Master of Research Thesis

Understanding the mechanics of the 'self' (and consciousness) through the use of sensory-substitution-devices

Research question

How does integration of sensory-substitution-devices (SSDs) into the body-schema provide information to generate a theory of the mechanics of the 'self' and its role in consciousness?

Hugo Rodriguez

(MRes 30111269)

Submitted on 24 April 2018

Faculty of Arts

Department of Philosophy

Macquarie University

Table of contents

Statement of Certification

This thesis is submitted as a requirement for the degree of Master of Research at Macquarie University. I declare that it has not been previously submitted, in part or in whole, in any formal course of study, and that it is entirely my original work. I certify that to the best of my abilities all sources of information have been acknowledged and all references documented.

Signed:

Abstract

The main purpose of this thesis is to offer grounds for a possible theory of the self and consciousness based on research on sensory-substitution-devices (SSDs) – devices that convert signals from one bodily sense, vision for example, into signals that can be read in a different modality, such as tactile vibrations. A main problem of traditional views of the self has been conflating the experiencer with the experienced and I propose, as a possible solution, the existence of two selves, an *observer-self* that emerges from the synergetic sum total of brain/body functioning, and an *observed-self*, the unconscious house of our perceptions, personal identity and history. I suggest that this is possible because the selves can be thought of as essentially processing-sensors with the former being the ontological first-person 'I', converting biological information into phenomenology, and the latter converting sensorial inputs into biological information.

The research method I adopted is philosophical. It comprises literature review to critically deduce how incorporation of SSDs into the body-schema transforms the observed-self of the user. The process involves four main factors, which when viewed together indicate that SSDs generate a novel sense. I differentiate the body-schema and the body-model and I identify four stages of SSDs integration, which changes the feeling of body ownership. Integration creates the separateness of 'attributions to oneself', that is, our observed-self, from 'distal attributions', which we perceive as non-self.

I then propose a *schematic representation of consciousness* that shows consciousness as a full-spectrum (including unconsciousness) and how brain activity generates our field-of-conscious-awareness, with effectors marking the boundaries of the observed-self. The schema depicts information as travelling *unidirectionally*; afferent downward signals from the environment (global-to-local), and efferent local-to-global upward output generated by our thoughts and intentions. SSDs modify the observed-self because they act as sensoria inputs as part of the downward process.

I close the thesis by suggesting that further research should explore how theories such as *autopoiesis* and *emergentism* can assist explaining and completing the model of consciousness outlined here.

Introduction

This thesis is an attempt to offer suggestions for a possible theory of the self and the role it plays in consciousness based on studies on sensory-substitution-devices (SSDs). In Chapter 1 I discuss traditional notions of the self and the need to consider consciousness as comprising two selves, an *observer-self* and an *observed-self*. I offer grounds to treat them essentially as processing-sensors with the latter arising from afferent/efferent nerves circuitries and sensing the environment, and the former perceiving this field-of-awareness phenomenologically. I then comment on how the observed-self operates within the body-schema and the body-model.

Chapter 2 is a review of studies on SSDs and their integration into the body-schema. Here I describe how degrees of integration can lead to 'brain exaptation' and modifications on the observed-self, and how these studies provide material for a hypothetical model of the observed-self and consciousness.

Chapter 3 is an attempt to provide evidence to support the notion that the observed-self is primarily located within the boundaries of the physical body. Based on this evidence I propose a schematic model of consciousness and the role that the observed-self plays in it. I close this Chapter with suggestions of how SSDs modify our cognitive capacities as well as the observed-self.

In the last Chapter I use a real-life example to illustrate how the model is assumed to work, and I offer tentative definitions for consciousness and the observed-self. I then identify likely strengths and weaknesses of the speculative model offered in this thesis, and suggest likely issues for further research based on these definitions.

What are sensory-substitution-devices?

The idea of sensory substitution was introduced by Paul Bach-y-Rita in the 1960s (Bach-Y-Rita et al., 1969) and the primary purpose was to use one sensory modality, touch for example, to process information typically accessed by another modality such as vision. The device constructed was a system comprising three main parts, which are also the fundamental components of all modern devices. A *sensor* such as a video camera, a microphone or an ultrasound gun first detects environmental objects in one modality, and then transmits these inputs to an electronic *coupling system* that interprets and converts this information into signals that can be perceived by a *stimulator***,** which executes them in a different modality. A typical SSD is a 'Tactile-visual-substitution-system' (TVSS). A video camera (the sensor) captures objects visually and the coupling system converts these signals into electro-vibration, 'tactile images' that can be then sensed through the stimulator as vibrations on the skin (Collins & Bach-y-Rita, 1973). Blind or blindfolded users of the device can quickly learn to detect simple targets and orient themselves by discriminating different stimulations on the skin. Since then, SSDs improved greatly and have been used in research fields including cognitive neuroscience, brain functioning, and rehabilitation. Modern devices use a variety of sensors and stimulators that include electromagnetic, electro-tactile, shape memory alloys (a piece of metal with properties to 'remember' its original shape), piezo-electric mechanisms (that generate an electrical charge in response to mechanical stress), compressed air, and constraint gauge systems (See Lenay et al., 2003).

Another popular SSD has been the image-to-sound device known as the 'vOICe' (the middle letters stand for 'oh I see'). Developed by medical doctor Peter Meijer (Ward & Meijer, 2010) the device transduces images from a camera into soundscapes comprising bleeps, whirrs and whistles received through earphones that the user, usually a blind person, uses to distinguish objects in the environment. Scanning left to right the instrument produces particular sounds which change in pitch for elevations and in loudness for brightness; the transduction is possible due to the fact that the visual cortex of even blind people can respond to the sound it produces. The modern vOICe looks a lot like a normal pair of sunglasses (with a mini-camera) and it is easy to wear.

SSDs are thus machines that carry out signal transduction. Through brain-machine interface they transmit signals to the brain that can be interpreted by different neural systems. In this manner the user can receive, process and respond to stimuli otherwise inaccessible by the defective sense. The device can substitute senses due to the great cross-modal plasticity of the brain to accommodate information in different modalities, taking information from one sensory input and adapting it to another modality, and allowing the user to perceive external objects, discriminate distal attribution, and thus navigate the surroundings. Detection of environmental stimuli can be quite realistic: in an experiment researchers zoomed in the object without letting the user know, and the response from the user was an instant defensive move back raising his arms; in another example a blind person was able to perform delicate work on an automatic production line using the device (Bach-y-Rita, 1972).

Research method and the scope of this dissertation

This thesis is primarily concerned with offering grounds to build a theory of consciousness and the observed-self based on SSDs research rather than with testing a hypothesis, and therefore my research approach is essentially qualitative. I have adopted an interpretative ontological and epistemological stand that provided me with the opportunity to not only critically analyse existing views and empirical evidence in the areas of SSDs, consciousness and the self, but also offer innovative concepts that can assist producing a novel approach. My thesis is essentially a 'top-down' reflective research effort to study deductively the narrower issue of the 'self' from the more global theme of the mind. Towards this end, I have reviewed papers that focus on consciousness, the self, and SSDs with the intention to produce evidence that the mechanics of the observed-self can be found in the process of SSDs incorporation into the body-schema.

Chapter 1: The concepts of Consciousness and the 'self'

1. Traditional views and their limitations

A long philosophical tradition was to consider the self as a non-innate social intersubjective construct and the product of cultural socialisation. Josiah Royce (1898), for example, argued that self-consciousness serves a social function, and George Mead (1962) believed that there is no primary self-consciousness other than the one generated by the social context. This tradition was then challenged. Jean-Paul Sartre (1956) amongst others, for example, argued that our experiences are 'self-given' acts that can be understood as independent pre-reflective intrinsic phenomenological features of consciousness, and expanding on this notion Shaun Gallagher (2000) proposed the existence of a 'minimal self', not subject to temporal extension – as opposed to a 'narrative self', which possesses continuity over time giving us self-identity.

A common feature in both traditions has been the notion that the self can be seen as a manner of self-evaluation or to 'relate to oneself' (Foucault, 1984), exhibiting in our ability to reflect on one's own thoughts and feelings. The general idea has been that self-consciousness is interacting reflexively on oneself, but this notion can be ontologically and epistemologically disputed. Consider the old problem that to be aware of oneself requires an 'observer', a homunculus. Although the notion of a homunculus has been largely dismissed, the mystery of 'who is the observer' has remained unanswered. I align myself with John Searle's (2005) suggestions that the existence of an intrinsic observer seems unavoidable, and I would propose that a possible way to offer a solution is by asking the question: If I am able to sense my experiences, am I essentially a *processing-sensor*? The basic rationale for this question is that if through bodily sensoria I am aware of the world around me and the state and feelings of my body, and I use my thoughts and feelings to process this information and make decisions, then I could be seen as a processing-sensor. If we accept this idea, it could be argued that it is illogical to presume that one can sense oneself because a sensor functions not to perceive its own parts but to receive stimuli and respond to them. In this dissertation, I will provide, based on sensory-substitution-devices research, arguments to respond to this question affirmatively, with the intention of promoting the belief that understanding the self as a processing-sensor offers basis for a mechanical[1] theory of the self and the roles it plays in consciousness. This model, though, requires the existence of two selves, an *observer-self* and an *observed-self*; I will argue that only the latter is 'the (unconscious) constructor' of our field-of-awareness, with the former being 'the (conscious) experiencer' of this construction.

The essence of the self has been traditionally studied presupposing the existence of a minimal self, somehow representing the first-person 'I'. Gallagher (2000) argued that this minimal conception of 'I' enjoys immunity in that it cannot be possibly mistaken for what is 'not me', and it is impervious because it is thought to be non-cognitive in the sense that it is uncontaminated by judgement or reflection or by an intention to identify oneself. Being non-cognitive, it could be conceived, to use Gallagher's words, as a 'pre-reflective point of origin' for our actions and experience; an elementary form of consciousness. This notion has been the focus of some attention. José Bermúdez (1995**)**, for example, proposed the term 'non-conceptual first-person content' to explain how through our interactions with the environment we gain information about ourselves – what Ulric Neisser (2006) called the 'ecological-self'. The gist of the argument has been that non-conceptual self-awareness seems to be innate and reducible to no more than a natural ability to perceive without cognisance – this claim gained support from evidence that neonate babies seem to possess a natural ability to imitate facial expressions non-reflectively (the assumption being that if the response is natural and reflexive it must be self-generated). Galen Strawson (1999) attempted to make further contributions to understanding an elementary form of self by arguing for a local phenomenological reflective hiatus-free self without history or continuity.

These views, however, have a shortcoming**:** they all fall short of explaining how a reflexive minimal self can generate phenomenal self-awareness, introspection, feelings and cognition. To remedy this, I suggest that the minimal self ought to be understood as a processing-sensor in its own right, an 'observer-self' emerging from the synergetic[2] sum total of bodily components functioning in unity. Throughout this thesis I will be arguing that functioning in 'closed-system' mode (see Glossary for a description of 'closed-system') this observer possesses the property to convert biological activity into phenomenological perception.

[1] The term 'mechanical' implies identifying the different constituents of consciousness and the self, how these components interact, and their primary functions.

[2] Synergy is understood as the process of feedback between two systems that is amplified by recursive loops, similar to how a sound is amplified when the microphone is too close to the speaker. It produces an effect that is greater than the sum of the separate parts, and the system as a whole can thus be seen as a new and separated operating system in its own right.

Considering an observer-self as a processing-sensor allows for the possibility that perceiving one's thoughts and feelings is an act of 'observing brain activity', essentially not unlike how we observe any other stimuli[3]. And if this is the case then the terms self-consciousness and self-reflection are misnomers because the observer-self can be construed as only being capable of detecting 'otherness'. I will now describe the hypothesised observer-self a bit more in depth.

1.1 The 'observer-self'

The observer-self is presumed to be responsible for transforming biological signals into phenomenological experiences, and I would suggest, in agreement with Searle (2005), that the main problem of current neurobiological research has been a mistaken conception of the self. Searle proposes the existence of an 'abstract' self and points out that to understand it we should first understand consciousness, a task that could be made simpler if we ignore for a moment scientific analysis and focus instead on the pragmatic role of consciousness. From this angle of investigation, consciousness could be basically described as our sentience during wakeful (and dreamlike) states, and the criteria in place for this abstract observer are different than those used for a neurobiological stand. Here we identify as important our subjective qualitative feelings perceived from a first-person perspective (a first-person ontology), which we sense as a total and unified conscious experience. Based on this appraisal, the observer-self could be identified as *the entity that feels the qualitatively subjective experience of a united conscious field,* and I contend that this experiencer emerges from the synergetic sum total of bodily functioning. I believe that since the amalgamation of bodily functioning is achieved with synergy[4], the observer-self emerges from it because the amalgamation is more than just the sum of its parts[5] and thus unique.

I would argue that this 'observer-self' is the elusive 'experiencer', the 'I' in the first-person ontology, and that it functions to process information captured in the form of neuro-biological signals, transforming them into our phenomenological experiences**,** which we perceive as a

---

[3] The observer-self is presumed to be a self-sustained 'autopoietic' system, and since, according with autopoiesis theory, systems cannot access internal components of other systems, all communication is conducted by observing their external properties. An observation however, is a subconscious release and absorbing of signals – through which two systems become structurally coupled. The expression 'the observer-self observing the brain' should thus be interpreted as: 'the experiencer (the 'I' in the first person ontology) exchanging subconscious information with the brain.

[4] See Corning, 1983 for a more in-depth description of synergy and it ubiquitous nature.

[5] Further research should explore the observer-self as an 'autopoietic' network of processes (a closed-system capable of maintaining itself) that cannot be described by using components that define a different system (Maturana & Varela, 1991).

unified conscious field and a narrative whole. Put differently, the body, functioning as a unique closed system, generates our perceptions, feelings and cognitions in biological form, but as an agglomeration it causes the emergence of a different and unique system, the observer-self, a processing-sensor in its own right with its own independent properties. Information is no longer perceived as biological signals but as an integrated field of awareness manifesting as unified subjectivity (See Searle, 2005 for a somewhat similar view) – through a process not unlike how a transducer converts energy from one form into another[6].

Having briefly considered a possible outline for an 'observer-self', we now need to determine how this self experiences the integration of SSDs into what we perceive as an observed-self.

1.2 The 'observed-self'

Our conscious mind is, in simple terms, what we are aware of; it exhibits as a united-field-of-awareness[7] that manifests in two broad fashions: attributions-to-oneself, which we perceive as a qualitative, unified subjectivity (Searle, 2005), that is as being 'me, the person that I am'; and distal-attributions, which we sense as 'not me'. The observed-self is the former, a part of our field-of-awareness that we perceive as our body and personal-identity, which separates 'the person that I am' from the outside world[8]. I theorise that when a SSD is integrated into the body-schema, the incorporation changes the ownership of the body by changing the afferent/efferent nerve signals, and the result is modifications on how the 'observer-self' perceives the 'observed-self'.

The traditional notion of the self is therefore comparable with what is referred here as the observed-self, and it is built by our perceptions of how the body changes over time, and in the body's history of relationships with the environment. This history is our memories, and since the self is essentially the narrative history of our memories (e.g. Heersmink 2017), objects that are incorporated into the body-schema gain a place in the memories of the observed-self and thus in the person's sense of individuality. Also, as a SSD becomes incorporated into the body-schema, it comes to be an integral part of the user's arsenal of strategies to deal with the

[6] The observer-self could be considered as a converter of energy, somewhat not unlike how the brain is the converter of environmental inputs into electro-chemical operations.
[7] This notion has been studied under the general paradigm of 'united-field-approach', adopted by a comparative small number of researchers (See Searle, 2005).
[8] In reality, the observed-self should be understood as our whole field-of-conscious-awareness, containing also the 'extended self' because what we perceive is interactions with the world. However, this argument is not only beyond the purpose of this thesis but considering it would complicate matters unnecessarily for the basic model proposed here.

world, reinforcing further the device's status as a component of the observed-self. Incorporated SSDs become established features of the individual's personal attributes and conscious field and, endowed with particular experiences and memories, they influence the person's sense of proficiency to interact with the environment.

On the surface these concepts could be seen as common-sense intuition, but to properly understand how SSDs assist generating a cogent theory of consciousness and the self, we need to look in more detail at 'the gap' between biology and phenomenology, and at the notion of an observer in charge of perceiving what we sense. It has been pointed out that to make sense of our conscious experiences, it is necessary to consider the existence of an entity possessing properties that include the ability to be conscious**,** to be capable of perceiving and assigning these perceptions to memory**,** to persist over time**,** to operate in the gap between the biological and the phenomenological**,** and to make decisions and act accordingly (Searle, 2005). The difficulty in identifying and explaining an observer with these characteristics has been allocating these attributes to only one 'self', without considering the different roles played by the two selves explained here. Searle, for example, contended that we cannot have both a conscious field *and* a self; using the analogy that the area between Mexico and Canada cannot consist of 50 States *plus* the USA because they are both the same, he treats consciousness and the self as one. Such interpretation makes it impossible to identify which part of the self has intentions and makes decisions and which part is observed, but this problem can be avoided by allocating the power to interpret and decide to the observer-self, with the observed-self being the result of these decisions – not unlike how USA as a Nation makes laws that control its States.

In this manner, the role of consciousness is to generate a whole field-of-awareness, whilst the role of the observer-self is to separate perceptions of the external world from perceptions of 'me'. Therefore, the attributes of being conscious, assigning perceptions to memory over time, and making decisions, belong solely to the observer-self[9] – this part of the self, for example, may elect to look at USA as a Nation with its culture, laws and a particular society, and in this case the perceived picture of USA is different than a collection of separate States. The ability to interpret and have intentions is thus crucial to describe the foundations of the observer-self.

[9] Future research should consider explaining the mechanics by which the observer-self complies with these roles. Once again, the works of Maturana and Varela (1991) on 'autopoiesis' may assist.

The history of our self-identity and the sense of being the same person over time are in reality histories of connections between 'me' (the observed-self) and the outside world, that is, between the 'private' and the 'extended' parts of consciousness. These impressions, I would argue, are housed in the observed-self in neuro-biological form, but they are experienced phenomenologically, appraised cognitively and controlled by the observer-self[10]. In Chapter 2, I will be arguing that when incorporated into the body, SSDs provide the user with new ways of interpreting the world, modifying the body-schema and the observed-self.

If this reasoning proves valid, it can provide the means to offer a tentative definition for the observed-self as: *the part of our conscious mind that we perceive as the person that we are, with our personal identity, personal attributes and history, that is, 'attributions to oneself'*.

1.3 Our sense of agency and of ownership of the body

Central to understanding the selves described above is the distinction between having a sense-of-agency and having a sense-of-ownership. The former refers to being aware that it is 'me' who is generating an action, and the latter is the sense that 'I' am the agent sensing the experiences. Gallagher described the process in a neurocognitive model of immediate self-awareness (Fig. 1); between an intention and the action executing it (represented in the 'Intended State' and 'Actual State' boxes in the diagram) there is a subconscious mechanism that compares efferent commands with motor intentions to allow for error corrections.

The sense of agency is produced by the anticipated sensory feedback (represented in the 'Forward Model' in the diagram) and the sense of ownership by the 'Predicted State'. The intricacies of this mechanism are not relevant for our purpose; what is relevant is that we analyse the epistemological connotations of Gallagher's conceptions, and see how they compare with what is proposed here.

The observer-self proposed here is presumed to be able to only perceive 'otherness', and therefore, since self-perception is not allowed, it is necessary to clarify how it accommodates the notions of agency and ownership. Let's consider first the sense of agency. Gallagher says that we feel a sense of agency when we perceive that we have caused the action, and he recognises that under certain circumstances such as trauma, illness or brain dysfunction one's

[10] I would suggest that a topic of further research should be to evaluate the merits of a hypothesis that the observer-self makes decisions when assessing thoughts and feelings generated by the brain and focussing and favouring (by attending to) some thoughts over others.

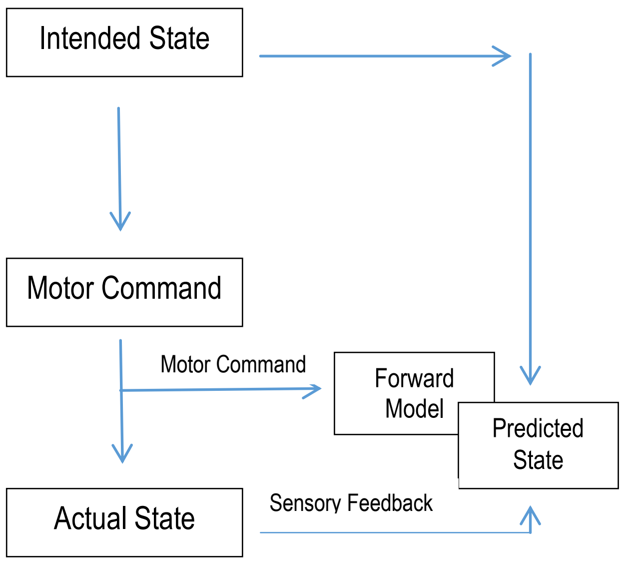

Fig. 1: ***Neurocognitive model of immediate self-awareness****. A pre-motor non-conscious comparator (Forward model) compares efferent copies and motor commands with motor intentions to detect and correct possible errors in the sense of agency and ownership.*

actions can be perceived as alienated or caused by others. He explains this phenomenon as a mismatch between an intention and the 'feedback comparators' in the afferent/efferent mechanism. According to Gallagher's view, the self in a disrupted comparator feedback produces alienation to the perception of being the orchestrator of an action, but the lack of agency could be interpreted with an alternative explanation. The model proposed here would suggest that what is disturbed is not the observer-self, but rather the observed-self that it perceives. Gallagher's self in a disruptive comparator feedback system is still what the observer-self captures; the only difference is that it now senses a different reality, an alienated observed-self.

The same reasoning would apply to our sense of ownership; whether we recognise things as owned by the body or not, they are still perceptions captured by the observer-self. Incongruities or pathologies resulting in misperceptions, according to my view, would indicate that the observer-self is allocating distal-attribution and attributions-to-oneself differently to how it otherwise does in healthy individuals because this is what it observes.

There is another alternative explanation that the model in this thesis needs to accommodate. Gallagher makes a distinction between a minimal self (a basic experience not extended over time) and a 'narrative' self (a historical 'self' extended over time), and he explains that it is the narrative-self that gives us self-identity. But the mechanics of how this notion of a minimal-self can generate a narrative-self seem unclear, or perhaps more puzzling, how a minimal-self that could be seen as 'non-cognitive' can produce cognition. I would suggest that these difficulties can be avoided if we consider the alternative explanation that a minimal self is in reality the proposed observer-self, and the narrative self the history of experiences brought about by the observed-self (I will expand on these notions later).

Another major distinction that we should make regarding the minimal-self is that the observer-self in my model is envisaged as possessing elementary cognitive capabilities (otherwise it would be impossible to comprehend how it could assess options and make decisions). I would argue that cognition manifests in our abilities to contemplate our thoughts, understand their cognitive contents, and make proper decisions accordingly.

1.4 The observer and observed selves are presumed to be essentially 'processing-sensors'[11]

A major pillar of the proposed theory of consciousness and the self, based on Sensory-Substitution-Devices (SSDs) research, is the presumption that the observer and the observed selves are essentially processing-sensors processing information internally, and this requires understanding their basic mechanics. A particular study in robotics conducted by Jun Tani (1998) may offer valuable information about how internal neurological mechanisms process sensory-motor information, and from this conceptualisation we can then make inferences about the existence of a similar mechanism for the observer and observed selves.

Tani offered a model of the structure of the self based on the results of experiments with a visual-based mobile robot (please note that Tani's description of the self as 'self-consciousness' corresponds to the notion of the observed-self as proposed in this thesis, and they should be taken as equivalent for the purpose of this analysis.) The essence of Tani's model is simple: the self comprises two major mechanisms, namely an internal mechanism acting as a 'subjective mind' that makes predictions about what to expect when observing the environment, and the actual observations of the objective world. When predictions by the subjective mind are accurate and they fit the observation sensed, there is little or no need for the self to intervene, and in this case the robot is supposed to be non-self-conscious. When

[11] This thesis proposes that the observed-self and the observer-self are processing sensors. In open-mode they detect signals incoming from other systems (sensor), and in in closed-mode they process this information internally (processing).

the observed world does not fit the predictions, there is conflict that requires learning and in this situation the robot exhibits self-consciousness. Tani concludes that in this manner, *self-consciousness of the robot arises in this moment of incoherence…directed to the conflict to be resolved* (Tani, 1998, 538).

The mechanics of self-consciousness in the robot appear to be similar to those observed in the olfactory systems in animal experimentation (Skarda & Freeman, 1987). It was found, for example, that when experiencing a new odorant, neural activity in the animal's CNS becomes more active to process the new stimulus, and the relation between the internal mechanism and what the animal senses continues changing through the learning process – just like in the robot.

If we align with Tani's interpretations of self-consciousness, it could be possible to infer that both the robot and a person exhibit self-consciousness in their own particular ways that emerges when the relationship between internal mechanisms and an object in the environment becomes incoherent. A second inference could be that the main purpose of self-consciousness (our observed-self) is to resolve conflicts caused by unexpected observations and thus adapt to environmental situations to reach intended goals. The implications of these interpretations for the model proposed here are significant because they suggest that an observer-self (the robot and a person's body as wholes) can be seen as processing-sensors that activate relevant internal mechanisms when required and for the purpose of connecting 'me' (the observed-self) with the world and thus evolve.

Another important concept suggested by Tani's experiments is that the observed-self is a system that can take states as both closed and open systems. In the open- mode, it exhibits a 'field-like' aspect and the property of acting as a sensor, detecting external entities. In the closed-mode, a 'local' aspect, it acts as a processor of this information. In Chapter 4, I will briefly describe how these properties are presumed to fit into the proposed model of consciousness and the observed-self based on SSDs research.

Having provided a possible rationale for the concept that the observed-self is basically a mechanism that in closed-mode processes information that intakes in the open-mode, I would like now to make some inferences regarding how a similar rationale may be used to explain the observer-self. Let's consider first the sensor aspect. Whilst sensors in the observed-self are the bodily senses, a sensor for the observer-self could be presumed to be the property of the synergetic sum total of all bodily functioning acting as a separate and independent operator – to perceive information from the observed-self and converting this biological information into phenomenological perceptions. In the case of the observed-self, sensed information is processed by the brain, but in the observer-self it has to be presumed to be 'me, the person that I am'; I would suggest that we process this information by attending to our thoughts and feelings and making decisions[12].

I would like to close this segment by noting that a similarity, which can be useful for future exploration, between the selves of Tani's robot and a human, as suggested in this thesis, is that both, robot and human, rely on sensors to represent the world and on an internal subjective mechanism to separate 'me' from 'not-me' and negotiate with the world.

1.5 The observed-self and the body-schema

If one is prepared to accept that the selves are essentially properties of the body and processing-sensors, the next step would be to find research efforts that may not conflict with this presumption, and that may offer grounds to expand on it. In this section, I am exploring this evidence, using it to describe the connection between the observed-self and the body-schema. One such possibility can be found on the works of the influential psychological researcher Jack Loomis (1992). Loomis attempted to explain the self by considering it as comprising the bodily-self (the body-schema), interoceptive sensations (e.g. hunger) and our subjective thoughts and feelings. He recognised a distinction between the phenomenal and the physical world, and the distinction was important because the impressions of 'me' and 'otherness' belong to the phenomenal world (the world as perceived by the observer-self). The reason that we perceive (see colours, for example) is not because stimuli like colours exist but due to our phenomenal ability to 'infer' the physical world (Campbell, 1966). A way to understand consciousness and the observed-self is thus by proposing that it is in our subjective experiences that we find the division between what could be referred to as *attributions to oneself* and the rest, referred to as *distal attribution*. Without this separation it would be impossible to separate ourselves from the world, live and evolve, and the observed-self makes this separation possible.

[12] These are bold assumptions and they should be interpreted only as possibilities to be explored in future studies.

The distinction between 'me' and 'otherness', however, is neither fixed nor with clear boundaries, but rather a constantly evolving process that allows a person to confer different degrees of distal attribution to different objects. Also, what is experienced as part of 'me' can still be felt as such even when it is not, as in the example of the phantom limb (See Simmel, 1966), and under the influence of psychoactive drugs or psychopathologies the boundaries between self and non-self can be altered (e.g. Kluver, 1967). I'll be addressing the flexible boundary nature of the observed-self a bit more in depth in Chapter 3.

Following the works of Loomis (1992), I would like to describe a rather simple theoretical model to explain the division between attributing something to oneself and to otherness. The perception of otherness is created when afferent signals from senses to the CNS and efferent signals from CNS to the periphery are lawfully related and include external objects (See also Bach-y-Rita, 1972). Attribution to oneself on the other hand occurs when this relation is unrelated or independent of the external stimulus, thus felt as internal or 'private'.

Loomis' model is based on the works of Erich von Holst (1954) and represents attributions to oneself (the observed-self) as afferent and efferent connections between 'effectors' (the periphery) and the brain; distal-attributions are 'extensions' of the effectors – Fig. 2 shows an adaptation of von Holst's schematic representation of this mechanism. Since the observed-self ends up in the effectors, what the model describes is essentially the notion that the self could be seen in the internal representation of the body-schema, and therefore, understanding the body-schema is crucial to describe and understand this private part of our consciousness.

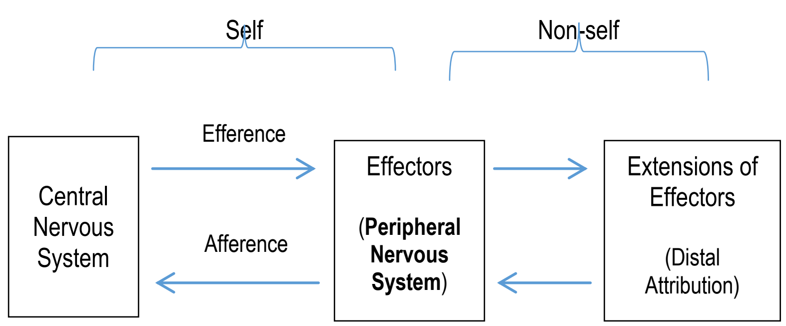

*Fig. 2: **Efferent/Afferent model**. The self is represented in the CNS-Effectors connections. Distal attribution (non-self) is represented by the Extension connections.*

Since initially coined by neurologist Henry Head in 1911, the term 'Body Schema' has undergone some reinterpretations but it has retained the essential meaning of *how the body is subconsciously represented in movement and space*. The representation of a body-schema changes over time, being modified by incoming sensory impulses, and since SSDs can act as senses, they affect the body-schema; the result is changes in both the properties of the body-schema and the perceived body and sense of individuality.

The properties of the body-schema have been identified (Haggard & Wolpert, 2005); they comprise seven fundamental attributes, all important to understand the observed-self. Perhaps the most relevant one for our purposes is 'supramodal', the faculty of the brain to integrate proprioceptive information (position, location, orientation and movement of body parts) to maintain the functioning of the body as a whole, and to combine different senses in the same neural representation of the body – a sensorial simultaneous participation. The body has to do this functional integration 'coherently' and with 'adaptability' (Carlson et al., 2010), that is, it has to resolve differences amongst sensoria to maintain coherent continuous organisation and adjust as quickly as possible to new developments. The property of 'adaptability' is particularly important because to evolve a person needs constant readjustment to environmental demands, and the body appears to have developed a good predisposition towards this end (Johnson, 2000).

Another property of the body-schema relates to the plasticity of the brain, and refers to the brain's ability to integrate information spatially within different regions that contain multisensory neurons capable of taking information from more than one modality. This allows the brain to integrate information transmitted interpersonally and from the environment to constantly update the body-schema.

These properties play a crucial role in how the body schema is transformed by SSDs; they will be expanded in Chapter 2. In the meantime, it would suffice to say that since the proposed model of the selves involves afferent/efferent sensory-motor interactions, the main functioning of brain networks to consider when explaining it is related to the role that neural circuits play in setting the conditions for the observed-self to communicate with and intake information from the world. When dealing with other persons, these circuits, for example, may involve mirror neurons to imitate actions (Chaminade et al., 2005).

The observed-self and the body schema are thus related in the sense that the body, as considered here, marks the boundaries of the private part of our field-of-consciousness. This is a central concept in this thesis and implies that a general attribute of the body is to incorporate non-bodily objects that have potential to be integrated. Incorporations have been typically studied as integrations of tools or prostheses; however, incorporations of SSDs follow similar mechanisms. They can be thought of as sophisticated tools; when integrated into the body they have the potential, like tools and prostheses, to alter the body-schema and function similarly to any natural sense, modify the body-schema and re-establish the boundaries of the observed-self and the perception of what constitutes 'me'. As we will see later, however, the process of incorporation is complex and integration difficult.

1.6 The body-model

The body-schema is not analogous to the concept of a *body-model*, and the distinction is important when trying to explain how the image of oneself is transformed together with bodily transformations. Whereas the body-schema refers to the unconscious process of the body in action, the body-model can be understood as a pre-existing *conscious* state of bodily affairs stable over time, giving the person's self-identity and personal attributes. Whilst the body-schema undergoes transformations, the body-model may impose constraints for these transformations to maintain stability of functioning (de Preester & Tsakiris, 2009).

Transformations of the observed-self by SSDs refer primarily to how the device can be integrated as part of the body-schema. But this integration should be seen within a process that is characterised by grades of integration and that should be understood in terms not only of integration but also of *incorporation*. Incorporation is basically a deeper and more enduring state of integration that includes incorporation into the body-schema, the body-model, and into the image of oneself. The mechanism could be explained in simple terms as follows: an external object, for example a cane used by a blind person, that is initially completely foreign to the body, may gain access to it and be integrated into the body-schema through regular use. At the beginning, this integration is superficial and perceived as merely an annex to the body, a detached 'tool', but not as incorporated into it, and certainly not felt as part of the observed-self. Through habitual use the cane is gradually integrated deeper into the body-schema of the blind person and in time it may acquire the status of being incorporated as part of the body-model. However, although at this deep level of integration the object is used subconsciously, it is still not felt as a 'natural' part of the self. The final status would be when the object has undergone such profound degree of incorporation that the person would perceive it to be just as natural as any other part of the body. Being integrated into the body-schema is thus a pre-requisite to being incorporated into the body-model and the observed-self.

Therefore, with regular use a non-bodily object could be felt as an extension of the body, but to feel totally incorporated it has to modify the body-schema *and* the body-model. This is the case because the incorporation has to generate a change in the sense of ownership of the body (de Preester & Tsakiris, 2009); modifications that do not produce this change of ownership and a re-organisation at this level may not be perceived as true transformations. And since the observed-self relates to essentially the highest levels of body-model, as it is argued in this thesis, the sense of an object being part of oneself should be found primarily in the transformations of the body-model, that is, in the perception of the biological body and its enduring attributes.

Bodily transformations and integration of SSDs into the body-model could also be considered from the point of view of *embodiment*. Described simply, embodiment is the phenomenological perception of an enduring body that manifests in a feeling of subjective sensorimotor unit of functioning (Merleau-Ponty, 1996). It exhibits as a complete perception of self-identity and of ‘me’ stable over time. This enduring feeling of unity over time can be noticed in the phenomenon of the phantom limb – the sensation that an amputated or missing limb is still attached. The feeling of detachment is denied because it has not yet caused the unity of embodiment or body-model to be modified, thus preserving the sensation of the limb still being attached – and thus still part of the observed-self. If the amputation is then replaced with a functional artificial limb, the body-model may in time be re-organised to accommodate the prosthesis, which may be integrated as part of the body-schema and the person’s life and thus potentially as part of the perceived self.

Perhaps a more dramatic example of the body-model and embodiment (and indirectly the observed-self) preserving its existing unity over time is the rubber-hand illusion. To produce this illusion a rubber hand is placed in front of the person as replacing the real hand, which is hidden behind a screen. By stroking the rubber and the real hands synchronously the person soon starts feeling the rubber hand as real (Botvinick & Cohen, 1998). The illusion can be taken as evidence that a non-corporeal object can be felt as integrated into the person’s representation of the body (Tsakiris & Haggard, 2005), and also that the body-model tends to be preserved over time – the person does not feel as having three hands (Longo et al., 2008). It offers as well indirect support to the claim that the observed-self is transformed concurrently with transformations in the body-model.

1.7 Neural correlates of phenomenological experiences

Therefore, the sense of having a self can be understood as an extension of the sense of ownership of the body. It is presumed to arise from how we represent our body (Tsakiris & Haggard, 2005) through a process that entails processing multisensory inputs (e.g. visual, tactile and postural – see Graziano & Botvinick, 2002) through cycles of afferent and efferent nerve circuitries (Loomis, 1992). A question that should then be addressed is: What are the neural correlates of the body-schema and the body-model, and how do these brain regions interact with afferent inputs from peripheral sensoria to generate the sensation of owning the body and being 'me'?

The neural correlates of body-schema have been long identified in maps of the brain where different areas process information from different modalities – this is the putative 'Body Schema' of classical neurology (See Maravita & Iriki, 2004 for an analysis of how this neural map is modified by the use of tools). The neural map for the body-model, however, is different; it has been investigated mostly in terms of body ownership. In general, results from this line of studies indicated that the perception of body ownership, associated with the impression of body-model, may be found in the right temporal and parietal lobes activity. Lesions in these areas were noted to generate a response of denial of ownership (Bottini et al., 2002), and electrical stimulation in this region produced 'out-of-body' experiences (Blanke et al., 2002). Also, an experiment conducted by Tsakiris et al. (2008) revealed that stimulation in the temporo-parietal area can reduce the effect of the rubber-hand illusion, changing the feelings of ownership of the rubber hand. These results can be interpreted as indicating that the body-model plays a crucial role in the sense of body-ownership and therefore in the observed-self.

If the body-schema and the body-model play a role in the modulation of the observed-self, it could be argued that together they house the impression of our self-identity, containing a reference description of the constantly modified state of 'me'. And a primary role of the body-model (as mentioned earlier) is to preserve stability by imposing constraints on what is perceived as foreign to the body (See Costantini & Haggard, 2007), but allowing at the same time for gradual updating of our sense of ownership and self (See Wolpert et al., 1998).

Based on this information, I would suggest that integration of SSDs into the body-schema can be thought of as a process that involves three broad components of consciousness: a distal-attribution of conscious awareness (perceptions of the external world), attribution to oneself (the observed-self), and the mechanisms that relate the two, which are responsible for internalising the SSD and integrating it into and as part of the observed-self.

Generalising from this assertion, distal attributions are considered here as the part of consciousness that is extended from the body and thus felt as ‘not me’; it is the feelings that the object in question is recognised as foreign even though it may ‘feel’ as belonging to the body[13]. A prosthetic arm, for example, may be felt as an integral part of the body and as functionally attached to it, but the person is still aware that it is a foreign object. The observed-self conversely exhibits in the feeling that the object is no longer foreign but it is recognised as a natural part of the body.

The observed-self is thus well protected from deep incorporations of non-corporeal elements because it is not easy to develop such deep intimate relation with external objects. As will be seen later, full integration into the observed-self is extremely rare. What is more common is partial integration, whereby a non-bodily artefact generates a strong feeling of body-ownership and body-belongingness but not sufficient to be cognitively recognised as being part of the ‘natural me’.

In this Chapter, I briefly addressed the mechanisms of consciousness as comprising an observer-self and an observed-self, and the roles that the body-schema and the body-model play in the mechanism responsible for incorporating non-bodily objects into the observed-self. In Chapter 2 I will review current research on SSDs and analyse the mechanics of integration of the devices into the body-schema, describing how the integration can lead to ‘brain exaptation’ and modifications of the observed-self.

[13] Note that this description of distal attribution goes further than the traditional notion of attributing a cause of proximal sensation to an external object. In this thesis I am presupposing that objects that are felt as proximal senses may still be treated ‘cognitively’ as distal.

Chapter 2: Sensory-substitution-devices and the observed-self

2. Research on SSDs can provide information to study the mechanics of the observed-self

The foundations of my theoretical framework as explained in Chapter 1 can be summarised as follows: The observed-self can be understood as perceptions of afferent/efferent nerve signals that represent occurrences within the boundaries of the body. Bodily senses undergo changes that can incorporate non-bodily objects such as SSDs, which can then be integrated into the body-schema and modify the body-model, transforming the observed-self. Therefore, to generate a theory of the mechanics of consciousness and the self, what is needed is an understanding of how SSDs transform the body-schema and how these transformations may result in neural modifications. Once these mechanisms have been identified, it should be possible to infer how our natural bodily senses develop in the course of evolution. With this in mind, this Chapter is intended to answer three fundamental questions: 1. How are SSDs integrated into the body and modify the body-schema? 2. What are the neural correlates of these modifications? 3. To what degree are SSDs incorporated into the body-model and transform the observed-self? I will first review literature on SSDs and then use this information to provide likely answers to these questions.

SSDs have typically been used to study distal-attribution (Auvray et al., 2005; Siegle & Warren, 2010; Kiverstein et al., 2015; Bach-y-Rita & Kercel, 2003) rather than the self, but they offer valuable information to generate a theory of the mechanics of consciousness and the observed-self. An essential concept in these studies is that when using a SSD, with practice the interface between the user and the extended device disappears and the device is integrated into the body-schema; the person no longer senses the device but what the device captures. Julian Kiverstein and Mirko Farina (2012) explained that the experience 'extends to incorporate the device'; also, a consequence of this extension is a recalibration of brain circuits (Clark, 2008) – there is for example evidence that continual use of an external device such as a rake can restructure the functioning of bimodal neurons (neurons that fire when activated by internal and external signals) in macaque monkeys (Maravita & Iriki, 2004).

2.1 Factors of SSD integration into the brain-schema

One consequence of SSDs integration into the body-schema is that it helps understanding how something external to the body becomes, and is felt as, part of the body. The device is initially perceived as an external object, but with practice and through habituation it becomes a proximal stimulating sensor basically no different to any natural bodily sense. The process, however, is complex and understanding it requires understanding the mechanisms that lead to the integration.

Kiverstein and Farina (2012) explained that the SSD becomes gradually amalgamated with the person forming a single integrated system and that therefore the substrate of consciousness should be seen as part neural and part technological. They rationalised that this allows consciousness to extend into the external world and incorporate it into the body. In line with this reasoning, I would clarify that consciousness is being *modified* rather than extended by the SSD, and that this modification entails the external object being incorporated into the observed-self, which then acquires a novel function[14].

Integration of a SSD is, though, gradual and it is commonly felt as not fully integrated (de Preester, 2011); therefore, the issue needs more thorough explanations. We can start by describing the constituents of the amalgamation and their role. Bach-y-Rita and Kercel (2003) argued that information from SSDs, for example a TVSS device in a blind person, is conveyed to the brain through a human-machine-interface, replacing information previously received by the defective sense. Practice transforms the TVSS into a proximal sense in a process that involves four main factors**:** a subjective impression or phenomenological experience that cannot be fully compared with existing modalities; the characteristics or properties of the object of perception; how different independent senses respond to stimulation; and how this response causes supplementary neurological processing of the perception (Grice, 1962).

Viewed jointly these factors indicate that although SSDs act upon existing modalities, they generate a novel one, a *new sense*, an issue that has received considerable attention. Auvray and Myin (2009), for example, addressed the question of whether the substitution of a sense is more in line with the substituting or the substituted sense and they reached the conclusion

[14] The reason that the term *modification* may be more appropriate than the term *extension* is that the evolution of the self does not necessarily involve extensions and it could even involve contractions.

that it is neither, but a novel development endowed with its own characteristics. Their argument has been supported by findings that newly formed perceptions are only approximate and not as vivid or neat as real ones; they are for example 'quasi-visual' experiences (Deroy & Auvray, 2014) or 'quasi-auditory' perceptions (Kiverstein & Farina, 2012) – quite different from bolder earlier claims of 'seeing though the ears or the skin'[15].

In brief, the perception is not fully visual, acoustic or tactile but a new perception incorporating different senses, and SSDs do not reduce perception to an existing modality but rather engage a cross-modal process leading to a novel ability (Deroy & Auvray, 2015).

2.2 SSDs can lead to brain 'exaptation'

One way to understand how the observed-self may be transformed by the integration of a SSD into the body-schema is to understand how the device acquires the status of a proximal stimulation sense, and since the process involves neural adaptations, we need to analyse the brain substrates involved in the changes. But there is a semantic issue that we need to address first because the term 'adaptation', as commonly used in the field of evolution, implies a teleologically-loaded suggestion of 'pre-adaptation', that is, an adaptation that 'imposes' certain consequences, which seems to conflict with the notion of natural selection, and the word *exaptation* was suggested as more appropriate (Gould & Vrba, 1982). An exaptation refers to a trait that can evolve because it serves a new particular function, and I believe that this is more descriptive of what takes place in brain regions during the process of SSDs integration. For example, a neural area that evolved to process auditory information can undergo transformations to process visual information in a manner not unlike how bird feathers that evolve as insulation were 'exapted' as means for flying. The importance of the term exaptation also lies on the fact that it refers not only to developments by natural selection but also to co-option adaptations that could involve non-natural artefacts such as SSDs. I would thus like to suggest that when referring to brain areas acquiring functions for which they did not originally evolve, such as auditory areas processing visual information, we refer to the new developments as brain exaptation. With this in mind, I would like now to analyse how SSDs are transformed into proximal stimulators and modify the body-schema.

[15] The term 'quasi' implies that perceptions through SSDs are unique and only partly resemble natural visual or auditory experiences.

2.3 The underlying mechanics of SSDs integration into the body-schema

Since integration of a SSD produces a novel sense, senses are not really substituted by other senses, but rather existing senses are used as 'raw material' in the creation of new ones. A crucial clarification to be made in this regard is that although sometimes brain areas tend to *defer* or *dominate*[16], deference or dominance is not as relevant as the fact that both senses integrate to create a novel one. SSDs integration can hence represent good examples of brain exaptation because they activate a mechanism by which different neural areas not specifically adapted for the new function are recruited to form new circuitries that from then on function to exclusively process information from the integrated SSD. In this manner, afferent/efferent cycles generate the 'quasi-vision' or 'quasi-auditory' sensation typical of the devices, and this provides good basis for explaining the mechanics that transform SSDs into proximal stimulators.

There are two factors that we need to take into account to explain the integration. One is that the process of SSD integration should be seen as more profound than simply integrating a tool or an artificial sense (as explained by Lenay et al., 2003). If they were taken as no more than artificial senses, and if their functions were limited to stimulate brain areas, then the process could be shortened and made simpler, but with less dramatic consequences – for example, by implanting a matrix of electrodes on the retina it would be possible to stimulate the visual cortex and thus avoid a detour through tactile stimulation to recreate vision (See Wyatt & Rizzo, 1996). But SSDs do a lot more than stimulate brain areas; they modify brain circuits and the result is a transformation of the body-schema and potentially the observed-self. Another crucial factor to consider is that the device has little value if the user is not in control of its actions (See Lenay et al., 2003). The real value of the SSD lies in allowing the user to actively control its movements and in this manner proactively contribute to the development of the new way to perceive the world. It is only through the user's intentional actions to explore the environment that the machine gains the state of proximal stimulation and becomes integrated into the body-schema. When integration is finally achieved, signals from the SSD generate profound changes in neural networks, and the cross-modal plasticity

[16] Cortical dominance occurs when cortical activation from a new sensorial source generates an experience that is normally associated with activity in that area - the cortical area is said to dominate, that is, to retain its normal qualitative expression. An example would be tactile perception dominating over visual perception producing the feeling of a phantom limb. In cortical deference, a particular area appears to take its qualitative expression from the non-standard input source, thus deferring to the new source. An example would be tactile neural areas complying with visual tasks in Braille.

of the brain then allows for the formation of an integrated new neural circuit adapted for this particular function – which is different to the circuitries of the substituting or the substituted sense. Once brain exaptation is in place, let's say, for example, a new way to 'see' through skin vibration, it can result in the recruitment of other relevant neural circuitries. This was in fact validated empirically; when the coupling device for tactile stimulation in a visual-tactile sense conversion experiment was moved from the chest to the back of the user or the camera moved from hand-held to the glasses, adaptation to the new situation was virtually instantaneous (Lenay et al., 2003).

Therefore, question 1, 'How are SSDs integrated into the body and modify the body-schema?' above could be answered as follows. Through use and habituation, SSDs lose their status of being distal attribution objects and are integrated as proximal stimulator in the body-schema. The integration leads to brain exaptation and the formation of novel neural circuits specialising on the newly adopted sense.

2.4 What are the neural correlates of SSDs integration?

This is the second question above, and I would like to address it by first considering research based on embodiment (van Gelder & Clark, 1998), and brain plasticity and brain-areas integration (e.g. Thompson & Varela, 2001). The modern trend has been to consider the mind as entailing embodiment and situatedness, that is, located in the interplay between brain, body and environment. This requires a methodology like 'Dynamic-Sensorimotor-Theory', which can be described as a process that takes place when movement is produced by the interaction of dynamic sub-systems located within the person and in the environment[17]. Applied to the observed-self, the interplay can be interpreted as suggesting that conscious perception should be found in the interactions between sub-systems that include neural substrates, the body-schema and the environment, and since all sub-systems are dynamic, all participants (including the SSD) in the embodied situation play crucial active roles. Non-linearity means that perception is not developed in a continuous manner or at a steady rate, but rather as dictated by deliberate changes in any of the intervening sub-systems. This indicates that neural substrates are both the cause and the consequence of objects that we perceive, and can

[17] The interplay between brain, body and environment is a notion closely related to the notion of dynamical systems account of sensorimotor contingencies, which explains perception as a form of 'know-how' that emerges from lawful regularities in the sensorimotor flow in an active and situated agent. See 3.4 for an explanation of how the mechanism for SSD integration into the body-schema could be explained as an unconscious ability to form sensorimotor contingencies.

be altered by SSDs – it also implies that the newly developed way to perceive the world is possible due to the properties of the device itself (as a potential proximal stimulator).

In line with the dynamic sensorimotor theory, the observed-self can be seen as part of a not ordered, unpredictable and irregular system of interaction between multiple interdependent components engaged in mutual interactions. Such argument is essentially similar to the model of dynamic consciousness proposed by Thompson and Varela (2001) and can explain the observed-self as a reciprocal two-way relation between sensory-motor and neural network activity within a cross-modal brain-body-world interaction system.

Also important for the purpose of this thesis is Thompson and Varela's claim that cognitive acts entail transient integration of *widely distributed* brain areas (See also Damasio, 1990). If large brain areas are involved and required in the formation of cognitive responses, it is highly likely that a similar process is needed to generate the impression of 'being oneself'. This assumption leads to the proposition that integration of SSDs interface, that is, the common boundaries between the SSD and the user that allows human and machine to interact, may not be sufficient to generate the sense that the device is a part of oneself because widely spread brain areas need to become integrated as well. Therefore, perceiving a SSD as part of 'me' is not easy; only when the newly formed neural network has become sufficiently structured and operational to process primarily the new sense, the impression that the SSD is part of the observed-self is presumed to be produced[18].

Brain substrates thus involve localised areas but also widespread connections; once in place the neural network is an established feature of the body-schema as well as of the body-model. Perhaps the most remarkable experiment on how tools can generate relevant brain substrate adaptations is the one reported by Maravita and Iriki (2004) – briefly mentioned earlier. The experimenters trained macaque monkeys to use a rake as a tool to get food pellets beyond their hands' reach, and measured the neurological changes in pre-motor parietal bimodal neurons from the intraparietal cortex (responsible for integrating somatosensory and visual information). Results showed that after training, some bimodal neurons in the visual receptive fields of the hand (which also included a small area around the hand) expanded to include the area of the entire rake.

[18] It could be suggested that brain areas compete for supremacy to gain access to the observed-self. Further studies should explore whether this competition could explain how neural centres such as the visual and auditory networks have become established through evolution.

The second question, ‘What are the neural correlates of SSDs integration?’ could thus be answered as follows. The neural substrates of the body-schema are modified together and in conjunction and cooperation with SSDs in a process that involves a non-random, that is, purposeful, intentional and dynamic interaction between brain, body and environment. The interaction can result in the SSD becoming a proximal stimulator and cause modifications in neural areas. As in cognitive acts, the integration is expected to require the involvement of widely distributed brain areas. Neural correlates should involve not only widespread neural circuits, but these circuits should function as a united sub-system to process information related exclusively to the use of the integrated SSD.

Although SSDs can cause modifications in the neural substrates of the body-schema, this is only indicative of objects being integrated into the body-model, but not into the observed-self. To find out how a SSD could be fully incorporated as part of ‘me’, we need to answer the third question: to what degree are SSDs incorporated into the body-model and transform the observed-self?

2.5 Gradients of integration

The reason that SSDs are integrated into the observed-self is that by influencing the body-schema they modify the sense of body-ownership; the difference between feeling the device as an extension of the body or as incorporated as part of the first-person ‘I’ is basically how much it is felt as ‘owned’ by the body. But this integration is subject to constraints by the body-model (the enduring and constant conscious perception of bodily states) and therefore, to understand the gradients of SSD integration we need to first explain how the sense of ownership reorganises the body-model.

The process of incorporation into the body-model could be hypothesised as involving four stages, with the sense of ownership being modified gradually through the stages. The initial stage takes place when the SSD is installed on the user’s body and gains some meaningfulness for the person; the second phase would take place when the device is transformed into a proximal stimulant. This would be followed by a stage when the SSD influences the body-schema and the body-model to the degree that it is felt essentially not different to any other bodily sense but still intellectually recognised as a foreign object. Only when a SSD is considered by the user as a ‘natural part of me’ is complete incorporation achieved, and this is the last stage. As we can see, as de Preester and Tsakiris (2009) pointed out, the gradients of integration from bodily extensions to body-incorporations are very profound.

The initial stage of integration occurs when interactions between the person and the SSD cause a basic sense of ownership that transforms the device from being a neutral object to acquire some meaning for the user – it becomes 'self'-relevant (See Beggan, 1992 and Belk, 1988 for explanations about how objects become psychologically meaningful). This stage of integration has been found to give rise to some degree of neural correlates that recruit cortical midline structures, in particular from the medial-prefrontal-cortex (Northoff et al., 2006; Northoff & Bermpohl, 2004).

During this first stage, the SSD is felt as a distal attribution tool, but with practice it gradually becomes a feeling of being an extension of the body and then as incorporated into the body. The transition, however, has been only vaguely explained (de Preester & Tsakiris, 2009), thus making the second stage rather difficult to understand, which is unfortunate because this is a crucial stage in the incorporation. In general, it could be assumed to take place as follows. Once the SSD gradually acquires the status of being owned by the user's body, it starts generating modifications in the neural substrates of the brain by recruiting neural coalitions mostly from the frontal and sensory areas. The circuitries that are most apt for the new perception are integrated; being the most appropriate they are also 'the best competitors' to fulfil the functions of evaluating and processing information captured by the device.

The process of integration should be seen as 'global' in the sense that it involves spread-out circuitries (Thompson & Varela, 2001), a process that has been studied under the name 'global workspace theory' (Baars, 2003). According to this theory, different neural coalitions compete for supremacy to new stimuli, and the best competitor, the one that 'wins the competition', gets 'broadcast' by frontal areas thus gaining access also to neural systems that '*report, reason, evaluate, decide and lay down episodic memories*' (See Kiverstein & Farina, 2012, 34, for a more detailed explanation). Global workspace has been explained with an analogy. The spotlight in a theatre stage illuminates only the main feature of the play whilst leaving all other still relevant factors in the (unconscious) dark. But the stage play involves everything, including the background. In brain activity, the 'spotlight' would be activation of the sensory cortex by the SSD, and 'the background' would be the broadcast (using cortico-cortical and cortico-thalamic fibres) to neural networks. The observed-self could perhaps be interpreted as uniting the overall architecture (the spotlight, the audience, and all that is involved in the stage play) and integrating the large number of components, including the neural network involved, the sensors, and, when incorporated, also the SSD (See Baars, 1997 for an explanation with similar connotations).

The result is that the body-schema is transformed, and it could perhaps be hypothesised that the reason that a SSD moves from being perceived as distal attribution to being sensed as proximal stimulant is that it plays a more central part in this global process.

The transformation could also be seen in the epistemic capacities of the user (the ability to acquire or develop new knowledge). Integration of the SSD has an effect on the person's knowledge and beliefs and thus influences epistemic functionality (See Matthen, 2005 for an explanation of how global broadcasting can influence capacities such as re-identification, classification, grouping and tracking).

As the SSD becomes more deeply integrated into the body-schema, through practice and habituation the third stage may follow. This could be seen as a move from the stage when the device is perceived as part neural and part technological (Kiverstein & Farina, 2012) to mostly neural, unconscious and more phenomenologically integrated. This phase of integration would be characterised by feeling the SSD as essentially not different to any other bodily sense except that it is still recognised intellectually as a non-natural object. It could perhaps be argued that the SSD has become so deeply incorporated into the body-schema that it has altered the body-model and it is now supervening on the brain – at this stage the brain can predict the consequences of the movement of the incorporated object (Kiverstein & Farina, 2012).

Studies on SSDs have mostly addressed integration only up to stage 3, with little attention to what transpires during the stage when the SSD is *intellectually* as well as phenomenologically accepted as part of the body; therefore, the final stage can only be a speculation, albeit an important one. The purpose of including this fourth phase of integration is twofold: firstly, I would like to suggest a possible criterion for full integration; and secondly, use this assumption to propose future research considerations.

Consider how a person may perceive a false denture, an artificial arm or a TVSS that has been used for a very prolonged period, and compare this with how a tattoo that was done during the childhood-amnesia period (that is before the age of approximately three and a half years) may be perceived. Which one would be recognised *intellectually* as well as phenomenologically a part of ‘me’ and which one wouldn’t? It could be argued that the denture, the artificial arm and the SSD are perceived as non-natural and therefore as not fully incorporated intellectually. The person’s beliefs therefore can make a difference, and intellectual incorporation cannot be achieved unless the belief that the object is a natural one is in place. Notwithstanding this apparent limitation, a SSD that has been used for a sufficiently protracted period can be felt as part of the observed-self, indistinguishable phenomenologically from any other sense, and it seems therefore in order to grant the device the status of full incorporation, but only at phenomenological levels.

Having intellectual property over a SSD, however, does make a difference (I will explain the reason more in depth in Chapter 3 with the example of Otto and Inga). Consider the hypothetical scenario where a device similar to a SSD is implanted during the childhood-amnesic period and that as a result this person grows up with this device being a fully intellectually and phenomenologically incorporated proximal sense. Consider further that the sensor of the device is portable and can be moved to other places to capture objects in distant locations. There is nothing theoretically that would prevent this individual from being present *in person* at the place where the sensor is, and the objects in that environment to be perceived as nearby objects. The theoretical repercussions of considering a stage four of integration could therefore be significant for a likely (not so) future scenario possessing such technology.

I would like to close this Chapter with a summary of the main findings from SSDs research, and point out some associated theoretical considerations that can be used to generate a model of the mechanics of consciousness and the observed-self.

2.6 How studies on SSDs can provide material for a model of consciousness and the observed-self

The most important research finding that provides empirical and theoretical material to conjecture a theoretical model for the mechanics of the observed-self are findings that, with practice, the interface between a SSD and the user fades away and the device is integrated into the body-schema of the person (Auvray et al., 2005; Kiverstein & Farina, 2012; Bach-y-Rita & Kercel, 2003). Considered together with the argument that mediated (by SSDs) and unmediated (natural senses) perception follow similar mechanisms (Auvray et al., 2005), these studies provide evidence to suggest that the observed-self can be understood as emerging from the perceptions captured by natural or artificial senses alike.

Another piece of evidence to consider is that SSDs can cause recalibration of brain circuits (e.g. Clark, 2008) and that with practice the device can restructure the functioning of neural paths (e.g. Maravita & Iriki, 2004). These findings are compatible with, and offer grounds to support the notion that the observed-self is the result of cycles of afferent and efferent signals between brain and periphery (Loomis 1992).

We can use this line of reasoning to suggest that the mechanics responsible for SSDs integration are similar to how our natural senses developed through the process of natural evolution. If this claim is valid, it would indicate that we generate our separation from the world through a sense when the brain has allocated a particular neural area that responds to the interface point of that sense, that is, when the sense has become a standard or 'conditioned' proximal sensor and thus part of the observed-self – in natural evolution, the elements for integration may be primarily internal to the body, but the integration could be understood as following similar mechanics as SSDs incorporations.

Another piece of evidence that could prove useful for the proposed model of consciousness and the observed-self is the criteria for amalgamation of SSDs described above (Grice, 1962). In Chapter 3, I will use this evidence to propose that what SSDs do is essentially assist the observed-self gain new modes of dealing with environmental demands.

Chapter 3: The mechanics of SSD integration into the observed-self

3. The foundations of the observed-self

In this Chapter, I describe the mechanisms responsible for integrating SSDs into the body-schema in preparation to then formulate the foundations for a theory of consciousness and the observed-self based on these mechanisms. Prior to addressing them, I would like to recap on the information presented in Chapters 1 and 2 and summarise the main hypothetical assumptions upon which the theory should be built. After that, I will consider theoretical and empirical evidence from SSDs research that can assist validating my assumptions.

The first and most elementary theoretical assumption is that the observed-self is the perception of *attributions to oneself* as separated from what we sense as distal attributions – for the purpose of interacting with the world and thus evolve. This notion is by no means new but it has been typically explained in ways different to how it is addressed here. Its essence, for example, has been explained in terms of a minimal-self, an immediate subject of experience unextended in time (Gallagher, 2000), and as a narrative-self, a subjective history of connected events essential for the person (Heersmink, 2017). It has also been described as a social intersubjective construct (e.g. Mead, 1962) even though it possesses the property of being independent (e.g. Gallagher, 2000). In Chapter 1, I pointed out that these traditions fall short of explaining how a reflexive and minimal self can give phenomenological experiences and that a remedy for this is to suggest the existence of an observer-self that possesses the property of sensing the observed-self. In this Chapter, I will expand on this suggestion and on the properties of the observed-self to incorporate objects such as SSDs.

A second crucial assumption is that the observed-self is generated by afferent/efferent neural circuitries, and that conscious awareness of it emerges from neural *efferent* substrates only; afferent signals are presumed to be unconscious. I hypothesise that when operating in an open-mode the observed-self senses the outside world, and in a closed-mode it processes this information internally. Based on these assumptions, I suggest a general *schematic representation of consciousness*, showing the place that the observed-self occupies in it.

The third and perhaps most critical conjecture for the purpose of this thesis is that since the observed-self is presumed to be the inside world of our periphery senses, bodily sensorial interfaces mark the boundaries of private-consciousness, and objects that are integrated into the body-schema can potentially become part of this self[19]. This claim presumes that integration of SSDs into the body transforms the observed-self in stages that range from initially attributing the SSD the status of distal attribution (not being part of 'me'), then becoming proximal stimulants progressively modifying the body-schema and the body-model and being experienced as 'me'.

A fourth and last main speculation is that SSDs generate modifications to existing neural circuitries and that as a result our cognitions also change.

3.1 The observed-self functions to separate 'me' from 'otherness'

A basic point of agreement amongst modern views is that selfhood is not innately fixed or automatic but developed through intentional interactions with the world (Zahavi, 2014) and this implies that there is a constantly changing demarcation between what has been studied under the notion of 'subjectivity' and the objective world. It is this demarcation that causes a person to feel as an individual with self-identity – and what inspired Sartre (1956) to believe that we express agency by giving ourselves experiences. There is thus intuitive consensus that a primary function of the self can be presumed to be interacting with the world, essentially satisfying the first theoretical assumption in section 3. above. I will now consider more detailed evidence, but this requires recapping momentarily on some crucial points.

Explaining the mechanics of two separate selves requires considering the self as rather differently to how it has been typically studied. In this dissertation, I hypothesise that the self is divided into an observed-self, of which we are consciously aware as representing 'me', and an intrinsic observer-self, of which we are unaware. Subjectivity would be perceiving the observed-self, and inter-subjectivity or objectivity is the perception of what this self perceives (Loomis' extensions of body effectors; Loomis, 1992).

Also, the emphasis is this dissertation is on explaining the difference between the hypothesised *observer-self* and what it observes, an issue that can be seen as similar to the traditionally rejected notion of a homunculus. There are, however, important reasons for not associating the observer-self with 'the man in the machine'. The observer-self is rather a direct representation of the first-person 'I', and manifests in the thoughts, feelings, actions

[19] Here I am basing my assumptions mostly on the works of Auvray et al. (2005); Kiverstein and Farina (2012), and Bach-y-Rita and Kercel (2003).

and perceptions that we have (explanations other than those offered in Chapter 1 about the observer-self would be outside the scope of this short thesis).

The separation between 'me' and otherness is thus explained as a separation between the observed-self and what this self observes; or put differently, the perception of 'me' is in reality the perception of the observed-self but not of the observer-self (which cannot be perceived, just like we cannot perceive our own eyes). Understanding the separation between the observed-self and otherness is therefore understanding what the intrinsic observer-self perceives as 'me' as separated from the world. But this reasoning requires that we overlook for a moment the traditional notions of a 'pre-reflective' minimal-self because presupposing the idea of a non-intentional and reflective self begs the difficult question of how it can then generate intentional rationality. Perhaps more importantly, considering the idea of a minimal self that can give us the experience of 'being conscious of something', but which is also itself experienced, an 'inner-consciousness' (Husserl, 1991) could be seen as conflating experience and the experiencer; it is also a confusing argument and possibly circular. Moreover, if when integrated a SSD becomes part of the self, then according to this reasoning it should be considered as 'an observer observing itself', which seems unacceptable. I would rather propose that to separate 'me' from 'otherness' we should accept the view that an intrinsic observer is required, and that we need to look for the evidence or its existence in a way that is different to a homunculus.

I suggest that a possible answer could be found in the argument that a self is a postulation of a unique point of view (Searle, 2005), which is similar to Nagel's (1974) claims that what it is like to experience something is only accessible to the organism having the experience (to feel what a bat feels for example, you would have to be one). It could be argued that if an object exists only from the point of view of the experiencer, then the notion of true objectivity comes into question because it renders objective observations unviable; nonetheless, this reasoning provides a possible answer for the self/non-self distinction proposed here. Searle claimed that the self cannot be consciously experienced, but that since we would not be able to perceive things consciously without it, it can be inferred to exist. He seems to be implying that the self itself is the intrinsic experiencer, a contention that appears to be compatible with my presumption of an observer-self acting as a processing-sensor. I would suggest that the best way to study this intrinsic self is not so much by analysing the notion of subjectivity and objectivity but by describing what the observer-self *is* and what it *does*.

Unfortunately, the answer is difficult to find in current work, but we can make some inferences from existing views. A line of study by Alvin Goldman (1970) focussed on the distinction between 'reflective' and 'non-reflective' consciousness: consciousness as a mental state whose object is the experiencer can be seen as reflective; and a mental state that does not focus on the experiencer's internal states can be considered as non-reflective. These terms correspond roughly to what Uriah Kriegel (2004) described as 'transitive self-consciousness' and 'intransitive self-consciousness' – transitive consciousness is consciousness with 'relational' properties whereas non-transitive consciousness has no relational properties. The distinction, though, does not resolve the issue of who the observer is because both reflective and non-reflective experiences are observations and do not identify the observer.

I would propose that to add clarity to this matter the crucial questions to address should be 'Who or what senses these states of consciousness?', and 'How does this observer observe and for what purposes?' Since available theories do not seem to address these questions in sufficient depth it is necessary to speculate. What the observer-self is was briefly described in Chapter 1. To recap, I conjectured that it is a separate entity in its own right that 'emerges'[20] from the sum total of bodily functions united with synergy. And I asserted that what we perceive as the observed-self is the awareness of what this observer perceives as occurring within the boundaries of the body – this would correspond in some sense to the 'non-reflective self'. 'Otherness' is awareness of extensions of bodily effectors, and would correspond somewhat to Goldman's 'reflective self'.

Having suggested tentatively what the observer-self *is*, we can use this notion to explain what it *does*. The answer I propose is that since the observer-self is a processing-sensor in its own right, it functions to capture stimuli from both the environment, which we perceive as distal attributions, and the body, which we attribute to ourselves. Based on this reasoning, Gallagher's (2000) minimal self should perhaps be understood as the faculty of being conscious of 'attributions to oneself', and Heersmink's (2017) narrative self as perceiving the history of the connections between the observed-self and otherness.

[20] The term 'emerges' refers to the *emergence-principle,* explained in point 3.2.

3.2 Evidence that the observed-self emerges from neural efferent substrates, and how the evidence can be used to generate a schematic representation of consciousness.

The assertions above suggest that the observer-self captures what we sense as our field-of-conscious-awareness. I would argue that the observer-self is the intrinsic intermediate agent connecting the private with the extended parts of the mind – for the purpose of interacting with the environment and thus evolve[21]. In this segment, I am exploring evidence to support the idea that the field-of-consciousness is the full spectrum, consciousness as a whole (including unconscious processes) generated by afferent/efferent neural circuitries, but that *conscious awareness emerges only from efferent neural substrates*. Based on these assumptions, I will suggest a tentative general 'schematic representation of consciousness' and infer the place that the observed-self occupies in it.

To understand the mechanism responsible for the perception of the observed-self it helps if we compare it with the notions of 'neural correlates of consciousness' and 'emergence'. When describing the neural correlates of consciousness, Chalmers (2000) pointed out that they are immersed in a one-way causal-explanatory relationship from neural to conscious experience – a causal approach. Conversely, adopting a radical-embodiment view, Thompson and Varela (2001) suggested correlates as two-way reciprocal relationships where local and global systems influence each other – a constitutive approach. Although seemingly contradictory, the apparent differences can be removed if we look at the process as comprising information always travelling *unidirectionally* with afferent signals (the input) always travelling global-to-local (body to brain to neurons), and efferent signals travelling local-to-global (neurons to brain to body). This line of reasoning is in agreement with well-established empirical evidence demonstrating that what we perceive consciously, including our decisions, comes *after* a 'readiness-potential' in brain activity (Libet, 1985), and therefore we must presume that our conscious field-of-awareness emerges from efferent signals or upward causation *only*[22].

The process of unidirectionality, however, cannot be restricted to CNS-periphery, but it must be presumed to incorporate all possible components, from elementary sub-atomic particles

[21] I am assuming that the observer-self should be understood as 'me the observer' or 'me, the person that I am'. I am also assuming that consciousness can be broadly divided into a 'private' part, corresponding to the observed-self, and an 'extended' part, corresponding to otherness.

[22] The proposed model in this thesis demands that downward causation must be a subconscious process because our conscious decisions are generated after and as a result of unconscious brain activity.

(the most basic building blocks of matter) to the most global (the whole universe), with the person and the situational environment being parts in this context. The model I propose to explain consciousness and the observed-self is based on this reasoning, and postulates that brain activity not only generates our field-of-awareness but also *changes the state of global systems*[23] (including the whole universe), and conversely, global systems influence the state of the brain (Fig. 3). In this manner, with information travelling unidirectionally, afferent signals intake global-to-local information[24], processing them internally and generating an efferent local-to-global response. During recursive loops brain, body and environment generate changes on each other whilst still being dynamically integrated constitutively as part of the same global system (See Palermos 2014 for a discussion on how continual mutual interactions and non-linear relations operate within a dynamical system theory).

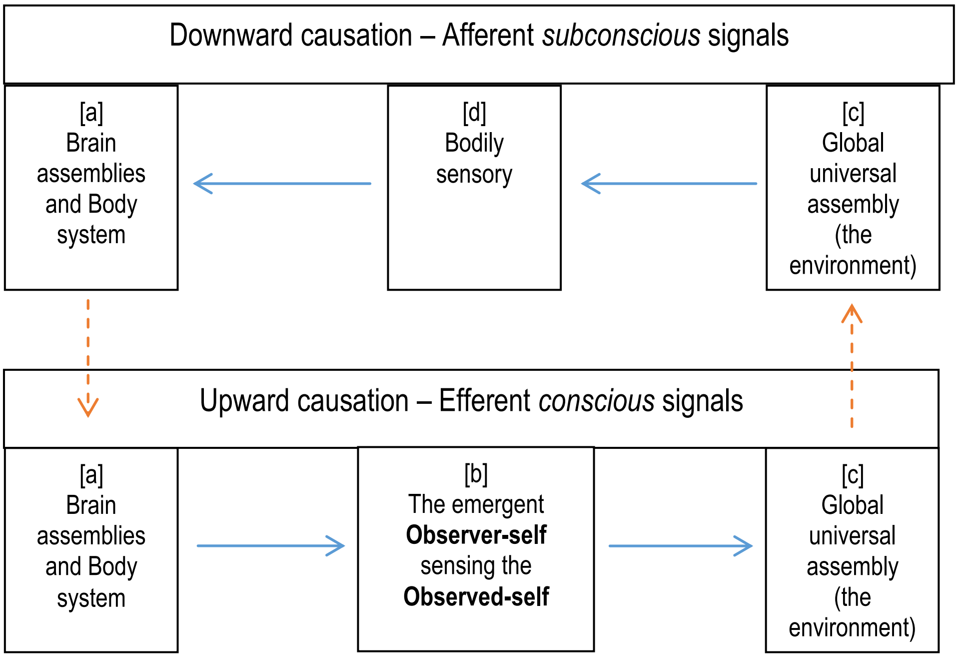

Fig. 3: ***Schematic representation of consciousness and the selves***. *Downward causation is comprised of subconscious information from global to local assemblies. Upward causation is comprised of conscious information from local to global assemblies. The observed-self emerges from local brain assemblies in information that is felt phenomenologically by the observer-self.*

[23] An assumption required by this model is that an observation changes what is observed.
[24] von Holst (1954) uses the term 're-afferent' to describe downward feedback signals.

The similarities between this reasoning and Thompson and Varela's concept of 'reciprocal causation' are apparent, but with a twist because my proposal requires ultimate feedback from an all-encompassing global universal system. Let's consider these apparent similarities within the concept of *emergence*[25].

A term seemingly coined by psychologist George Henry Lewes in the nineteenth century, 'emergence' is associated with an even older concept, probably attributed initially to Aristotle**,** 'a structure is more than the sum total of its components', and denotes that properties of a system are unlike properties of any of its component parts. This is a central concept in the claims I make in this thesis and an important aspect in the schematic representation of consciousness and the observed-self in Fig. 3. It denotes that although the brain, body and environment are systems that emerge from the parts they contain, they are independent systems with their own qualitative attributes and properties. The state of the brain, for example, emerges from the collective behaviour of local brain assemblies, but as a holistic system it generates efferent signals that represent the brain as a unit. Efferent brain signals together with and as part of the collective behaviour of all other bodily parts give emergence to an independent body-system (Fig. 3, [a]). A crucial assumption in this thesis is that the observer-self is this qualitative emergence, and that the observed-self is what this emerging system perceives as conscious awareness ([b]). The body as an intrinsic observer then influences the environment ([c]) whose feedback subsequently dominates local bodily sensory ([d]) and through afferent signals the state of the brain and body ([a]). In this manner, brain-body-environment relationships are all parts of the same constitutive coupled-dynamic-system, which exhibit self-organisation and emergent processes at multiple levels (See Kelso, 1995; Port & van Gelder, 1995 for similar views on multilevel emergence).

The dynamics of the system can be envisaged as comprising two broad stages. In the downward phase, unconscious signals from global environmental assemblies find their way to the inside of the brain, 'enslaving' its component parts and changing their behavioural alternatives. In the upward phase, these changed brain sub-components modify the brain state as a whole[26], which in turn contributes to changing the state of the body-schema, the body-model and the consciously observed-self. We, that is, the first-person 'I' or observer-self, are

[25] The 'Emergence' paradigm is closely related to 'Emergentism', a worth-researching (for future studies) discipline now resurfacing as a possible method to explain mental causation from physicalist frameworks (Moore, 2015).
[26] The process of brain components changing the state of the whole brain is similar to what has been referred to as 'collective variables' or 'order parameters' (Thompson and Varela 2001).

oblivious of these subconscious processes, and what we sense is the complete phenomenological field-of-conscious-awareness that ‘emerges’ from it. When we decide to attend to events outside the boundaries of the effectors, we are essentially attending to distal-attribution objects, and when we decide to introspect, we perceive attributions to ourselves[27].

3.3 How the proposed model explains integration of SSDs into the body-schema

The diagram in Fig. 3 indicates that since information within the dynamic whole travels unidirectionally in loops, all participant sub-systems influence the rest (and in turn, themselves). Through our intentions ([b]), for example, we can consciously and voluntarily alter brain activity ([a]); albeit this requires first modifying the global environment ([c] – See Leopold & Logothetis, 1999 and Penfield & Jasper, 1954 for evidence that people can purposely generate neuronal biases and control brain activity in binocular rivalry and epilepsy). This is thus an enactivist claim because it recognises that experience arises from an interaction with the environment; it is not something that just occurs to a person but the product of what the person does with intent.

Users of SSDs enact the experiences produced by the devices by actively and purposely participating in the integration of the device into the body-schema; more specifically, by displaying intentions and skills that produce the feedback needed to ‘know how’ to use it. It is conscious intents that cause the SSD to instigate the neural modifications required to alter the body-schema; when these modifications occur they cause exaptation of an already established meta-modality (See Nagel et al., 2005 for empirically based related comments).

The same process applies to the use of tools, albeit with a lesser degree of integration. Let’s consider for a moment how the model in Fig. 3 can accommodate enactive integration of tools into the body-schema. There is evidence that with regular use, the tip of a tool, a pair of tongs for example, can be felt as a functional extension of the body (Riggio et al., 1986) and can activate regions in the intraparietal sulcus (Inoue et al., 2001). The diagram would explain the mechanics as follows: A conscious intention ([b]) to use the tool acts as the catalyst for an upward causation influencing the state of the environment, including the state of the tool it contains ([c]). As a result, the tool becomes more ‘self-relevant’ and psychologically meaningful for the person (See Beggan, 1992 and Belk, 1988 for a description of an analogous process in other circumstances) impacting on the person’s senses

[27] See Thompson and Varela 2001, box 1, for a description of an emergence process that can be viewed as having similar connotations.

([d]) more significantly and causing increasingly profound alterations in neural substrates ([a]). Repeated loops lead to modifications of the body-schema and the tool becoming incorporated into the plastic neural representations of the body.

A similar process should apply to more complex mechanisms such as SSDs, and in fact this has been supported by modern research. Huttenlocher (2009), for example, reported evidence that, due to the cross-modal plasticity of the brain and the property of cortical areas to take multisensory functions, SSDs can unmask (unconsciously, I would argue) cortical potential to process information from different sensorial modalities. Another piece of evidence suggested that SSDs using audition to perceive visual inputs activate the visual cortex (de Volder et al., 1999) – the activation has been viewed as the result of a two-way reciprocal relationships system where local (neural) and global (environmental) systems influence each other, an interpretation that is quite similar to the model proposed here, except that our model implies loops of information travelling always in one direction.

The model in Fig. 3 can be seen from a different angle as a closed-loop system that when engaging the brain it produces *sensation* (sensorial stimuli triggered by the SSD), *decision* (brain feedback that manifests as our thoughts and feelings) and *reaction* (our behaviours and further intentions, for example, to continue using the SSD). Within this sensation-decision-reaction sequence closed-loop, a SSD acts upon a pre-existing neural network capable of responding to afferent inputs emerging from the device ([c-d] – sensation), which results in generating efferent outputs tailored to accommodate the newly adopted sense ([a-b] – decision). Changes in afferent/efferent signals and the body-schema lead to reactions and new decisions ([b-c] – action).

The model in Fig. 3 also accommodates the suggestion that what causes a SSD to lose its distal-attribution status and become a proximal stimulator could be that with practice the device changes efferent signals that control body activity from being voluntary to involuntary, and, further, ignores the SSD/body interface (See von Holst, 1954 for an explanation of these mechanics). The stimulator of the device is initially perceived consciously ([b]) then becoming unconscious ([d]).

3.4 Effectors mark the boundaries of the observed-self

Perhaps the most critical conjectures in this thesis are that effectors mark the boundaries of the observed-self and that SSDs gain access to the body-schema and the observed-self by acquiring status as proximal stimulators, thus functioning similarly to how natural senses do. Since our senses (eyes, taste buds or smell receptors, for example) are proximal stimulators (Harman, 1990) we are not aware of them, only of what they capture, and one way to study proximal stimulation is by manipulating the sensory interface. This is difficult with our natural senses, but as SSDs produce similar effects (Siegle & Warren, 2010) studying their process of integration are appropriate means to study these critical assumptions.

One question (not so crucial to this thesis) that Siegle and Warren (2010) raised was whether the strategy for integration is primarily *cognitive* (we learn to 'infer' what the device does) or *perceptual* (the device acts as a normal sense and we reflexively attend to what it does). Their results favoured the latter and support the notion that SSDs prompt us to instinctively rather than rationally sense the environment through the device, thus demonstrating more similarities between SSDs and our natural senses[28].

The mechanism for SSD integration into the body-schema could also be explained as an unconscious ability to form 'sensorimotor contingencies' (O'Regan & Noë, 2001). According to this view, awareness of distal objects is produced by the intrinsic knowledge of the laws that govern perception (body movements, for example, allow us to unconsciously use these laws to locate objects and judge their distance). Users of SSDs achieve integration when, by unconsciously gaining knowledge of these laws, they cease to consciously pay attention to the SSD's stimulator (e.g. on the skin) attending instead only to the objects that the device captures (Siegle and Warren refer to the former as 'proximal attention' and the latter as 'distal attention'). Once this intrinsic knowledge is attained, it is maintained and generalised; if the stimulator of the device is placed on another part of the body, for example, the transference to the new site can be quickly learned.

However, the most crucial concept in SSD-body integration is probably that it changes the sense of ownership of the body. Different to tool use, which at best may be felt as extensions of the body, SSDs can produce true incorporation because they modify the feeling of ownership itself. De Preester (2011) considered these modifications as occurring primarily at the levels of *motor*, *sensory* and *cognitive* capacities. Motor capacities are those related to motor actions, such as moving around, and sensory capabilities refers to how the user of a SSD perceives the device. Cognitive extension is not directly related to how the body's capacities are altered, but rather to the cognitive/intellectual consequences of the

[28] There is now wide acceptance that SSDs and natural senses follow similar processes.

modifications. And all are matters of *embodiment relations* because they alter our sense of body-model and how we use our body.

Studying how SSDs modify the observed-self by modifying the status of the interface requires understanding the distinction between what represent extensions and what represent incorporations, because only the latter generate the feeling of being 'me'. The difference lies in the degree of integration, which is reflected in the 'transparency' of the device; low transparency is an indication that the device is quite detached from the body (Ihde, 1979) – Maravita and Iriki's (2004) experiment training macaque monkeys to use a rake as a tool would be an example of low transparency[29]. Conversely, if removing an integrated device produces an acute sense of loss it is an indication of deep incorporation – a prosthetic arm, for example, could be a lot more transparent than the fork or the knife that we use for eating.

Therefore, effectors connected to the SSDs can be presumed to delineate the boundaries of the observed-self by regulating the degree of transparency and ownership of the device. An intermediate state of integration would be when the SSD is felt as part of the body but still recognised as an artificial annex, whereas a fully incorporated device would be one that the user not only feels as part of the sensorimotor apparatus of the body but essentially not different to any other sense (the differences will be explained in a bit more retail in point 3.5). A good example of deep incorporation can be seen the rubber-hand illusion (described earlier). What creates the illusion that the rubber hand is part of the observed-self is not simply the synchronicity of strokes, but that other factors are also respected. The rubber hand has to look like a hand (body identity), be placed in the right position as a body part (body specificity), and be of the same handedness (anatomical constraint). Similarly, creating the feeling of ownership over a SSD requires that the device be 'accepted' into the normativity of the body-model; only then it can gain access to be an integral part of the observed-self.

There is another factor that we need to take into consideration to understand how effectors mark the boundaries of the observed-self; SSDs are not really extensions but rather incorporations that change the state of the body-model. What generate the effector-limits of the observed-self are the established neural networks caused by the incorporation; it is the property of these incorporation-related networks to maintain stability over time that produce the feeling that the arm has been replaced in the rubber-hand illusion, and that a phantom-

[29] The rake used by the monkeys maintained low transparency probably because being temporary the action did not seem to have created a feeling of deep body-ownership of the tool.

limb is still felt attached to the body. It is for the same reason that the body-model tends to prevent objects that do not replace parts of the body to become incorporated (de Preester, 2011) and that obtaining full SSDs incorporation is difficult. Not surprisingly, since the neural substrates that create the phantom-limb experience or the rubber-hand illusion are already in place, these effects are felt quite rapidly, whereas incorporation of SSDs, which require the creation of new neural substrates, is achieved quite slowly and usually cumbersomely.

The success of integrating SSDs into the body-model is thus dependent upon how they become incorporations rather than extensions of the body. Notwithstanding this, it is degrees of integration that leads to incorporation, and to properly understand how the observed-self operates within the boundaries of the effectors we need to consider evidence supporting the stages of integration. Stage I takes place when the SSD device is installed and the device is still treated as a completely external object. The second stage ensues when the device gains some transparency leading to the status of proximal stimulation. This is a more typical stage of 'embodiment relations' (Ihde, 2002) or 'human-machine' relation, when the feeling is perceiving through the machine in a partial-transparency relation (de Preester, 2011) but with the machine being not deeply incorporated into the body-schema. Examples would include spectacles, microscopes, or hearing aids, but also actions such as feeling the car as a symbiotic extension of oneself (Ihde, 1979). Stage II can be presumed to start when the SSD has gained some status of proximal stimulation producing some but not very pronounced feeling of changes in the ownership of the body. De Preester (2011) sees the transition as *exteriorization* because attention is more focussed on distal attribution; she also sees this as changing from what could be labelled as tools or instruments to be felt as perceptual 'prosthesis'.

The typical feature of stage II would thus be to perceive the SSD as receding to a background and being gradually replaced by the experience of what it captures, although still possessing primarily properties of a tool. During this stage, the device could be seen as generating unconscious changes in the body-schema but not so pronounced changes in the body-model.

Stage III then follows, and it is characterised by the SSD influencing deeply not only the body-schema but also the body-model. This would be rather similar to de Preester's (2011) criterion of a 'qualitative experience', the feeling that the SSD is generating a new sensation that does not belong to any known sensory modality[30]. De Preester explained that a new way of experiencing is considered as a 'noetic novelty' due to adding a dimension in mental activity, and this would be a crucial development through this stage of integration. If removal of a prosthesis (or in our case a SSD) *"leads to a loss in the domain of perceptual experience at the noetic side"* (de Preester, 2011, 133), it is an indication that the body-model, and by default the observed-self, has incorporated the device. At this stage, the stimulator of the device would be an established incorporation into the observed-self.

Transition to stage IV occurs when this phenomenological experience is accompanied by the belief that the device enjoys complete intellectual ownership, that is, when it is believed to be a natural sense, and this has received little empirical attention to date.

3.5 Integrated SSDs lead to modifications in cognitive capacities as well as in the observed-self

As Andy Clark and David Chalmers (1998) suggested, a crucial question that needs answering is: where do the mind and our cognitive capacities stop and the rest of the world starts? The modern trend favours the notion that cognitions extend as part of a coupled system comprising the person plus external entities, and SSDs represent an important part in this system. Extended cognitions related to a SSD; however, have to be considered within the degree of integration of the device because different degrees of integration signify different levels of cognition, and an appropriate way to understand this process and determine how SSDs transform cognition and the observed-self is by considering the notorious example of Otto and Inga[31] (Clark & Chalmers, 1998). From a functionalist point of view the subjective experience of remembering (e.g. how Otto and Inga *feel* when remembering the address of the Museum) is irrelevant, but it could be argued (in agreement with de Preester, 2011) that how Otto and Inga interpret this experience is different.

The key factor comes down again to ownership, in this case ownership of a belief (intellectual ownership), and how a belief may allocate a feeling of being 'me' as different to being 'mine' (but not 'me'). The impression of being 'me' is an indication of profound ownership and integration of the artefact, whereas 'mine' can reflect more superficial

[30] For de Preester this demarcates the distinction between perceiving the device as a tool or as prosthesis.

[31] Otto suffers from memory loss and relies on his notebook for information, as opposed to Inga who relies on her own memory. The functionalist view (e.g. Clark & Chalmers, 1998) is that both are examples of extended cognition because what is important is only the functionality factor (Otto's notebook functions cognitively just as Inga's memory does).

integrations. If Otto believes that the notebook is his but not ‘he’, then this would have a direct effect on his cognition and impression of himself, giving him the perception that the book is still an external object. And indeed, it does appear that although Otto’s extended cognition incorporates the notebook, he still recognises it as ‘otherness’. Conversely, Inga believes that the source of information is from a natural part of her brain and this makes her memories part of her ‘self’. I thus agree with de Preester that Inga has a stronger sense of ownership over her memories than Otto – I would contend that whilst Otto’s book would be an example of stage II or perhaps III of SSD integration, Inga’s memories would be compared with stage IV.

Admittedly these inferences may not be very robust because the two examples belong to different fields of study, one perceptual phenomenality (sensorial) and the other cognitive (extended cognition), which have not received empirical consideration together, but they do bring into attention the fact that SSDs alter cognitive capacities and that this alteration has direct effect on the perception of the observed-self. In a world that is moving fast towards humanistic cyborgisation, the influence of external devices on cognition is bound to gain more research attention.

The important concept to keep in mind for the purpose of this thesis is that entities that are perceived as profoundly owned by the body are considered more as part of the observed-self and are processed differently, compared with those that are less so. Available evidence, however, relates mostly to how physical objects are incorporated into the body and therefore cognitive integration is but a consideration for future studies.

In this chapter I attempted to provide an explanation of the mechanisms responsible for integrating SSDs into the body-schema and the body-model and the resultant modifications in the sense of body-ownership and the observed-self. Chapter 4 will address ways in which this information can be used to establish the bases for a theory of consciousness and the observed-self based on SSDs research.

Chapter 4: Theoretical foundations for a theory of consciousness and the observed-self based on SSD research

4. The mechanics of consciousness and the observed-self, and how they fit in the schematic model (Fig. 3)

In this chapter, I will propose some grounds upon which a theory of the observed-self based on SSDs research can be constructed. Prior to this, though, it is necessary to suggest a more precise definition of the observer-self, suitable for the proposed theoretical framework. I will attend to this by considering primarily to the works of Searle (e.g. Searle 2005). I selected these works because they seem to offer concepts that are quite compatible with arguments in this thesis that the impression of being 'me' can be found in the internal mechanisms of the body (mostly in the afferent-efferent nerve circuitries). After that, I will re-examine the schematic model described in Fig. 3 and explain how, within this schema, a SSD modifies the observed-self, offering an alternative to current theoretical models to study consciousness and the self, and proposing possible issues for further research. I will then address the strengths and weaknesses of the model and summarise my claims.

Trying to define the self has traditionally been a daunting task and I provided arguments that the reason has been conflating the 'experiencer' and what is 'experienced'. Searle (2005) tried to overcome this problem and he proposed a description of the self that in some ways resembles the one proposed here, although, as we will soon see, there are fundamental differences. I hope to assist resolving these differences by explaining a bit more in detail the mechanics of consciousness and the observed-self within the schematic representation in Fig. 3. To assist understanding these explanations I would like to first briefly refer to views on how the self has been somewhat equated with the concept of *personal-identity*.

The concept of the self has been typically equated with personal-identity, in the hope that by resorting to the four primary criteria for describing it, the self would also be explained. The first criterion for personal-identity concerns identifying the body as being that of the same person over time, which goes together with a second criterion of the body exhibiting coherence of continuity. Another condition was that consciousness is recorded in memory, and a fourth one was continuity of personality. Although these criteria are suitable for explaining our perceptions of personal-identity, they do not contribute much to identify the mechanics of whom the observer is, which prompted Searle to say: '*I have reluctantly come to the conclusion that the nature of human consciousness requires the postulation of a non-Humean* [not based on physical experiences] *self*'(Searle, 2005, 10). What is required, therefore, is to separate the experiencer from the experienced, and for this I have offered distinct explanations for the existence of an observer-self and an observed-self. Armed with this distinction, we can now move on to consider the fundaments of SSDs integration into the body-schema and offer more precise explanations of the mechanics of consciousness and the observed-self based on this integration.

To generate a theory of the observed-self based on SSDs research it is important to identify relevant concepts that have already proven valid by consensus, and that can be applied to the model. The most critical of these concepts are that with practice, SSDs become proximal stimulators (Grice, 1962) that modify our conscious perceptions, and that their functioning follows similar mechanisms as our natural senses (Auvray et al., 2005). From these pieces of evidence, it is possible to suggest some basic grounds for the proposed theory: *A phenomenological experience (qualia) can be thought of as the product of a mechanism that starts with sensory inputs, and since SSDs can act as sensory inputs they contribute to our conscious experiences and by default to the observed-self.*

Less researched but still quite persuasive is the assumption that *an experience emerges from activity in the afferent/efferent circuitries, and that the sense of oneself is the perception of activity restricted to areas within the body and internal to the effectors.* As proposed by Loomis (1992), and based on the works of von Holst (1954), the idea that the self can be located within the afferent/efferent nerve signals independently of extensions of the effectors does not appear to have been disputed and it offers grounds to argue that the observed-self emerges from internal bodily mechanisms that can include incorporation of SSDs.

Another relevant (and well-established) piece of evidence is that our phenomenological perceptions are the result of brain activity that exhibits as 'readiness-potential' (Libet, 1985), which occur prior to become conscious, thus indicating that our experiences should be considered as the product of an upward local-to-global mechanism, brain to person.

Based on this evidence, it is possible to suggest a sequence of unidirectional mechanical events represented as follows in Fig. 3:

Bodily-sensory [d]>>> Brain-assemblies [a]>>>observed-self/qualia [b]

Having described these theoretical concepts, the next task is to construct, from the evidence, the grounds for a possible theory of consciousness and the observed-self – somewhat similarly to how a palaeontologist reconstructs a whole skeleton from a few pieces of available bones. A crucial hypothetical piece in this puzzle would be the assumption that the synergetic amalgamation of internal bodily mechanisms causes the *emergence* of an independent and self-sustained system that can be identified as the ontological first-person 'I'. Functioning in 'open-mode', this observer senses, amongst other things, biological activity, and functioning in 'closed-mode' it processes the thoughts, feelings and actions generated by the brain and body (represented in ([b]) – the 'observer-self' sensing the 'observed-self' – in Fig. 3).

Since information is presumed to travel unidirectionally, the diagram depicts the first-person, 'I', as conveying information only upwards to more global entities, generating changes in the situational environment as a whole. This means that the thoughts and intentions that we entertain should have ripple repercussions in the state of existence as a whole. The diagram postulates that once the state of the universal system has been modified by these messages, the cycle closes and a downward global-to-local response is generated, causing emergent changes in all its component sub-systems, including impacting on our bodily sensoria ([d]) and subsequently modifying brain assemblies ([a]).

What this cycle indicates is that the thoughts that we attend to more meaningfully (which represent our choice) are the catalyst for the response that in the following downward stage of the loop produce our next thoughts and intentions – which we perceived as a modified field-of-conscious-awareness. In Fig. 3 this mechanism is represented as the sequence:

Observed-self [b]>>>Global-universal-assembly [c]>>>Bodily-sensory [d]>>>Brain-assemblies [a]>>>Observed-self [b]

It should be noted that the model in Fig. 3 also requires that global-to-local information (Downward causation) travel unconsciously, and that conscious perception results *only* from local-to-global connections (Upward causation).

Viewing these mechanics conjointly, I would like to summarise some possible theoretical elementary grounds to produce a theory of consciousness and the observed-self based on SSDs research.

Consciousness is essentially the interaction between an observer-self and an observed-self, both acting as processing-sensors (further research should explore the notion that the latter incorporates as well the narrative history of connections with the outside world). The ontological first-person 'I' is the observer-self that emerges from the amalgamated functioning of all bodily parts, including integration of SSDs, united with synergy. In its open-mode, this observer intakes information from biological processes, and in the closed-system mode it processes this information giving rise to phenomenological experiences – we perceive this as our field-of-awareness and as the SSD being part of the observed-self. These experiences are then the catalyst for environmental changes, whose feedback subsequently prompts neural modifications and new thoughts and intentions in a continual loop of interactions between body and world.

Having described the essence for a hypothetical model, I would like to illustrate with a real-life example how the theory is postulated to work, and offer some suggestions for future research.

4.1 How the theory is assumed to work

Emilie Gossiaux (Fig. 4) is a painter and sculptor who became blind in 2010 after a tragic motor vehicle accident. Following her strong vocational interest in fine arts, she learned to continue painting using the BrainPort, a SSD that converts visual images into electrical pulses that can be felt as stimulations of the surface of the tongue[32]. I will use Emilie's example to illustrate how the proposed model in Fig. 3 is assumed to work.

Fig. 4: Emilie Gossiaux using BrainPort to paint

The model presupposes that when Emilie was first introduced to the idea of using the BrainPort, the device had the status of distal-attribution for her, but with potentiality to

[32] A brief history of Emilie's case can be accessed on https://www.youtube.com/watch?v=1xYi9oZMVWI

become a proximal-stimulator, and be integrated into her body-schema. As she contemplated the possibility of using the machine, her thoughts and intentions ([b]) acted as the catalyst to make upward local-to-global modifications to her environment ([c]), containing the BrainPort. This caused the device to acquire some psychological significance for her, becoming 'self-relevant' (Beggan, 1992; Belk, 1988) and setting up global-to-local conditions for Emilie to start using it. Further intentions during recurrent loops, enhanced the level of meaningfulness of the BrainPort, generating a stronger will to use it, and with practice she became more proficient at perceiving objects through it. In time, the device gained status of being a bodily proximal stimulator, and this led to its integration into Emilie's body-schema ([a]).

The model in Fig. 3 suggests that Emilie received downward sensorial inputs from the SSD subconsciously, reaching her brain through unnoticed afferent signals. After being processed internally by the brain, efferent signals then made her consciously aware, first of the BrainPort, and then of what it was capturing, and loops of afferent/efferent cycles generated the feeling that the device was being gradually 'owned' by her body.

I hypothesise that efferent signals in Emilie's nervous system ([a-b]) complied with two primary functions. One was to provide her with conscious thoughts, feelings and the necessary follow-up intention and actions to continue using the BrainPort with increased proficiency. The other function was to act as the means to increase psychological significance of the device ([b-c]), leading to more profound integration of it into her body-model.

How did this mechanism alter the state of Emilie's '(observed) self'? And what general conclusions about consciousness and the self, based on the graph in Fig.3, can we extract from her example? When interviewed about the use of the BrainPort Emilie described her feeling as follows: *'…using the brain port I just started drawing more simply; eventually I just felt like I started painting again…'* (Gossiaux, 2015). It appears thus that recurrent b>c>d>a>b loops resulted in a new way of perceiving her artwork, and that this assisted Emilie to appraise her self-image and aptitude differently than before using the device.

The model also shows that the [b] link was responsible for creating the separation between what she observed as being herself and the external world. I would suggest that a similar mechanism is responsible for how our senses generate our field-of-awareness and the feeling of being conscious of oneself, as separated from the outside world.

4.2 Strengths and weaknesses of the proposed model

This thesis is an effort to produce grounds for a novel view of the self and consciousness, and a weakness is that it focussed on inferences from current work in progress without adhering to or following up any particular approach. The robustness of my claims depends, essentially, solely on the strength of logical inferences; if these interpretations are found to be inconsistent or in disagreement with empirical results, or if they violate accepted views, the model could prove to be weak.

Also, due to the limited scope of the thesis it was not possible to explain in sufficient detail how the link between SSDs and the self is supposed to work (this was a concern expressed by one of the examiners of my research proposal). I tried to moderate this problem by offering basic explanations of the observer and the observed selves, and use the real-life example of Emilie Gossiaux to show how the mechanics of the proposed model are meant to operate.

Another shortcoming of this thesis is that by focussing on the mechanics rather than on the functionalities of consciousness, it contributes little to expand on the idea of a minimal-self or on the traditional notions of self-reflection and self-consciousness (which are treated here only as referential). Also, my views are partly inspired on the works of Loomis (1992), which have received little mainstream scientific attention in the philosophical community; I believe, though, that these works have not been seriously challenged and that they encompass the essence of the observed-self as considered in my thesis.

In terms of the strengths of the proposed model I believe that I have produced valid inferences to assert that integration of SSDs into the body-schema does alter the state of the self and that SSDs can thus be used to study consciousness and the observed-self. I also believe that this thesis has provided some grounds for future research to further explore consciousness as composed of two selves, which may lead to understanding how the self is formed and operates in the natural course of evolution.

4.3 Considerations for future research

The model in Fig. 3 is speculative as well as incomplete, and a pressing line of investigation, which I would propose as a doctorate dissertation, is to provide valid information to develop it further and assess its validity. There are two theoretical approaches that may offer important grounds for this purpose: *autopoiesis* and *emergentism*.

Initially proposed as a biological theory to explain living organisms (Maturana & Varela, 1991), autopoiesis is the study of self-sustaining systems capable of creating their own components. As a unity, an autopoietic structure is the product of operations linking its components parts, but its properties cannot be compared with theirs. It operates as a circular system of transactions where the whole cannot function without the parts, or the parts without the whole, thus endowing the unit with 'operational closure'. The theory proved to be adequate as a research tool in domains including sociology, political sciences, Darwinism, psychology and consciousness, and I believe that it can provide valuable information to explain the mechanisms of consciousness as depicted in Fig. 3.

A 'unit of Consciousness' in this schema is represented as a full cycle, from [b] to [a]. Operating in closed-mode, the cycle is hypothesized to be the mechanism that gives us our conscious field of awareness. Operating in open-mode, components within this cycle, such as our body and entities in the environment, structurally couple with each other, all contributing to the consciousness cycle.

The schema presupposes that our observations change the state of what we observe, and like autopoiesis theory, it infers that these changes occur as a result of *transactions* (as opposed to exchange of information) between independent self-creating systems. A transaction is believed to take place when the observation causes perturbations in the 'external state' of the observed structure (not unlike how neurotransmitters activate a neuron); once activated, the structure engages its own internal components, and the resulting new state of the structure then acts as the triggering signal causing perturbations in more global systems. As the cycle completes, it modifies the state of the observer, and the two systems become *structurally coupled* (See Maturana & Varela, 1991 for an explanation of this mechanism in traditional autopoietic terms).

If these inferences are proven valid, it could be argued (also in agreement with autopoietic theory) that conscious decisions are generated internally by *all* systems at both, local (e.g. brain/ body) and global (e.g. society) levels[33].

Another theory that can provide important contributions to explain Fig. 3 schema is *emergentism*. Pioneered by Samuel Alexander (1920) and Conwy Lloyd Morgan (1925) in

[33] I would suggest that decisions are prompted by 'propensities', a property of all autopoietic systems that exhibits as a tendency to favour certain states.

the early 20th Century, emergentism is essentially a method to explain autonomous mental causation from within a physicalist framework. A physical system (e.g. the body) can be described in terms of the intrinsic properties of its lower-level elementary components, which in agglomeration generate the higher-level substantial unity (Kim, 1999). Emerging systems however have supervenient properties that are distinct, additional, unexpected and novel (a bicycle for instance has properties not found in any of its parts).

Emergentism implies that neither lower nor higher level structures can exist without each other and that one causes the other. The mechanism is represented within the circularity of consciousness in Fig. 3, where brain activity [(d]) causes conscious decisions ([b]) leading to environmental changes ([c]) that in turn modify the state of the brain. Through 'preferential neighbourhood relations' (Varela et al., 1974) lower level systems couple structurally, giving emergence to an all-encompassing consciousness and its boundaries, which we perceive as a complete field-of-awareness.

I would suggest that all constituents of consciousness, physical and non-physical, can be studied as independent emergent sub-systems. The state of the brain for instance emerges from the collective behaviour of its local neural assemblies, and the state of the body from the amalgamation of all bodily components; likewise, societal structures such as the legal system, the economy or the government, can be seen as emerging into autopoietic systems due to interaction amongst people and communication (Luhmann, 1986).

The proposed model would possibly describe consciousness as comprising three primary systems: a private and an extended consciousness, and the observer that senses and connects them. Our private consciousness is what the observer-self perceives as the observed-self ([b]), giving us the sense of self-identity and of being 'the person that I am', constant over time. The extended consciousness is what we sense as 'distal attributions' (Siegle & Warren, 2010) and constitute 'otherness'.

Autopoiesis and emergentism theories may offer the platform to explain phenomenology, the 'hard' problem of consciousness, because they may assist in elucidating relations between a physical structure and the metaphysical system that emerges from it. This however, may require a multidisciplinary approach incorporating physics. In his article 'More is different', Phillip Anderson (1972) offers crude attempts to formulate a non-reduction (to specific laws) theory to explain the shift from quantitative to qualitative involving 'phase transition', a concept worth exploring towards this end. A hypothesis to be considered would be that consciousness and the sense of self are emergent bodily phenomena that obey laws that are not necessarily those governing subservient components, and that a conscious experience is in essence *transduction* from the mode of energy of one system (e.g. electro-chemical) into another mode of energy in a different system (e.g. feelings and subjective perception).

4.4 Summary

In this dissertation I attempted to offer suggestions to generate the grounds for a possible theory of the self and the role it plays in the general scheme of consciousness based on research on sensory-substitution-devices (devices that convert information from one sensory modality into a different modality). To overcome the problem of how a reflective minimal-self can generate phenomenal experiences, I proposed the existence of an *observer-self* 'emergent' from the synergetic sum total of brain/body functioning that captures and converts activity of an *observed-self*, arising from afferent/efferent biological signals. The selves are presumed to be essentially processing-sensors, with the former being the ontological 'first-person' and the latter the biological house of our perceptions, personal identity and history.

Adopting an interpretative top-down reflective research method I reviewed literature to deduce how SSDs transform the observed-self when they are incorporated into the user's body-schema in a process that involves four main factors, which together indicate that SSDs generate a novel sense. I identified the neural correlates of these modifications and the four stages of SSDs integration into the body-model and ownership of the body, providing evidence that the observed-self functions to separate attributions to oneself from distal attributions. I then hypothesised a general *schematic representation of consciousness.* The schema shows that the field-of-consciousness is the full spectrum (including unconscious processes) generated by brain/body activity, but that conscious awareness emerges only from efferent neural substrates; effectors mark the boundaries of the observed-self. The model depicts information as travelling *unidirectionally* involving a downward process of afferent global-to-local signals and an upward process of local-to-global efferent signals. SSDs modify the observed-self because they act as sensory inputs as part of the downward process.

Finally, I suggested that a pressing need for further research should include exploring how theories such as autopoiesis and emergentism could assist explaining and completing the consciousness cycle hypothesised in Fig. 3.

Glossary

Afferent signals Unconscious Global-to-local signals that travel unidirectionally from a system to the system's subcomponents. Afferent signals from the environment for example, travel downwards affecting the state of the body's periphery, in turn influencing the brain and subsequently brain constituents.

Afferent-efferent loops Loops of signals travelling unidirectionally from global-to-local and local-to-global systems.

Agency of the body The awareness that it is 'me', my body, who is generating an action.

Attributions to oneself Perceptions that we attribute as belonging to our body, personal identity, personal attributes and personal history. We perceive attributions to oneself as manifestations of the 'self'.

Autopoiesis Self-sustaining systems capable of creating their own components. The properties of an autopoietic system cannot be compared with those of its component parts. The system as a whole cannot function without its parts, or the parts without the whole, thus endowing the unit with 'operational closure'. Consciousness is presupposed to function as an autopoietic system.

Body-model A pre-existing state of bodily affairs stable over time that manifests as conscious perceptions of one's body and oneself, including our personal attributes and self-identity. Different to the body-schema (see body schema), which is an unconscious process, the body model allows the person to be aware of bodily changes, imposing restrictions to these changes in order to preserve stability of bodily functioning over time.

Body-schema A subconscious process of how the body is represented in movement and space. The body-schema undergoes transformations by incoming sensory impulses, but we only become aware of these changes through conscious perceptions of the body-model (see body model). Artefacts external to the body that cause modifications to the body-schema can modify the body-model and create the feeling that the artefact is incorporated into the body – thus altering the perceived boundaries of the body.

Closed-system An 'autopoietic' (see autopoiesis) network of processes capable of maintaining itself by generating and managing its own component parts. In closed-mode a system is capable of detecting and processing information from its own elements. Consciousness is presumed to be an autopoietic system; operating in closed-mode, it generates our field of awareness containing all our perceptions, feelings and cognitions. The body is also thought to be an autopoietic system. Since closed-systems are self-contained and causally circular, we are not consciously aware of bodily processes; we become aware of them when they give emergence to a different and unique system with its own properties (referred here as 'the observer-self). We sense this unique emerging system as being 'me', the person that 'I am'.

Distal attribution Distal attributions are the part of our field of awareness that we perceive as 'otherness' or 'not me'. They are felt and recognised as foreign to the body and as not belonging to the self. With habitual use a distal object (prosthesis for example) may change the distal status and become an 'attribution to oneself' (see attributions to oneself), being perceived as part of the self.

Efferent signals Conscious Local-to-global signals that travel unidirectionally from internal components of a system to the system encompassing them. Efferent signals from brain substrates for example, influence the state of the brain as a whole, which in turn alters the state of the body and subsequently the environment.

Emergence The term 'Emergence' implies that a structure is created from more than the sum of its components. The properties of the emergent structure are not like and cannot be compared with the properties of any of its component parts. An emergent system is independent and 'autopoietic' (see autopoiesis). This thesis presupposes that consciousness is a system emerging from the body's component parts, whose properties are unlike those of the body or brain.

Observed-self The observed-self is a term introduced in this thesis to describe the part of the conscious field of awareness that we perceive as being oneself. It is what we sense as being 'me', one's own body, personal identity and subjective experiences, separated from what is 'not me'. The observed-self exhibits as attributions to oneself (see Attributions to oneself); it is proposed to emerge from afferent/efferent nerve signals travelling between brain and periphery effectors. Extensions outside the boundaries of the body effectors do not belong to the observed-self; they are signals that travel outside the boundaries of the body, and represent the non-self or ‘otherness’.

The observed-self is an unconscious process (essentially comparable to the body-schema); it becomes conscious when the ‘observer-self’ (see observer-self) senses it. Sensing these unconscious signals by the observer-self gives us our field of awareness and the perception of separation between us and the world. Operating in a ‘local’ closed-mode (see closed-system), the observed-self is presumed to process information received through bodily sensoria. In open-mode (see open-system) it exhibits a ‘field-like’ aspect, releasing processed information that is then sensed phenomenologically by the ‘observer-self’.

Observer-self The observer-self is a term proposed here to describe ‘the observer or experiencer’, the elusive ‘I’ in the first-person ontology. Hypothesized to emerge from the synergetic sum total of brain/body functioning, transforming biological signals into phenomenological experiences, it exhibits in our field of awareness and in our ability to separate the perception of ‘me’ from the perception of ‘otherness’. Since the amalgamation of bodily parts is achieved with synergy, the observer-self emerges as a unique and independent system with its own properties, allowing it to convert subconscious electro-chemical bodily energy into phenomenological perceptions. This thesis presupposes that in ‘closed-mode’ (see closed-system) the observer-self perceives and manages its sub-components (internal and external to the body) through our conscious thoughts, feelings and actions. In ‘open-mode’ (see open-system) it releases this processed information subconsciously influencing other systems.

The observer-self is presumed to function as a sensor and thus to only be able to perceive what it senses, and not itself. Therefore, it senses internal components of the body, including the brain, in essentially the same manner as it senses environmental stimuli.

Open-system A system operates in open-mode when it communicates with other systems, as opposed to communicating with its own components, thus exhibiting ‘field-like’ characteristics. After having processed information internally in closed-mode (see closed-system), an entity influences other systems by releasing information in open-mode; resulting on the two systems becoming structurally coupled. It is hypothesized that consciousness operates as a system; after receiving information subconsciously from another system, we process this information in closed-mode (with our conscious thoughts and feelings) before releasing the result of this process in open-mode. The concept of open-system implies that although systems, such as a person, are activated by signals incoming from other agents, these signals do not have direct access to the internal components of the receiving structure (somewhat like neurotransmitters do not directly affect the components of the receiving neuron), thus rendering all systems autopoietic (see autopoiesis).

Proximal sense The term 'proximal sense' refers to bodily senses that are fully incorporated as parts of the body and thus felt as owned by the body. A typical example would be the eyes. A non-proximal sense conversely, is not felt as belonging to the body; an example could be the fork and knife we use to eat. A foreign object, prosthesis for example, may be initially non-proximal but with practice and habituation it can become a proximal sense.

Proximal stimulator Proximal stimulator is a term essentially synonymous to proximal sense. Stimulators are whatever stimulate proximal senses (see proximal sense); they are by definition entities external to the body. When a stimulator becomes incorporated as part of the body, it gains the status of proximal stimulator, functioning as a proximal sense. A cane used by a blind person is an example of a stimulator not belonging to the body; with time and practice however, the person may feel the tip of the cane as essentially not different to the tip of a finger, signalling the change of the cane status from being a distal stimulator to a proximal stimulator.

References

Alexander, S. (1920). *Space, time, and deity: the Gifford lectures at Glasgow, 1916-1918* (No. 2). London: Macmillan.

Anderson, P. W. (1972). More is different. *Science*, *177*(4047), 393-396.

Auvray, M., Hanneton, S., Lenay, C., & O'Regan, J. K. (2005). There is something out there: distal attribution in sensory substitution, twenty years later. *Journal of integrative neuroscience*, *4*(04), 505-521.

Auvray, M., & Myin, E. (2009). Perception with compensatory devices: from sensory substitution to sensorimotor extension. *Cognitive Science*, *33*(6), 1036-1058.

Baars, B. J. (1997). Neural hypotheses derived from Global Workspace Theory: Elements of a cognitive neuroscience of consciousness. *Cognitive Studies*, *4*(3), 5-14.

Baars, B. J. (2003). The global brainweb: An update on global workspace theory. *Science and Consciousness Review*, *2*.

Bach-y-Rita, P. (1972). *Brain mechanisms in sensory substitution.* Academic Press.

Bach-y-Rita, P., Collins, C. C., Saunders, F. A., White, B., & Scadden, L. (1969). Vision substitution by tactile image projection. *Nature*, *221*(5184), 963.

Bach-y-Rita, P., & Kercel, S. W. (2003). Sensory substitution and the human–machine interface. *Trends in cognitive sciences*, *7*(12), 541-546.

Beggan, J. K. (1992). On the social nature of nonsocial perception: The mere ownership effect. *Journal of personality and social psychology*, *62*(2), 229.

Belk, R. W. (1988). Possessions and the extended self. *Journal of consumer research*, *15*(2), 139-168.

Bermúdez, J. L. (1995). Ecological perception and the notion of a nonconceptual point of view. *The body and the self, 153474.*

Blanke, O., Ortigue, S., Landis, T., & Seeck, M. (2002). Neuropsychology: stimulating illusory own-body perceptions. *Nature*, *419*(6904), 269.

Bottini, G., Bisiach, E., Sterzi, R., & Vallarc, G. (2002). Feeling touches in someone else's hand. *Neuroreport*, *13*(2), 249-252.

Botvinick, M., & Cohen, J. (1998). Rubber hands 'feel' touch that eyes see. *Nature*, *391*(6669), 756.

Campbell, D. T. (1966). Pattern matching as an essential in distal knowing. *The psychology of Egon Brunswik*, 81-106.

Carlson, T. A., Alvarez, G., Wu, D. A., & Verstraten, F. A. (2010). Rapid assimilation of external objects into the body schema. *Psychological Science*, *21*(7), 1000-1005.

Chalmers, D. J. (2000). What is a neural correlate of consciousness. *Neural correlates of consciousness: Empirical and conceptual questions*, 17-40.

Chaminade, T., Meltzoff, A. N., & Decety, J. (2005). An fMRI study of imitation: action representation and body schema. *Neuropsychologia*, *43*(1), 115-127.

Clark, A. (2008). *Supersizing the mind: Embodiment, action, and cognitive extension*. Oxord University Press, USA.

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, *58*(1), 7-19.

Collins, C. C., & Bach-y-Rita, P. (1973). Transmission of pictorial information through the skin. *Advances in Biological and Medical Physics*, *14*, 285–315.

Corning, P. A. (1983). *The synergism hypothesis: a theory of progressive evolution*. McGraw-Hill Companies.

Costantini, M., & Haggard, P. (2007). The rubber hand illusion: sensitivity and reference frame for body ownership. *Consciousness and cognition*, *16*(2), 229-240.

Damasio, A. R. (1990). Synchronous activation in multiple cortical regions: A mechanism for recall. *Seminars in Neuroscience*, *2*, 287–296.

de Preester, H. (2011). Technology and the body: The (im) possibilities of reembodiment. *Foundations of science*, *16*(2-3), 119-137.

de Preester, H., & Tsakiris, M. (2009). Body-extension versus body-incorporation: Is there a need for a body-model?. *Phenomenology and the Cognitive Sciences*, *8*(3), 307-319.

de Volder, A. G., Catalan-Ahumada, M., Robert, A., Bol, A., Labar, D., Coppens, A., & Veraart, C. (1999). Changes in occipital cortex activity in early blind humans using a sensory substitution device. *Brain research*, *826*(1), 128-134.

Deroy, O., & Auvray, M. (2014). A crossmodal perspective on sensory substitution. *Perception and its Modalities, 327-349.*

Deroy, O., & Auvray, M. (2015). Beyond vision: The vertical integration of sensory substitution devices. In D. Stokes, M. Matthen & S. Biggs (eds.), *Perception and Its Modalities*. Oxford University Press (2015).

Foucault, M. (1984). *The Foucault Reader*. *New York* (Vol. 1). Pantheon.

Gallagher, S. (2000). Self reference and schizophrenia. *Exploring the self*, 203-239.

Goldman, A. (1970). *A Theory of Action*. New York: Princeton University Press.

Gossiaux, E. (2015). *Interview with Emilie Gossiaux*. Accessed on https://www.youtube.com/watch?v=1xYi9oZMVWI

Gould, S. J., & Vrba, E. S. (1982). Exaptation—a missing term in the science of form. *Paleobiology*, *8*(1), 4-15.

Graziano, M. S., & Botvinick, M. M. (2002). How the brain represents the body: insights from neurophysiology and psychology. *Common mechanisms in perception and action: Attention and performance XIX*, 136-157.

Grice, H. P. (1962). Some remarks about the senses. *Analytical philosophy (first series)*. 248-268. Ed. R. J. Butler.

Haggard, P., & Wolpert, D. M. (2005). Disorders of Body Scheme. *Higher-Order Motor Disorders*, 1–7.

Harman, G. (1990). The intrinsic quality of experience. *Philosophical perspectives*, *4*, 31-52.

Heersmink, R. (2017). The narrative self, distributed memory, and evocative objects.

Philosophical Studies, 1-21.

Husserl, E. (1991). *On the Phenomenology of the Consciousness of Internal Time (1893–1917).* Translated by J. B. Brough. Sprnger+Science and Business Media.

Huttenlocher, P. R. (2009). *Neural plasticity: The effects of environment on the development of the cerebral cortex*. Harvard University Press.

Ihde, D. (1979). Technics and praxis. Dordrecht: D. *Boston: Reidel*.

Ihde, D. (2002). *Bodies in technology* (Vol. 5). University of Minnesota Press.

Inoue, K., Kawashima, R., Sugiura, M., Ogawa, A., Schormann, T., Zilles, K., & Fukuda, H. (2001). Activation in the ipsilateral posterior parietal cortex during tool use: a PET study. *Neuroimage*, *14*(6), 1469-1475.

Johnson, S. H. (2000). Thinking ahead: the case for motor imagery in prospective judgements of prehension. *Cognition*, *74*(1), 33-70.

Kelso, J. (1995). *Dynamic Patterns: The Self–Organization of Brain and Behavior*. MIT Press.

Kim, J. (1999). Making sense of emergence. *Philosophical studies*, *95*(1-2), 3-36.

Kiverstein, J., & Farina, M. (2012). Do sensory substitution devices extend the conscious mind?. *Consciousness in Interaction: The role of the natural and social context in shaping consciousness*, *86*, 19-40.

Kiverstein, J., Farina, M., & Clark, A. (2015). Sensory Substitution. In M. Mohan (Ed) *The Oxford Handbook of Philosophy of Perception.* New York: Oxford University Press.

Kluver, H. (1967). *Mscal and the Mechanisms of Hallucination*. Chicago: The University of Chicago.

Kriegel, U. (2004). Consciousness and self-consciousness. *The Monist*, *87*(2), 182-205.

Lenay, C., Gapenne, O., Hanneton, S., Marque, C., & Genouëlle, C. (2003). Sensory Substitution : Limits and Perspectives. *Touching for Knowing Cognitive Psychology of Haptic Manual Perception*, *19*, 275–292.

Leopold, D. A., & Logothetis, N. K. (1999). Multistable phenomena: changing views in perception. *Trends in cognitive sciences*, *3*(7), 254-264.

Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and brain sciences*, *8*(4), 529-539.

Longo, M. R., Schüür, F., Kammers, M. P., Tsakiris, M., & Haggard, P. (2008). What is embodiment? A psychometric approach. *Cognition*, *107*(3), 978-998.

Loomis, J. M. (1992). Distal attribution and presence. *Presence: Teleoperators & Virtual Environments*, *1*(1), 113-119.

Luhmann, N. (1986). The autopoiesis of social systems. *Sociocybernetic paradoxes*, *6*(2), 172-192.

Maravita, A., & Iriki, A. (2004). Tools for the body (schema). *Trends in cognitive sciences*, *8*(2), 79-86.

Matthen, M. (2005). *Seeing, doing, and knowing: A philosophical theory of sense perception*. Oxford University Press on Demand.

Maturana, H. R., & Varela, F. J. (1991). *Autopoiesis and cognition: The realization of the living* (Vol. 42). Springer Science & Business Media.

Mead, G. H. (1962). *Mind, Self and Society: From the Standpoint of a Social Behaviorist*. Chicago: University of Chicago Press.

Merleau-Ponty, M. (1996). *Phenomenology of perception*. Motilal Banarsidass. Delhi.

Moore, D. (2015). Supervenient Emergentism and Mereological Emergentism, *Axiomathes, 25*(4), 457-477.

Morgan, C. L. (1925). Emergent evolution. *Mind*, *34*(133), 70-74.

Nagel, S. K., Carl, C., Kringe, T., Märtin, R., & König, P. (2005). Beyond sensory substitution—learning the sixth sense. *Journal of neural engineering*, *2*(4), R13.

Nagel, T. (1974). What is it like to be a bat?. *The philosophical review*, *83*(4), 435-450.

Neisser, U. (Ed.). (2006). *The perceived self: ecological and interpersonal sources of self knowledge* (Vol. 5). Cambridge University Press.

Northoff, G., & Bermpohl, F. (2004). Cortical midline structures and the self. *Trends in cognitive sciences, 8*(3), 102-107.

Northoff, G., Heinzel, A., De Greck, M., Bermpohl, F., Dobrowolny, H., & Panksepp, J. (2006). Self-referential processing in our brain—a meta-analysis of imaging studies on the self. *Neuroimage*, *31*(1), 440-457.

O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences, 24*(5), 939–973.

Palermos, S. O. (2014). Loops, constitution, and cognitive extension. *Cognitive systems research*, *27*, 25-41.

Penfield, W., & Jasper, H. (1954). *Epilepsy and the functional anatomy of the human brain*. Little, Brown & Co. Oxford, England.

Port, R. F., & van Gelder, T. (Eds.). (1995). *Mind as motion: Explorations in the dynamics of cognition*. MIT Press.

Riggio, L., de Gonzaga Gawryszewski, L., & Umilta, C. (1986). What is crossed in crossed-hand effects?. *Acta Psychologica*, *62*(1), 89-100.

Royce, J. (1898). *Studies of Good and Evil: A Series of Essays Upon Problems of Philosophy and Life*. D. Appleton.

Sartre, J. P. (1956). *Being and Nothingness*. Trans. Hazel E. Barnes (New York: Philosophical Library, 1956), 330.

Searle, J. R. (2005). The self as a problem in philosophy and neurobiology. *The lost self: Pathologies of the brain and identity*, 7-19.

Siegle, J. H., & Warren, W. H. (2010). Distal attribution and distance perception in sensory substitution. *Perception*, *39*(2), 208-223.

Simmel, M. L. (1966). Developmental aspects of the body scheme. *Child Development*, *37*(1), 83-95.

Skarda, C. A., & Freeman, W. J. (1987). How brains make chaos in order to make sense of the world. *Behavioral and brain sciences*, *10*(2), 161-173.

Strawson, G. (1999). The self. *Models of the Self*, 1-24.

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., & Varela, F. J. (2001). Radical embodiment: neural dynamics and consciousness. *Trends in cognitive sciences*, *5*(10), 418-425.

Tsakiris, M., Costantini, M., & Haggard, P. (2008). The role of the right temporo-parietal junction in maintaining a coherent sense of one's body. *Neuropsychologia*, *46*(12), 3014-3018.

Tsakiris, M., & Haggard, P. (2005). The rubber hand illusion revisited: visuotactile integration and self-attribution. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(1), 80-91.

Varela, F. G., Maturana, H. R., & Uribe, R. (1974). Autopoiesis: the organization of living systems, its characterization and a model. *Biosystems*, *5*(4), 187-196.

van Gelder, T., & Clark, A. (1998). Being There: Putting Brain, Body and World Together Again. *The Philosophical Review*, *107*(4), 647.

von Holst, E. (1954). Relations between the central Nervous System and the peripheral organs. *The British Journal of Animal Behaviour*, *2*(3), 89–94.

Ward, J., & Meijer, P. (2010). Visual experiences in the blind induced by an auditory sensory substitution device. *Consciousness and Cognition*, *19*(1), 492–500.

Wolpert, D. M., Goodbody, S. J., & Husain, M. (1998). Maintaining internal representations: the role of the human superior parietal lobe. *Nature neuroscience*, *1*(6), 529-533.

Wyatt, J., & Rizzo, J. (1996). Ocular implants for the blind. *IEEE spectrum*, *33*(5), 47-53.

Zahavi, D. (2014). Consciousness, Self-Consciousness, and Selfhood. In *Self and Other Exploring Subjectivity, Empathy, and Shame*, 10–24. Oxford University Press.