiting the relevance of bodily action to a mere causal factor in the development of cognitive capacities. The discussion of extended cognition reconsidered an important aspect of human cognition, namely tool-use, and invited embodied cognitive science to clarify how humans are able to incorporate tools. In the context of BCI-use we have seen that extended cognition cannot sufficiently accommodate this task because it involves a trivial conception of the body reducing it to a replaceable component of an information processing mechanism.

I have shown that the phenomenological concept of the body offers a non-trivial and not merely developmental approach to the body. On the phenomenological account human existence can never be decoupled from an actual engagement with the world and the body is the means to realize this engagement. We also learn from phenomenology that a body is not only an objective body, but also a subjective and lived body and that this involves both experiences of the world (hetero-affection) and an immediate awareness of the body itself (self-affection). This conception of embodiment helped to account for the physiological restrictions in the sensorimotor approach while also providing a plausible explanation of how external tools can play a role in cognition in terms of incorporation into the habit body. However, the discussion has also lead to a further refinement of the challenge, namely to assess how bodily self-awareness and world-directed experience are interrelated. In addition I also began to outline a much more fundamental issue for embodied cognitive science: the need to address the role of the body in social interaction.

Discussing the different approaches to embodiment in the context of LIS and BCI has illustrated that understanding the body and its role for the human mind cannot be achieved by choosing a singular explanatory perspective, considering either the nature of bodily structures (sensorimotor approach), the subjective dimension of embodiment (phenomenology) or interaction with tools (extended cognition). All of these perspectives seem to be required –they are not independent from each other but inextricably linked. Considerations of non-biological bodily structures are needed to account for incorporation of tool use and the incorporation of tools requires considerations of subjectivity.

The second step of my investigation was thus to provide evidence that the best candidate framework to account for the question of how the body matters for human cognition is the enactive framework. Based on the enactive approach I have proposed an alternative approach to the body that would allow integrating the various perspectives on embodiment in a coherent way.

According to this, operational and subjective descriptions of embodiment are integrated in a continuously active and world-engaged bodily existence. Bodily existence is at the heart a centre of self and world-directedness. The maintenance of this existence involves various kinds of bodily action that are not limited to the biological realm, movement or overt bodily activity. Importantly, the maintenance of bodily identity cannot be divorced from subjectivity because bodily experiences allow the cognitive system to appreciate interactions with the environment as relevant for that goal. The body is thus the means by which a cognitive system can be a cognitive system and it is also the reason why a cognitive system could evaluate anything as meaningful.

I have applied the enactive concept of the body to the case of LIS and BCI-use. I have shown that its refined conception of bodily action and its integrating perspective allow to account for the various challenges posed by LIS and BCI-use. It was unable however to determine the role of the body in social interaction.

The discussion of the enactive body and LIS revealed a persisting general shortcoming of current embodied cognitive science: the body-body-social problem. Human cognitive systems cannot be sufficiently explained in terms of embodiment as this view presupposes a split between the individual (that is equated with the body or organism) and the social environment. From this perspective we do not do justice to the fact that humans are not isolated subjects, but deeply embedded in a social environment, constantly having to negotiate our own existence in a world of other subjects. The important question to be addressed is thus how embodiment relates to sociality and human identity.

In a third step in the investigation I have thus explored a solution to the body-body-social problem by elaborating on the enactive approach to autonomy in terms of a social perspective. Enactivism already contends that sociality matters for cognition. In this thesis I shed light on this issue by suggesting that sociality increases not only sense-making abilities but plays a role in the emergence of individual human identity as such. The maintenance of human identity is not solely achieved by embodied means but is co-constructed in a dynamics of social interaction processes, which I specified as processes of distinction and participation. The most encompassing starting point for cognitive science to approach human beings is therefore not the body or organism. We need a multiperspectival account of the complex dynamics of individual and intersubjective processes that together give rise to an individual's identity as a cognitive system. In addition I have outlined several implications that socially enacted autonomy has for the key concepts of the enactive approach to cognition, e.g. precariousness, normativity, consciousness and the role of the body. Precariousness at the human level involves an overlap between individually and jointly enacted processes. The maintenance of human autonomy is thus not achieved by individual means but vitally dependent on a continuous engagement with others. Interactions with the environment are evaluated based on an intrinsic goal that is socially defined: humans strive to maintain an identity that is distinct from others while at the time remaining open and able to engage with them (distinction and participation). This finds reflection on the phenomenological level in that humans experience themselves not only as separated individuals but also as being directed and open towards others.

I have indicated how this new perspective on the nature of human mind helps to transcend the prevalent dichotomy between the individual and others and also to shed new light on the question of what a human body is and what role it plays for cognition. The human body is not reducible to the organism, it incorporates non-biological tools, and bodily action is neither restricted to movement nor overt behavioural action. However, if human autonomy is in part socially co-constructed then these characteristics should be seen in the light of sociality and the individual's intrinsic goal to maintain a social existence. It is possible at times to adopt the perspective of the living body itself to describe a felt sense of embodied identity. However, this does not mean that at the organizational level the body is entirely independent of the world. Bodily world- and self-awareness are integrated in an engagement with the world that is socially enacted.

I have suggested that the level of social autonomy coincides with the human self. The

human self is a self that distinguishes itself from the social environment while at the same time remaining open to it. Conceiving of the self as genuinely social set the basis for the last step of my investigation: to provide further support for the newly developed approach to autonomy by applying it to the context of psychiatry and the question of how cognitive science can account for mental disorders. I proposed that the basic reason why humans can develop mental problems is that they have to maintain a balance between being distinct and participating social individuals. The realization and negotiation of this balance is however achieved in dynamical interactions with others, the quality of which can at times deviate from the individual's intentions. I thus proposed to consider mental disorders not as a dysfunction of the normal self but as an expression of the continuous human struggle to maintain an existence that is vitally dependent on social interaction dynamics whose outcome is never sure. Based on that I have proposed to consider schizophrenia in terms of an extreme imbalance between the two basic processes of socially enacted autonomy, distinction and participation. I offered an alternative interpretation of classical symptoms, such as bodily alienation, social isolation and hyperreflexivity. The symptoms are expressions of a profound imbalance between the two dimensions of distinction and participation. They indicate that the patient still strives to be a distinct and participating self, but that she is not or has not been able to fulfill this goal given the circumstances of her present or past social embeddedness.

This thesis has raised several novel questions and I have outlined ways to answer them. However, to fully acknowledge the role of sociality for cognition requires a fundamental shift for cognitive science that will obviously invite manifold elaborations in future research. There are open issues and questions that arise directly for the enactive framework to cognition. But based on the ideas presented in this thesis we can already point to possible avenues for research in a practical and empirical context.

A definition of human autonomy in terms of social dynamics of distinction and participation invites us to reappraise the enactive key concepts and their interrelations in light of social interaction processes. In that vein we should work out in more detail the complex dynamics of intrinsic normative individual action and the autonomy of the social interaction process that the individual systems are involved in. How exactly do the two kinds of autonomies, i.e. the organismic autonomous system and the autonomy of the social interaction processes affect each other? What is the evolution of their joint dynamics over time and how are they altered in different social contexts (including pathologies)?

In line with this thesis I would suggest to approach these questions based on a back and forth between conceptual clarification and empirical research and practice. One possibility on the conceptual side is to reappraise phenomenological considerations in social phenomenology and anthropology to help inform the enactive perspective on social autonomy. With regards to the operational description of the dynamics involved in socially enacted autonomy we need to better understand how lower levels of autonomy relate to social autonomy. What is the contribution of bodily and neurological processes to realizing a socially enacted autonomous network? And how does sociality affect neurological or bodily processes? For this purpose one might reconsider existing empirical studies in light of the present proposal and assess whether it allows to integrate or re-interpret findings that seemed unrelated or remained unsatisfactorily explained. On the other hand the present proposal could inspire the development of new experimental designs and studies, for instance to improve patients' quality of life. In case of LIS we could modify the qualitative assessment of the patient's perspective and experience of bodily impairment by including questions that target the patient's self-awareness as a social being. This could be linked with a detailed assessment of the concrete social dynamics underlying the several social relations that patients are involved in (family, doctors, caregivers).

A similar approach might be adopted for the context of psychiatry and the treatment of mental disorders. One could apply the new conceptual distinctions to different pathologies and see whether they help to reassess several different symptoms as being linked at a more basic level. For example, rather than limiting the explanation of autism to disturbances at the neurological or

even sensorimotor level (Gallagher, 2006) one might compare autism with symptoms in schizophrenia in light of an imbalance between processes of distinction and participation. This may allow situating both disorders on a spectrum of disturbances in self-other relations. More concretely, one could assess how the temporal dynamics of social interactions is altered in autism and test whether it can be improved when adjusting the speed and quality of reactions to the patient's needs.

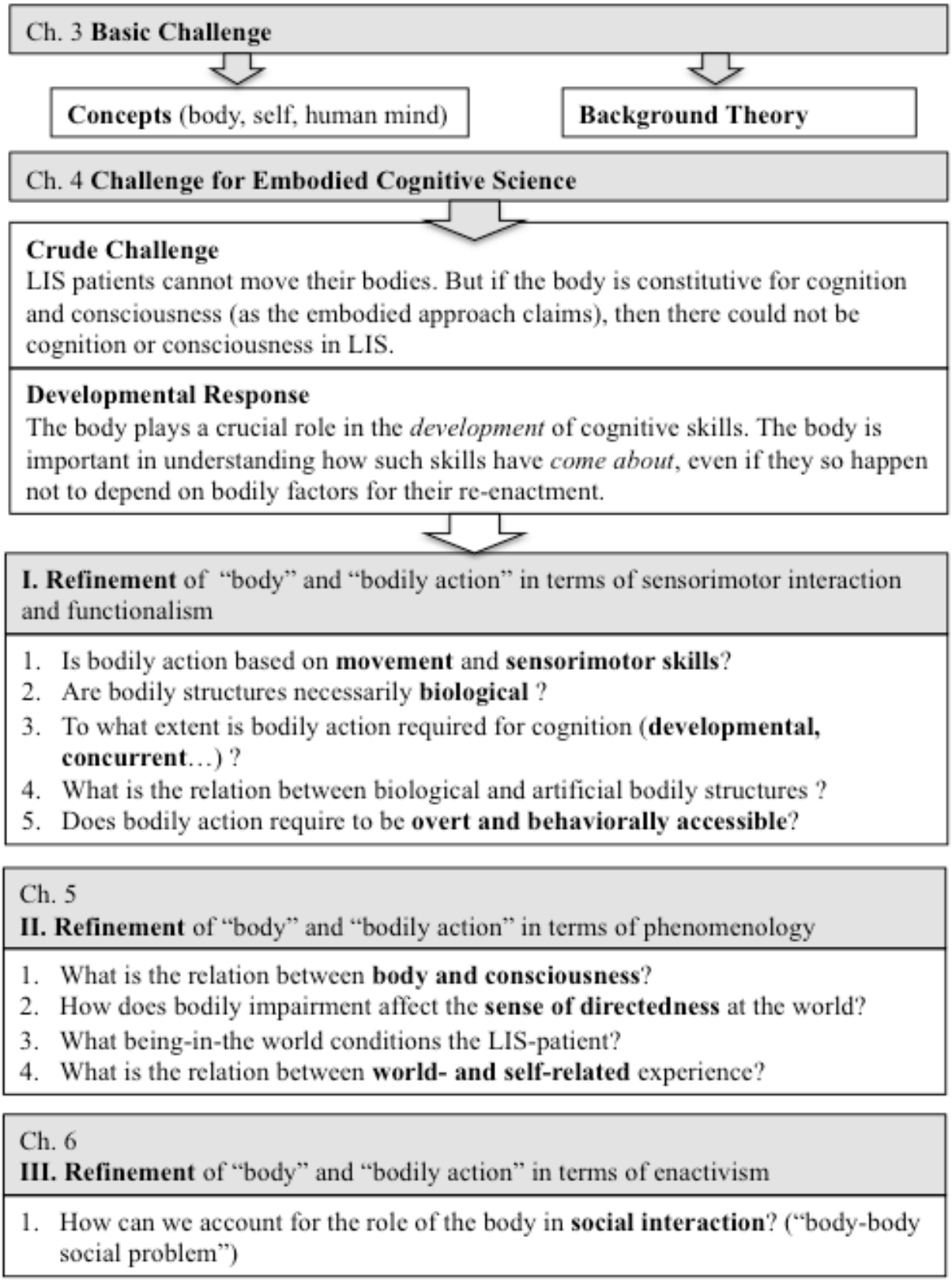
Reconsidering autism and schizophrenia in light of the present proposal also requires shedding light on an issue that I have not considered in this thesis, namely the relation between social autonomy and affect and emotions. In line with the present proposal, emotions would be considered as emerging in and for shared social engagement. They play a role in the regulation and co-regulation of our social, not merely bodily existence. Enactive cognition could explore how emotions relate to normativity and both self-affective and world-directed social experience and assess disorders in terms of regulation difficulties in both types of experiences.

Whether or not psychopathology can be usefully explained in terms of the present proposal could of course be seen in the actual therapeutic practice. If the maintenance of human existence depends on interactions with others and is not achieved by means of the organism in isolation, one might develop new forms of body-centred psychotherapy that re-establish a patient's disturbed sense of bodily self-presence in a joint process with others. The conjecture would be that these will be more effective for improving the patient's self-awareness and re-enabling her social abilities than individual therapy forms. This might apply for both the very interaction processes between therapist and patient (Seikkula et al., 2008, Seikkula and Olson, 2003) and other therapeutic means, such as somatic and dance therapy.

Contact improvisation dancers, for example, are always in bodily contact and they explore in spontaneous and joint interaction processes various ways in which bodies can engage and move together (Kaltenbrunner, 1998). The dance could be seen as a minimal form of the dynamics that arises based on the interplay of individual and interactive social processes. Crucially, in contact improvisation dancers "understand" each other in terms of bodily sensations and they experience their own agency and impulses to connect to the partner by touching or being touched. Touch is an important aspect factoring in human development and we overlook the fact that it is through direct bodily contact that human beings first encounter each other (Field, 1995, Ardiel and Rankin, 2010). It would be interesting to explore the therapeutic potential of contact improvisation dance by researching how interactions mediated by bodily touch can affect or improve not only the person's sense of embodiment but also her self-awareness as a social being. Based on the present thesis it could be hypothesized that repeated experiences of continuous resonance between the dancer's intended moves and the quality of dynamics of the dance can lead to positive affirmation of their sense of identity as being both a distinct and participating subject.

Last but not least, it would be interesting to look at socially enacted autonomy not only with regards to later stages of development or in cases of breakdown but also with respect to the initial development of early human autonomy. One avenue to pursue this endeavor could consist in examining to which extent studies in early infant development provide support for the proposal that human mind arises as continuous back and forth between social processes of distinction and participation.

Appendix: Overview of the Challenges



References

Adams, F., and K. Aizawa. (2009.) *The bounds of cognition*. Wiley: Blackwell.

Adams, F. and Aizawa, K. (2010). Defending the bounds of cognition. In Richard Menary (Ed.). *The Extended Mind*. Aldershot, Hants: Ashgate.

Albahari, M. (2007) *Analytical buddhism: The Two-tiered Illusion of Self*. New York: Palgrave Macmillan.

Anderson, M.L. (2003). Embodied cognition: a field guide. *Artificial Intelligence*. 149 (1): 91–130.

Ardiel, E. and Rankin, C. (2010). The importance of touch in development. *Journal of Paediatrics and Child Health* 15 (3): 153–156.

Azevedo, S.Z. (2006). "Perfect health" and the disembodiment of the self. An approach to Michel Henry's thought. In: Tymieniecka, A.-T. (Ed.), *Analecta Husserliana* L XXXIX: 323–338.

Bach-y-Rita P. (1972). *Brain Mechanisms in Sensory Substitution*, New York: Academic Press.

Baerveldt, C. and Verheggen, T. (1999). Enactivism and the experiential reality of culture: rethinking the epistemological basis of cultural psychology. *Culture Psychology* 5: 183–206.

Barbaras, R. (2011). Life and exteriority – the problem of metabolism, In: J. Stewart, O. Gapenne and E. A. Di Paolo (Eds.), *Enaction: Towards a New Paradigm for Cognitive Science*. Cambridge, MA: MIT Press: 89–123.

Barandiaran, X, Rohde, M., and Di Paolo, E. (2009). Defining agency: individuality, normativity, asymmetry and spatio-temporality in action. *Adaptive Behavior* 17: 367–386.

Barandiaran, X. and Egbert, M. (in press). Norm-establishing and norm-following in autonomous agency. *Artificial Life*.

Bauby, J.D., (1997). *The Diving Bell and the Butterfly*, London: Harper Perennial.

Bauer, G. F. Gerstenbrand, and E. Rumpl, (1979). Varieties of locked-in syndrome. *Journal of Neurology* 221: 77–91.

Benjamin, J. (1995). Recognition and destruction: An outline of intersubjectivity. In *Like Subjects, Love Objects: Essays on Recognition and Sexual Difference*. New Haven: Yale University Press.

Bennett, M. R., and Hacker, P.M.S, (2003). *Philosophical Foundations of Neuroscience*. Hoboken: John Wiley & Sons.

Bensch, M., Karim, A., Mellinger, J. Hinterberger, T., Tangermann, M., Bodgan, M., Rosenstiel, W. and Birbaumer, N. (2007). Nessi: An EEG-controlled web browser for severly paralyzed patients. *Computational Intelligence and Neuroscience*. doi:10.1155/2007/71863.

Birbaumer, N. et al. (2002). Brain-computer interfaces for communication and control. *Clinical Neurophysiology* 113: 767–791.

Birbaumer, N. et al. (2003). The thought-translation device (TTD): neurobehavioral mechanisms and clinical outcome. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 11 (2): 120–123.

Birbaumer, N. (2006). Breaking the silence: brain-computer interfaces (BCI) for communication and motor control. *Psychophysiology* 43 (6): 517–532.

Birbaumer, N. and Cohen, L.G. (2007). Brain–computer interfaces: communication and restoration of movement in paralysis. *Journal of Physiology* 579 (3): 621–636.

Blanke, O. and Metzinger, T. (2010). Full-body illusions and minimal phenomenal selfhood. In Fuchs, T., Sattel, H.C., and Henningsen, P. (Eds.), *The Embodied Self*. Stuttgart: Schattauer: 21–35.

Boudette, R. (2006). Yoga in the treatment of disordered eating and body image disturbance. How can the practice of yoga be helpful in recovery from an eating disorder? *Eating Disorders* 14: 167–170.

Braddon-Mitchell, D. and Jackson, F. (1996) *Philosophy of Mind and Cognition*. Oxford: Blackwell.

Brandom, R.B. (1994). *Making It Explicit. Reasoning, Representing, and Discursive Commitment*. Cambridge, MA: Harvard University Press.

Brooks, R.A. (1991). Intelligence without representation. *Artificial Intelligence* 47 (1–3): 139–160.

Bruno, M.-A., Pellas, F. and Laureys, S., (2008a). Quality of life in locked-in syndrome survivors, *Yearbook of Intensive Care and Emergency Medicine*: 881–890.

Bruno, M.-A., Bruno, M. A., Pellas, F., Bernheim, J. L., Ledoux, D., Goldman, S., Demertzi, A., et al. (2008b). Life with locked-in syndrome, *Revue Médicale De Liège* 63 (5–6): 445–451.

Bruner, J. and Kenney, H. (1966). The development of the concepts of order and proportion in children. In: Bruner, J.S., Oliver, R. and Greenfield, P.M. (Eds.), *Studies in Cognitive Growth*. New York: Wiley.

Carruthers, P. and Smith, P. K. (Eds.) (1996). *Theories of Theories of Mind*. Cambridge: Cambridge University Press.

Chisholm, N. and Gillett, G. (2005). The patient's journey: living with locked-in syndrome, *British Medical Journal* 331: 94–97.

Clark, A., and Chalmers, D. (1998). The extended mind. *Analysis* 58: 7–19.

Clark, A. (1997). *Being There: putting brain, body and world together again*. MIT Press

Clark, A. (1998) Magic Words: How Language Augments Human Computation. *Language and Thought*. In Boucher, J. and Carruthers, P., (Eds.) Cambridge: Cambridge University Press: 162-83

Clark, A., and R. Grush. (1999). Towards a cognitive robotics. *Adaptive Behavior* 7 (1): 5–16.

Clark, A. and Toribio, J. (2001). Sensorimotor chauvinism? Commentary on O'Regan, J.K. and Noë, A., A sensorimotor account of vision and visual consciousness, *Behavioral and Brain Sciences* 24: 979–980.

Clark, A. (2001). *Mindware*. Oxford: Oxford University Press.

Clark, A. (2008a). *Supersizing the mind: embodiment, action, and cognitive extension*. Oxford: Oxford University Press.

Clark, A. (2008b). Pressing the flesh: a tension in the study of the embodied, embedded mind?. *Philosophy and Phenomenological Research* 76: 37–59.

Clark, A. (2009). Spreading the joy: why the machinery of consciousness is (probably) still in the head. *Mind* 118: 963–993.

Cole, J. (2009). Impaired embodiment and intersubjectivity. *Phenomenology and the Cognitive Sciences* 8: 343–360.

Cole, J. (2010). Agency with impairments. In: Gallagher, S. and Schmicking, D. (Eds.) *Handbook of Phenomenology and Cognitive Science*. Springer: 655-670.

Colombetti, G. and Torrance, S. (2009). Emotion and ethics: an inter-(en)active approach. *Phenomenology and the Cognitive Sciences* 8: 505–526.

Dainton, B. (2004). The self and the phenomenal. *Ratio* 17 (4): 365–89.

De Haan, S. (2010). Comment: the minimal self is a social self. In: Fuchs, T., Sattel, H.C., and Henningsen, P. (Eds.), *The Embodied Self*. Stuttgart: Schattauer: 12–17.

Damasio, A. (2006). *The Feeling of What Happens*. Wilmington, MA: Mariner Books.

De Jaegher, H., Di Paolo, E. and Gallagher, S. (2010). Can social interaction constitute social cognition?. *Trends in Cognitive Science* 14 (10): 441–447.

De Jaegher, H., and Di Paolo, E. (2007). Participatory sensemaking: an enactive approach to social cognition. *Phenomenology and the Cognitive Sciences* 6 (4): 485–507.

De Jaegher, H. and Di Paolo, E. (2008). Making sense in participation: an enactive approach to social cognition. In: Morganti, F., Carassa, A., Riva G. (Eds.), *Enacting Intersubjectivity: A Cognitive and Social Perspective on the Study of Interactions*. Amsterdam, IOS Press: 33–49.

De Jaegher, H. and Froese, T. (2009). On the role of social interaction in individual agency. *Adaptive Behavior* 17(5): 444–460.

De Jaegher, H. (2010). Enaction versus representation: an opinion piece. In: T. Fuchs, H. Sattel and P. Henningsen (Eds.), *The Embodied Self: Dimensions, Coherence and Disorders*. Stuttgart: Schattauer.

Demertzi, A., Vanhaudenhuyse, A., Bruno, M-A., Schnakers, C., Boly, M., Boveroux, P., Maquet, Moonen, P., Laureys, S. (2008). Is there anybody in there? Detecting awareness in disorders of consciousness. *Expert Review of Neurotherapeutics* 8 (11): 1719-1730.

Dennett, D. (1992). The self as a center of narrative gravity. In: F. Kessel, P. Cole and D. Johnson, (Eds.), *Self and Consciousness*. Hillsdale, NJ: Erlbaum.

Diedrich, L. (2007). *Treatments: language, politics, and the culture of iIllness*. Minneapolis, MN: University of Minnesota Press.

Di Paolo, E. (2005). Autopoiesis, adaptivity, teleology, agency. *Phenomenology and the Cognitive Sciences* 4: 429 – 452.

Di Paolo, E. and Iizuka, H. (2008). How (not) to model autonomous behaviour. *Bio Systems* 91 (2): 409–423.

Di Paolo, E. (2009). Extended life. *Topoi* 28: 9–21.

Di Paolo, E., Rohde, M., and De Jaegher, H. (2010). Horizons for the enactive mind: values, social interaction and play. In: J. Stewart, O. Gapenne and Di Paolo, E. A. (Eds.), *Enaction: Towards a New Paradigm for Cognitive Science*. Cambridge, MA: MIT Press: 33–87.

Dobkin, B. (2007). Brain-computer interface technology as a tool to augment plasticity and outcomes for neurological rehabilitation. *The Journal of Physiology* 579: 637–642.

Drayson, Z. (2009). Embodied cognitive science and its implications for psychopathology. *Philosophy, Psychiatry & Psychology* 16 (4): 329–340.

Dreyfus, H. (1972). *What Computers Can't Do: A Critique of Artificial Intelligence*. New York: Harper and Row.

Dreyfus, H. and Dreyfus, S. (1996). The challenge of Merleau-Ponty's phenomenology of embodiment for cognitive science, In: Weiss, G. and Haber, F. (Eds.)*, Perspectives on Embodiment: the Intersections of Nature and Culture*. Routledge: 103–121.

Dudzinski, D. (2001). The diving bell meets the butterfly: identity lost and re-membered, *Theoretical Medicine and Bioethics* 2: 33–46.

Dworschak, M. (2009). Meine zweite Geburt. In: *Der Spiegel* 48, November 23, 2009.

Edelman, G.M. (1992). *Bright Air, Brilliant Fire: On the Matter of the Mind*. New York: Basic Books.

Eisenberger, N.I. (2011). Why rejection hurts: what social neuroscience has revealed about the brain's response to social rejection. In Decety, J. and Cacioppo, J. (Eds.), *The Handbook of Social Neuroscience*. New York, NY: Oxford University Press: 586–598.

Ekman, P. (1993). Facial expression and emotion. *American Psychologist* 48 (4): 384–392.

Fenton, A. and S. Alpert. (2008). Extending our view on using BCIs for locked-in syndrome. *Neuroethics* 1: 119–132.

Field, T.M., Woodson, R., Greenberg, R. and Cohen, D. (1982). Discrimination and imitation of facial expression by neonates. *Science* 218 (4568): 179–181.

Field, T. (1995). *Touch in Early Development*. London: Routledge.

Flynn, M.R, and Bush, J.W.M. (2008). Underwater breathing: the mechanics of plastron respiration. *Journal of Fluid Mechanics* 608: 275–296.

Fodor, J. (1975). *The Language of Thought*. Cambridge, Massachusetts: Harvard University Press.

Fodor, J. (1987). *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, MA: MIT Press.

Froese, T. and Di Paolo, E. (2009). Sociality and the life-mind continuity thesis. *Phenomenology of the Cognitive Sciences* 8 (4): 439–463.

Frith, C. and Frith, U. (1999). Interacting minds - A biological basis. *Science* 286: 1692–1695.

Fuchs, T. (2002). Der Begriff der Person in der Psychiatrie. *Nervenarzt* 73: 239–246.

Fuchs, T. (2005). Corporealized and disembodied minds: a phenomenological view of the body in melancholia and schizophrenia. *Philosophy, Psychiatry & Psychology* 12 (2): 95–107.

Fuchs, T. (2007). Fragmented selves: temporality and identity in borderline personality disorder. *Psychopathology* 40: 379–387.

Fuchs, T. and De Jaegher, H. (2009). Enactive intersubjectivity: participatory sense-making and mutual incorporation. *Phenomenology of the Cognitive Sciences* 8: 465–486.

Fuchs, T. and Schlimme, J. (2009). Embodiment and psychopathology: a phenomenological perspective. *Current Opinion in Psychiatry* 22: 570–575

Fuchs, T. (2011). The brain as a mediating organ. *Journal of Consciousness Studies* 18 (7-8): 196–221.

Gadamer, H.G., (1960/2007). *Wahrheit und Methode*. In: Figal, G. (Ed.), *Klassiker Auslegen, Bd. 30 – Hans-Georg Gadamer: Wahrheit und Methode*. Berlin: Akademie-Verlag.

Gallagher, S. and Cole, J. (1995). Body schema and body Image in a deafferented subject. *Journal of Mind and Behavior* 16: 369–390.

Gallagher, S. (2000). Philosophical conceptions of the self: implications for cognitive science. *Trends in Cognitive Science* 4 (1): 14–21.

Gallagher, S. (2001). The practice of mind: theory, simulation or primary interaction?. *Journal of Consciousness Studies* 8 (5-7): 83–108.

Gallagher, S. (2006). *How the Body Shapes the Mind*. New York: Oxford University Press.

Gallagher, S. and Schmicking, D. (Eds.) (2010). *Handbook of Phenomenology and Cognitive Science*. Dordrecht: Springer.

Gallagher, S. (2010). Phenomenology and non-reductionist cognitive science. In: Gallagher, S. and Schmicking, D. (Eds.). *Handbook of Phenomenology and Cognitive Science*. Dordrecht: Springer.

Gallese, V. and Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences* 2 (12): 493–501.

Gapenne O., Rovira, K., Ali Ammar, A. and Lenay C. (2003). Tactos: special computer interface for the reading and writing of 2D forms in blind people. In: Stephanidis, C. (Ed.), *Universal Access in HCI, Inclusive Design in the Information Society*. London: Lawrence Erlbaum Associates: 1270–1274.

Ginsburg, C. (2005). First-person experiments. *Journal of Consciousness Studies* 12 (2): 22–42.

Gillett, G. and Chisholm, N. (2007). Locked in syndrome, PVS and ethics at the end of life. *Journal of Ethics in Mental Health* 2 (2): 1–6.

Goenka, S. N. (2000). *The discourse summaries: talks from a ten-day course in Vipassana meditation*, Pariyatti.

Goode, G.B. (1962). Sleep paralysis. *Archives of Neurology* 6 (3): 228–234.

Gosseries, O., Bruno, M.A., Vanhaudenhuyse, A., Laureys, S. and Schnakers, C. (2009). Consciousness in the locked-in syndrome. In: S. Laureys and G. Tononi (Eds.), *The Neurology of Consciousness*. Oxford: Elsevier: 191–203.

Gould, S.J. and Lewontin, R.C. (1979). The spandrels of San Marco and the panglossian paradigm: a critique of the adaptationist programme. *Proceedings of the Royal Society of London. Series B. Biological Sciences* 205 (1161): 581 –598.

Grassian, S. (1983). Psychopathological effects of solitary confinement. *American Journal of Psychiatry* 140 (11): 1450–1454.

Gurwitsch, A. (1964). *The field of consciousness*. Pittsburgh, PA: Duquesne University Press.

Hanna, R. and Thompson, E. (2003). The mind-body-body problem. *Theoria et Historia Scientiarum: International Journal for Interdisciplinary Studies* 7 (1): 23–42.

Hanna, R. and Maiese, M. (2009). *Embodied Minds in Action*. New York: Oxford University Press.

Harvey, I., Di Paolo, E., A., Wood, R., Quinn, M., and Tuci, E. (2005). Evolutionary robotics: a new scientific tool for studying cognition. *Artificial Life* 11: 79–98.

Haugeland, J. (1985). *Artificial Intelligence: The Very Idea*. Cambridge, MA: MIT Press.

Havelange, V. (2011). The ontological constitution of cognition and the epistemological constitution of cognitive science. In: Stewart, J., Gapenne, O. and Di Paolo, E. A. (Eds.). *Enaction: Towards a New Paradigm for Cognitive Science*. Cambridge, MA: MIT Press: 335–361.

Hayward, J. (1998). A Rdzogs-chen buddhist interpretation of the sense of self. *Journal of Consciousness Studies* 5 (5–6): 611–626.

Hegel, G.W.F., (1807/1986). *Phänomenologie des Geistes*. Frankfurt am Main: Suhrkamp.

Heidegger, M. (2001/1927). *Sein und Zeit*. Tübingen: Niemeyer.

Heidegger, M. (1929–30/1992). *Die Grundbegriffe der Metaphysik. Welt – Endlichkeit – Einsamkeit*. Frankfurt am Main: Vittorio Klostermann.

Henry, M. (1995). Le corps vivant. *Cahiers de l'Ecole des sciences philosophiques et religieuses*, 18; repris In : *Auto-Donation*. Prétentaine, 2001.

Henry, M. (1999). Material Phenomenology and language (or, pathos and language). *Continental Philosophy Review* 32 (3): 343–365.

Henry, M. Qu'est-ce que cela que nous appelons la vie? (1978). *Philosophiques* 5 (1): 133-150.

Henry, M. (1988). The critique of the subject. *Topoi* 7: 147–153.

Henry, M. (2004). *Auto-donation: Entretiens et conférences*. Paris: Editions Beauchesne.

Hermans, H.J.M. Kempen, J.G. and Van Loon, R.J.P. (1992). The dialogical self. *American Psychologist* 47 (1): 23–33.

Hume, D. (1739). *A Treatise of Human Nature*, Book I, Part IV, Of Personal Identity. Retrieved: 19.3. 2012 from http://oll.libertyfund.org/index.php?option=com_staticxt&staticfile=show.php%3Ftitle=342&Itemid=27.

Humphrey, N. (2007). The society of selves. *Philosophical Transations of the Royal Society B* 362*:* 745–754.

Hurley, S. (1998). *Consciousness in Action*. Cambridge, MA: Harvard University Press.

Hutchins, E. (2011). Enaction, imagination, and insight, In: Stewart, O., Gapenne, J. and Di Paolo, E. (Eds.), *Enaction: Towards a New Paradigm for Cognitive Science*. Cambridge, MA: MIT Press: 425–450.

Husserl, E. (1992/1930*)*. Ideen zu einer reinen Phaenomenologie und phaenomenologischen Philosophie 1: Allgemeine Einfuehrung in die reine Phaenomenologie. Text nach Husserliana III/1 und V. In: Ströker, Elisabeth. (Ed.). *Bd. 5 Gesammelte Schriften/ Edmund Husserl*. Hamburg: Meiner.

Hutto, D. (2005). Knowing what? Radical vs. conservative enactivism. *Phenomenology and the*

Cognitive Sciences 4: 389–405.

Hutto, D. (2010). Radical enactivism and narrative practice: implications for psychopathology. In: Fuchs, T., Sattel, H.C., and Henningsen, P. (Eds.). *The Embodied Self.* Stuttgart: Schattauer: 43–66.

Inci, S., and T. Özgen. (2003). Locked-in syndrome due to metastatic pontomedullary tumor—case report. *Neurologia Medico-Chirurgica* 43: 497–500.

Ikäheimo, H. (2009). A vital human need: recognition as inclusion in personhood. *European Journal of Political Theory* 8: 31–45.

James, W. (1890). *The Principles of Psychology*, Vol. 1 (Holt and Macmillan, New York, London), chapter 15: The Perception of Time. Retrieved: 19.03.2012 from 'Classics in the History of Psychology. An internet resource developed by Christopher D. Green. Retrieved: 25.07.201 from http://psychclassics.asu.edu/James/Principles/prin15.htm.

Jeannerod, M. (2009). The sense of agency and its disturbances in schizophrenia: a re-appraisal. *Experimental Brain Research* 192: 527–532.

Jonas, H. (1966). *The Phenomenon of Life: Toward a Philosophical Biology*. Chicago: University of Chicago Press.

Kaltenbrunner, T. (1998). *Contact Improvisation*. Aachen: Meyer und Meyer Verlag.

Koch, S., Morlinghaus, K. and Fuchs. T. (2007). The joy dance. Specific effects of a single dance intervention on psychiatric patients with depression. *The Arts in Psychotherapy* 34: 340–349.

Krueger, J. (2011). The who and the how of experience. In: Zahavi, D., Thompson, E. and Siderits, M. (Eds.). *Self, No-Self? Perspectives from Analytical, Phenomenological, and Indian Traditions*. Oxford: Oxford University Press: 27–55.

Krueger, J. (forthcoming). Empathy, enaction, and shared musical experience. In: Cochrane, T., Fantini, B., and Scherer, K.R. (Eds.), *The Emotional Power of Music: Multidisciplinary Perspectives on Musical Expression, Arousal and Social Control*, Oxford: Oxford University Press.

Kübler, A., Furdea, A., Halder, S., and Hösle, A. (2008). Brain painting – BCI meets art. In: Müller-Putz, G. R., Brunner, C., Leeb, R., Pfurtscheller, G. and Neuper, C. (Eds.), *4th International Brain-Computer Interface Workshop and Training Course*. Graz University of Technology, Austria: Verlag der Technischen Universität Graz: 361–366.

Kuhn, T. (1962/1970). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Kurthen, M., Moskopp, D., Linke, D. and Reuter, B. (1991). The locked-in syndrome and the behaviorist epistemology of other minds. *Theoretical Medicine and Bioethics* 12: 69– 79.

Kyselo, M. (2011). Locked-in syndrome and BCI – towards an enactive approach to the self. *Neuroethics* (electronic publication ahead of print). doi: 10.1007/s12152-011-9104-x.

Kyselo, M. and Walter, S. (2009). Review: supersizing the mind. *Philosophical Psychology* 22: 803 – 807.

Kyselo, M. and Di Paolo, E. (in preparation). Through the enactive eye: Locked-in syndrome as a challenge for embodied cognition.

Lakoff, G. and Johnson, M. (1999). *Philosophy in the flesh: the embodied mind and its challenge to Western thought*. New York: Basic Books.

Laureys, S., Pella, F., Van Eeckhout, P., Ghorbel, S. Schnakers, C. Perrin, F., Berré, J., Fayonville, M., Pantke, K.., Damas, F., Lamy, M. Moonen, G. and Goldman, S. (2005). The locked-in syndrome: what is it like to be conscious but paralyzed and voiceless? *Progress in Brain Research* 150: 495–511.

Laureys, L, Perrin, F. and Brédart, S. (2007) Self-consciousness in non-communicative patients. *Consciousness and Cognition* 16 (3): 722–741.

Laureys, S. and Boly, M. (2007). What is it like to be vegetative or minimally conscious?. *Current Opinion in Neurology* 20: 609–613.

Lavigne, J.-F. (2009). The paradox and limits of Michel Henry's concept of transcendence. *International Journal of Philosophical Studies* 17 (3): 377–388

León-Carrión, J. P. van Eeckhout, and M., Domínguez-Morales, M.R. (2002). The locked-in syndrome: a syndrome looking for a therapy. *Brain Injury* 16: 555–569.

León-Carrión, J. P. van Eeckhout, and M. R. Domínguez-Morales (2005). The locked-in syndrome: a challenge for therapy. *Acta Neurochirurgica Supplementum* 93: 213–216.

Lewis, D. (1980). Mad pain and martian pain. In: I. Block. N. (Ed.), *Readings in the Philosophy of Psychology*. Cambridge, MA: Harvard University Press: 216–222.

Lotze, M. and Moseley, G.L. (2007). Role of distorted body image in pain, *Current Rheumatology Reports* 9 (6): 488–496.

Luhmann, N. (1992). *Die Wissenschaft der Gesellschaft*, Frankfurt am Main: Suhrkamp.

Lulé, D., Zickler, C., Häcker, S., Bruno, M.A., Demertzi, A., Pellas, F., Laureys, S, Kübler, A. (2009). Life can be worth living in locked-in syndrome. *Progress in Brain Research* 177: 339–351.

Lutz, A. and Thompson, E. (2003). Neurophenomenology: Integrating subjective experience and brain dynamics in the neuroscience of consciousness. *Journal of Consciousness Studies* 10 (9–10): 31–52.

Lutz, A. (2002). Toward a neurophenomenology as an account of generative passages: A first empirical case study. *Phenomenology and the Cognitive Sciences* 1 (2): 133–167.

Malcolm, N. (1962). Knowledge of Other Minds", in Chappell, V.C. (Ed.) *The Philosophy of Mind*, Englewood Cliffs: Prentice-Hall.

Maturana, H. R. and Varela, F. J. (1980). Autopoiesis and cognition: The realization of the living. *Boston Studies in the Philosophy of Science* 42, Dordrecht: D. Reidel.

Maturana, H. and Varela, F. J. (1987). *The tree of knowledge: The biological roots of human understanding*. Shambhala: Boston, MA.

McCarthy, J. and Hayes, P.J. (1969). Some philosophical problems from the standpoint of artificial intelligence. *Machine Intelligence* 4: 463–502.

McCulloch, W. S. and Pitts, W. (1943). A logical calculus of ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics* 5: 115–133.

McGinn, C. (1984). The concept of knowledge. *Midwest Studies in Philosophy* 9 (1): 529–554

Mead, G.H. (1934). *Mind, Self and Society*. Morris, C.W. (Ed.), Chicago: University of Chicago Press.

Menary, R. (2010). *The extended mind*. MIT Press.

Merleau-Ponty (2002/1945). *Phenomenology of Perception*. London: Routledge.

Metzinger. T. (2004). *Being No One: The Self-Model Theory of Subjectivity*. Cambridge, MA: MIT Press.

Minsky, M. and Papert, S. (1969). *Perceptrons: An introduction to computational geometry*. Cambridge, MA: MIT Press.

Murphy, R.F. (1990). *The Body Silent*. New York: Norton.

Newell, A. and Simon, H. (1963) GPS: A program that simulates human thought. In: E. Feigenbaum and J. Feldman (Eds.). *Computers and Thought*: 279–293.

Nizzi, M-C., Demertzi, A., Gosseries, O., Bruno, M-A., Jouen, F., Laureys, S. (2012). From armchair to wheelchair: how patients with a locked-in syndrome integrate bodily changes in experienced identity. *Consciousness and Cognition* 21 (1): 431–437.

Noë, A. (2004). *Action in perception*. Cambridge, MA: MIT press.

Núñez, R. (2004). Do real numbers really move? Language, thought, and gesture: the embodied cognitive foundations of mathematics, In: Iida. F, Pfeifer, R., Steels, L. and Kuniyoshi, Y., (Eds.), *Embodied Artificial Intelligence*. Berlin Heidelberg: Springer: 54–73.

Olson, E.T. (1998). There is no problem of the self, *Journal of Consciousness Studies* 5 (5–6): 645–657.

O'Regan, K. and A. Noë. (2001). A sensorimotor account of vision and visual consciousness, *Behavioral and Brain Sciences* 24: 939–1031.

Owen, A., Coleman, M.R., Boly, M., Davis, M.H., Laureys, S. and Pickard, J.D. (2006). Detecting awareness in the vegetative state. *Science* 313: 1402.

Parnass, J. and Sass, L.A. (2010). The spectrum of schizophrenia. In: Fuchs, T., Sattel, H.C., and Henningsen, P. (Eds.), *The Embodied Self*. Stuttgart: Schattauer: 227–243.

Perlis, D. (1999). Consciousness as self-function. In: Gallager, S. and Shear, J. (Eds.),. *Models of the Self*. Exeter: Imprint Academic: 131–149.

Perrin, F., Schnakers, C., Schabus, M., Degueldre, C., Goldman, S., Brédart, S., Faymonville, M.E., Lamy, M., Moonen, G., Luxen, A., Maquet, P. and Laureys, S. (2006). Brain response to one's own name in vegetative state, minimally conscious state, and locked-in syndrome. *Archives of Neurology* 63 (4): 562–569.

Petitmengin, C. (2006a). L'énaction comme éxperience vécue. *Intellectica* 43: 85–92.

Petitmengin, C. (2006b). Describing one's subjective experience in the second person: an interview method for the science of consciousness. *Journal of Consciousness Studies* 14 (3): 54–82.

Pieper, B. and Clénin, D. (2010). Verkörperte Selbst- und Fremdwahrnehmung sozialen Handelns. Eine praktisch-theoretische Forschungsperspektive. In Böhle, F. und Weihrich, M. (Eds.), *Die Körperlichkeit sozialen Handelns. Soziale Ordnung jenseits von Normen und Institutionen*. Bielefeld: transcript-Verlag: 261 296.

Plessner, H. (1981). *Die Stufen des Organischen und der Mensch. Gesammelte Schriften IV*. Frankfurt am Main: Suhrkamp.

Plum, F. and Posner, J.B. (1966). *The Diagnosis of Stupor and Coma*. Philadelphia: F. A. Davis.

Port, R. and van Gelder, T., (Eds.) (1995). *Mind as Motion: Explorations in the Dynamics of Cognition*. Cambridge, MA: MIT Press.

Prinz, J. (2009). Is consciousness embodied?. In: Robbins, P. and Aydede, M. (Eds). *The Cambridge handbook of situated cognition*. Cambridge: Cambridge University Press, Retrieved: 19.03.2012 from http://www.passions.de/situated_cognition.pdf#page=423.

Putnam, H. (1975). Philosophy and our mental life. *Mind, Language, and Reality*. Cambridge Cambridge: University Press: 291-303.

Reddy, V. (2003). On being the object of attention: implications for self-other consciousness. *Trends in Cognitive Sciences* 7 (9): 397–402.

Reid, L. (1998). Wittgenstein's ladder: the tractatus and nonsense. *Philosophical Investigations* 21 (2): 97–151.

Rietveld, E. (2008). Situated normativity: the normative aspect of embodied cognition in unreflective action. *Mind* 117 (468): 973–1001.

Rohde, M. (2011). *Enaction, Embodiment, Evolutionary Robotics. Simulation Models for a Post-Cognitivist Science of Mind*. Paris: Atlantis Press Book.

Rohrer, T. (2007). The body in space: dimensions of embodiment. In: Ziemke, T., Zlatev, J. and Roslyn, M.F. (Eds), *Body, Language and Mind 1. Embodiment*. Berlin/New York: Mouton de Gruyter: 339–377.

Röhricht, F. (2009). Body oriented psychotherapy. The state of the art in empirical research and evidence-based practice: a clinical perspective. *Body, Movement and Dance in Psychotherapy* 4 (2): 135–156.

Rosenblatt, Frank (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Reviews* 65: 386–408.

Rowlands, M. (2009). Enactivism and the extended mind. *Topoi* 28 (1): 53–62.

Rummelhart, D., Smolensky, P. McClelland, J. and Hintin, G. (1986). Schemata and sequential thought processes in PDP models. In J. McClelland, D. Rummelhart, and the PDP Research Group. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition,* 2. Cambridge, MA: MIT press.

Rupert, R. D. (2004). Challenges to the hypothesis of extended cognition. *Journal of Philosophy* 101: 389–428.

Rupert, R.D. (2010). Cognitive systems and the supersized mind. *Philosophical Studies* 152 (3): 427–436.

Schnakers, C., Majerus, S., Goldman, S., Boly, M. Van Eackhout, P., Gay, S., Pellas, F., Bartsch, V. Peigneux, P., Moonen, G., and Laureys, S. (2008). Cognitive function in the locked-in syndrome. *Journal of Neurology* 255: 323–330.

Schütz, A. and Luckmann, T. (2003) *Strukturen der Lebenswelt*. Stuttgart: Utb.

Schwartz, A.B, (2007). Useful signals from motor cortex. *Journal of Physiology* 579 (3): 581–601.

Searle, J. (1980). Minds, brains and programs. *Behavioral and Brain Sciences* 3: 417–424

Seikkula, J. and Olson, M. (2003). The open dialogue aporoach to acute psychosis: its poetics and micropolitics. *Family Process* 42 (3): 403–418

Seikkula, J. (2008). Inner and outer voices in the present moment of family and network therapy. *Journal of Family Therapy* 30: 478–491

Silberstein, M. and Chemero, A. (2012). Complexity and extended phenomenological-cognitive systems. *Topics in Cognitive Science* 4 (1): 35–50.

Smithers, T. (1997). Autonomy in robots and other agents. *Brain and Cognition* 34: 88–106.

Sprevak, M. (2009). Extended cognition and functionalism. *Journal of Philosophy* 106 (9): 503–527.

Sprevak, M. (2010). Inference to the hypothesis of extended cognition. *Studies in History and Philosophy of Science* 41: 353–362.

Stawarska, B. (2009). *Between You and I: Dialogical Phenomenology*. Athens: Ohio University Press.

Stephan, A. (2006). The dual role of 'emergence' in the philosophy of mind and in cognitive science. *Synthese* 151: 485–498.

Stevens, S. (1951). *Handbook of Experimental Psychology*. New York: Wiley

Strawson, G. (1999). The self. In: Gallagher, S. and Shear, J. (Eds.). *Models of the Self*. Exeter: Imprint Academic.

Svenaeus, F. (2000). The body uncanny – further steps towards a phenomenology of illness. *Medicine, Health Care and Philosophy* 3 (2): 125–37.

Schwartz, J. and Estrint, J. (2004). Living for today, locked in a paralyzed body, In: *The New York Times*, November 7, 2004.

Tani, J. (1998) An interpretation of the "self" from the dynamical systems perspective: a constructivist approach. *Journal of Consciousness Studies* 5 (5-6): 516–542.

Thompson, E. and Varela, F. (2001). Radical embodiment: neural dynamics and consciousness. *Trends in Cognitive Science* 5: 418–425.

Thompson, E. (2005). Sensorimotor subjectivity and the enactive approach to experience. *Phenomenology and the Cognitive Sciences* 4: 407–427.

Thompson, E. (2007). *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*. Cambridge, MA: Harvard University Press.

Thompson, E. and Stapleton, M. (2009). Making sense of sense-making: reflections on enactive and extended theories. *Topoi* 28: 23–30.

Thompson, E. and Henry, A (2011). Witnessing from here: self-awareness from a bodily versus embodied perspective. In: Gallagher, S. (Ed.), *The Oxford Handbook of the Self*. Oxford: Oxford University Press.

Thompson, E., Lutz, A. and Cosmelli, D. (2005) Neurophenomenology: an introduction for neurophilosophers. In: Brook, A. and Akins, K. (Eds.). *Cognition and the Brain: The Philosophy and Neuroscience Movement*. New York and Cambridge, Cambridge University Press: 1–66.

Thompson, E. and Cosmelli, D. (in press). Brain in a vat or body in a world? Brainbound versus enactive views of experience.

Toombs, S.K. (1988). Illness and the paradigm of lived body. *Theoretical Medicine* 9: 201–226.

Toombs, S.K. (1993). *The Meaning of Illness: A Phenomenological Account of the Different Perspectives of Physician and Patient*. Dordrecht: Kluwer.

Torrance, S. (2005). In search of the enactive. *Phenomenology and the Cognitive Sciences* 4 (4): 357–368.

Townsend, J.T. and Busemeyer, J. (1995). Dynamic representation of decision-making. In: Port, R. and van Gelder, T. (Eds.). *Mind as Motion: Explorations in the Dynamics of Cognition*. Cambridge, MA: MIT Press.

Van den Berg, J. (1972): *A different existence. Principles of phenomenological psychopathology*. Pittsburgh, PA: Duquesne University Press.

Varela, F.J. (1979). *Principles of Biological Autonomy*. New York: Elsevier North Holland.

Varela, F.J. (1995). Resonant cell assemblies: a new approach to cognitive functions and neuronal synchrony. *Biological Research* 28: 81–95.

Varela, F.J, and Shear, J. (1999). First-person methodologies: what, why, how?. *Journal of Consciousness Studies* 6: 1–14.

Varela, F.J., Thompson, E., and Rosch, E. (1993). *The embodied mind: cognitive science and human experience*. Cambridge, MA: MIT Press.

Varela, F.J. 1997. Patterns of life: Intertwining identity and cognition. *Brain and Cognition* 34: 72–87.

Van Gelder, T. (1998). The dynamical hypothesis in cognitive science. *Behavioral and Brain Sciences* 21 (5): 615–665.

Vogeley, K., Bussfeld, P., Newen, A., Herrmann, S., Happe, F., Falkai, P., Maier, W., Zilles, K. (2001) Mindreading: neural mechanisms of theory of mind and self-perspective. *NeuroImage* 14:170–181

Von Uexküll,. J. (1958). *Streifzüge durch die Umwelten von Tieren und Menschen. Bedeutungslehre.* Hamburg: Rowohlt.

Vygotsky, L. (1986). *Thought and Language*. Cambridge, MA: MIT Press.

Walter, S. (2010). Locked-in syndrome, BCI, and a confusion about embodied, embedded, extended, and enacted cognition. *Neuroethics* 1: 61–72.

Weber A., and Varela, F. (2002). Life after Kant: natural purposes and the autopoietic foundations of biological individuality. *Phenomenology and the Cognitive Sciences* 1: 97–125.

Wheeler, M. and Clark, A. (1999). Genic representation: reconciling content and causal complexity. *British Journal for the Philosophy of Science* 50: 103–135.

Wheeler, M. (2005). *Reconstructing the Cognitive World: The Next Step*. Cambridge, MA: MIT Press.

Wheeler, M. (2008). Cognition in context: phenomenology, situated robotics and the frame problem. In: Ratcliffe, M. and Gallagher, S. (Eds.), *International Journal of Philosophical Studies* 16 (3), special issue on Situated cognition: perspectives from phenomenology and science: 323–349.

Wheeler, M. (2010a). Mind, things, and materiality. In: Renfrew, C., and Malafouris, L. (Eds.), *The Cognitive Life of Things: Recasting the Boundaries of the Mind*, Cambridge: McDonald Institute for Archaeological Research Publications.

Wheeler, M. (2010b). In defense of extended functionalism. In: Richard Menary (Ed). *The extended mind.* Cambridge, MA: MIT Press.

Wheeler, M. (in press) Mind in life or life in mind? Making sense of deep continuity. *Journal of Consciousness Studies.*

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review* 9 (4): 625–636.

Wilson, R. A. (2004). *Boundaries of the Mind: The Individual in the Fragile Sciences: Cognition*. Cambridge, New York: Cambridge University Press.

Wittgenstein, L. (1953/1967). *Philosophical Investigations*. Translated by G. E. M. Anscombe. Oxford: Blackwell.

Wolpaw, J. N. Birbaumer, D. McFarland, G. Pfurtscheller, and T. Vaughan. (2002). Brain-computer interfaces for communication and control. *Clinical Neurophysiology* 113: 767–791.

Wolpaw, J. (2007). Brain–computer interfaces as new brain output pathways. *Journal of Physiology* 579 (3): 613–619.

Young, I. M. (1980), Throwing like a girl: A phenomenology of feminine body comport-ment motility and spatiality. *Human Studies* 3 (2):137-156.

Zahavi, D. (1999). Michel Henry and the phenomenology of the invisible. *Continental Philosophy Review* 32: 223–240.

Zahavi, D (2005). *Subjectivity and selfhood: investigating the first-person perspective*. Cambridge, MA: MIT Press.

Zahavi, D. (2007). Subjectivity and immanence. In: Grøn, A., Damgaard, I. Overgaard, S. (Eds.), *Subjectivity and transcendence.* Mohr Siebeck, Tübingen: 133–147.

Zahavi, D. (2009). Is the self a social construct? *Inquiry: An Interdisciplinary Journal of Philosophy* 52 (6): 551–573.

Zahavi, D. (2010). A distinction in need of refinement. In: Fuchs, T., Sattel, H.C., and Henningsen, P. (Eds.). *The Embodied Self.* Stuttgart: Schattauer: 3–11.

Zaner, R. (2003). Sisyphus without knees: exploring self-other relationships through illness and disability. *Literature and Medicine* 22 (2): 188–207.

Ziemke, T. (2003). What's that thing called embodiment? *Proceedings of the 25th Annual Meeting of the Cognitive Science Society*. Hillsdale, NJ: Lawrence Erlbaum: 1305–1310.