NEUROPHILOSOPHY OF SELF-CONSCIOUSNESS:
Clarifications for the Connection between Minimal and Reflective Form

Heidi Haanila

Neurophilosophy
of Self-consciousness:

Clarifications for the Connection
between Minimal and Reflective Form

Heidi Haanila

UNIVERSITY OF TURKU

Written by
MSocSc Heidi Haanila
University of Turku,
Faculty of Social Sciences,
Department of Philosophy, Contemporary History and Political Science,
Philosophy
Doctoral Programme of Social and Behavioural Sciences

Supervised by
Docent Valtteri Arstila
University of Turku
Professor Olli Koistinen
University of Turku
Professor Antti Revonsuo
University of Turku

Reviewed by
Doctor Julian Kiverstein
University of Amsterdam
Doctor Miira Tuominen
Stockholm University

Opponent
Doctor Julian Kiverstein
University of Amsterdam

Chairperson (custos)
Docent Valtteri Arstila
University of Turku

Abstract

Selfhood has a central role in people's lives. The self is present as the subject of experience, thinker of thoughts, and agent of action. This doctoral study elaborates on the concept on self-consciousness by examining the ways through individuals experience self. Self-consciousness is typically considered to have two forms: minimal and reflective. Minimal self-consciousness refers to the self as a subjective perspective or the subjectivity of consciousness that is present in all conscious mental states. By contrast, reflective self-consciousness refers to the self as an object of thoughts and experiences. The key finding of this study is that minimal and reflective self-consciousness are closely connected and that the whole of self-consciousness can be expounded only by acknowledging their dynamic interrelations. Until now, these interrelations have been understudied, but this study provides an approach to clarify their manifestation and significance. The research methodology of this study is neurophilosophy, which aims at multidisciplinary cooperation between philosophers and empirical scientists.

This dissertation consists of two parts. The first part lays conceptual foundations by providing detailed definitions of the concepts of minimal and reflective self-consciousness, examining their connections, and investigating their neural mechanisms. The analysis strengthens the phenomenological conception of minimal self-consciousness as a constitutive feature of experience that involves experiential, embodied, and typically affective features. In terms of reflective self-consciousness, the analysis highlights different modes of self-conscious thinking. The two forms of self-consciousness rely on different neural processes; however, the processes also co-occur and are intertwined.

The second part of this dissertation applies the conceptual model developed in the first part to concrete cases of self-consciousness. The analysis of these cases reveals that the conceptual model is not only theoretically interesting but also useful in practice for understanding instances of self-consciousness. The considered cases are altered states of consciousness in Cotard syndrome, depersonalization, and meditation, which are characterized by extensive changes in the dynamics of self-consciousness. These states elicit the features of typical self-consciousness and provide further support to the idea of the mutual influence of minimal and reflective self-consciousness on each other. In addition, this study indicates that the connections in the structures of self-consciousness are related to mental well-being.

KEYWORDS: self, self-consciousness, minimal self-consciousness, reflective self-consciousness, philosophy of mind, neurophilosophy, Cotard syndrome, depersonalization, meditation

Tiivistelmä

Minuus on arkikokemuksemme olennaisimpia piirteitä. Minä on elämässä läsnä kokien, ajatellen ja toimintaa ohjaten. Tämä väitöskirjatutkimus edistää minuuden monitahoisuuden ymmärtämistä keskittyen tarkentamaan käsitystä itsetietoisuudesta, eli siitä miten koemme minuuden. Itsetietoisuus jaetaan yleisesti kahteen muotoon: minimaaliseen ja reflektiiviseen. Minimaalinen itsetietoisuus viittaa minään kokemuksen subjektina tai subjektiivisuutena, joka on läsnä kaikissa kokemuksissa. Reflektiivinen itsetietoisuus puolestaan viittaa kykyyn ottaa minä ajattelun kohteeksi. Väitöskirjani keskeinen tutkimustulos on, että minimaalinen ja reflektiivinen itsetietoisuus muodostavat tiiviisti linkittyneen kokonaisuuden, jota voidaan ymmärtää vain huomioimalla niiden välinen molemminpuolinen vuorovaikutus. Tutkimukseni tarjoaa uusia tapoja selventää tuon vuorovaikutuksen ilmenemistä ja merkittävyyttä. Tutkimusmetodina toimii neurofilosofia, joka hyödyntää sekä perinteistä mielenfilosofista analyysia että empiiristen tieteiden tuottamia tietoja.

Väitöskirja koostuu kahdesta osasta. Ensimmäisessä osassa määrittelen minimaalisen ja reflektiivisen itsetietoisuuden käsitteet ja analysoin niiden välisiä yhteyksiä. Tutkimukseni vahvistaa käsitystä, jonka mukaan minimaalinen itsetietoisuus on tietoisuuden rakenteellinen ainesosa, joka sisältää kokemuksellisia, kehollisia ja yleensä affektiivisia piirteitä. Lisäksi korostan reflektiivisen itsetietoisuuden sisältävän useita eri tekijöitä, jotka vaikuttavat itseen kohdistuviin ajatuksiin. Erittelen kummankin itsetietoisuuden muodon mahdollistavia neuraalisia prosesseja, mutta samalla nostan esiin, kuinka nämä prosessit myös esiintyvät yhdessä ja vaikuttavat toisiinsa.

Väitöskirjan toisessa osassa sovellan ensimmäisessä osassa kehittämääni käsitteellistä mallia erilaisiin konkreettisiin kokemuksiin itsetietoisuudesta. Näitä tapauksia tarkastelemalla osoitan, että tuo malli ei ole ainoastaan teoreettisesti kiinnostava, vaan tarjoaa käytännön hyötyä erilaisten itsetietoisuuden tapausten ymmärtämiseen. Analysoin muuntuneita tajunnantiloja Cotardin syndroomassa, depersonalisaatiossa sekä meditaatiossa, joille tunnusomaista on juuri minäkokemuksen laaja-alainen muutos. Nämä muutokset tuovat esiin itsetietoisuuden normaalin rakenteen ja antavat lisätukea esittämälleni mallille minimaalisen ja reflektiivisen itsetietoisuuden välisestä vuorovaikutuksesta. Lisäksi analyysi osoittaa itsetietoisuuden osatekijöiden välisen tasapainon olevan yhteydessä psyykkiseen hyvinvointiin.

ASIASANAT: minuus, itsetietoisuus, minimaalinen itsetietoisuus, reflektiivinen itsetietoisuus, mielenfilosofia, neurofilosofia, Cotardin syndrooma, depersonalisaatio, meditaatio

Acknowledgements

I have always been fascinated by the richness of human consciousness, and thus, it feels a privilege to have had the opportunity to conduct research on the philosophy of mind. Writing this dissertation has been a great journey into the philosophy of mind, into academia, and into my self. I have had the delight of having wonderful company along the journey, and I warmly thank everybody who has been involved.

First of all, I want to thank my supervisors, Docent Valtteri Arstila, Professor Olli Koistinen, and Professor Antti Revonsuo. It has been an honor to work with you. A common thanks go to all of you for your encouragement in finding and determining my own way as a philosopher.

Valtteri has provided me with irreplaceable support. He has commented on numerous versions of my texts and plans, and broadened my knowledge of philosophy, writing, argumentation, and practicalities in academia. He has also been my role model as a neurophilosopher, with his expertise, skills, and patience for building connections between different disciplines and bringing philosophy to new places. In addition, he is always kind to people and I have been lucky to become his friend. I am also very happy for the opportunity to continue to work with him. A big thank to you Valtteri—you have been a super supervisor!

Olli has presented the viewpoint of a more traditional philosophy of mind, history of philosophy, and philosophy of action and will, which already caught my interest in the first stages of my university studies. Olli has a great character and I want to thank him for setting a fine example as a philosopher—through asking questions, emphasizing the importance of definitions, and being willing to understand as well as to share his understanding with others.

Antti enabled a genuine multidisciplinary framework for this dissertation through his expertise in psychology. As a master student, I was fascinated by his Consciousness and Brain course, and from then on I dreamed of studying consciousness more. I am impressed by the way he has mastered the wide field of the science

of consciousness. I am thankful to Antti for this multidisciplinary model and insight into my work, and for the opportunity to be a part of his research group.

I also thank Doctor Julian Kiverstein and Doctor Miira Tuominen for their valuable comments on the manuscript of this dissertation, which will help me in my future work. I am deeply grateful to Julian for kindly serving as a discussant in the doctoral defence as well as for his earlier ideas related to my work, which assisted me considerably in developing the argumentation in the dissertation.

As I was writing this dissertation, colleagues and friends provided indispensable support. First, I would like to mention three people who have played especially large roles: Anne Salminen, Nils Sandman, and Sanna Mattila. Particularly at the beginning of my doctoral studies, Anne and Nils were the most involved. They commented on the first versions of my texts and presentations, and we shared countless discussions, insights, and chess games.

At first, Nils set me an excellent academic example and taught me about new fields of science and the fun of academia. In the long run, I grew to respect him for his versatility and his interest in using his expertise for good. Nils, thank you for making my life richer with your sharp and sympathetic thinking and through sharing your ideas, travel reports, and art.

Anne was already my friend when we were master students and later, she was my co-author. I really enjoyed all of our reflections on the wonders of life and capacities of the mind. In particular, I remember our coffee breaks in the park when the sun was shining and how we contemplated the dimensions of self. Anne, thank you for always being kind and supportive and demonstrating courage in choosing your own path.

Sanna played a crucial role in the final stages of writing this dissertation. She helped me through the pandemic period during which we shared the same life situation, and it was a great advantage to be able to share the joy and challenges of combining parenthood and finishing a PhD. In addition to simply cheering me up, Sanna is admirably precise in a dialectical outline of arguments,

and she has reminded me of the importance of metaphysics. Thank you Sanna for keeping my spirits up.

Warm thanks go to Pii Telakivi, who I was lucky to meet at a summer school in Copenhagen back in 2014. She has acquainted me with many interesting people as well as with the ideas of enactivism. We have worked together and had fun. Thank you Pii for your friendship and for extending my mind.

I have had the delight of working in the philosophy unit at the University of Turku, which is characterized by a warm and welcoming atmosphere. I want to thank everyone I have met there over the years! In particular, I want to thank Hanna-Mari Salonen, who has always been an invaluable help with all manner of practical things. I am grateful to Polaris Koi and Susanne Uusitalo for introducing me to new things.

In addition, I have had the privilege of being a member of the Consciousness Research Group. The group has offered me multidisciplinary know-how that I could not have learned from books, and it has been great to get acquainted with all of the people there, who are highly competent and friendly. Thank you all, in particular Roosa Kallionpää, Katja Valli, Jarno Tuominen, Pilleriin Sikka, and Niina Salminen-Vaparanta.

Warm thanks go to all of my friends for support in sharing PhD student life. Thank you Visa Helenius for your unwavering endorsement of humanism. Thank you Mika Hämäläinen for kindly counseling me at the beginning of my PhD. Thank you Katariina Kangassalo for all of the hot chocolate and sympathy. Big thanks go to Suvi Holm and Tiia Sudenkaarne. Thanks also go to my former officemates Sini Nikunen and Oki Lindgrén for their awesome peer support and Friday beers at Proffan Kellari.

In addition, I am grateful to Jelena Issajeva, Teemu Mäntylä, Elisa Aaltola, Eerik Lagerspetz, Hemmo Laiho, Juha Räikkä, Helena Siipi, Matti Saarni, and Jani Sinokki. Thanks also go to my more recent colleagues Jenna Aarnio, Mireille Isaro, Laura Puumala, and Mikko Puumala for being young, bright, and full of philosophical interest.

Furthermore, the small network of empirically oriented philosophers of mind in Finland has been important to me. In

My profound gratitude goes to Evelyn, who has provided my self with whole new dimensions, giving me joy, energy, and perspectives that I could not had imagined.

Table of Contents

PART I

PART II

Abbreviations

1P	first-person
1PP	first-person perspective
3P	third-person
ASC	altered state of consciousness
CS	Cotard syndrome
DPD	depersonalization/derealization disorder
MSC	minimal self-consciousness
RSC	reflective self-consciousness
SRP	self-related processes
SSP	self-specific processing
AIC	anterior insular cortex
CMS	cortical midline structures
DMN	default mode network
PFC	prefrontal cortex
MPFC	medial prefrontal cortex
DMPFC	dorsomedial prefrontal cortex
DLPFC	dorsolateral prefrontal cortex
MMPFC	middle medial prefrontal cortex
VMPFC	ventromedial prefrontal cortex
VLPFC	ventrolateral prefrontal cortex

PART I

Prologue: Sofia's walk in the forest and her self-reflection

Sofia walks deeper into a forest. It is a bright, sunny day in July, and Sofia is looking for places where blueberries grow. She knows approximately where they are but wants to be sure so she is prepared for the best berry-picking season. At the same time, she is enjoying the walk through the forest. She has not been in the countryside for a while, and now the forest looks enchanting; she is overwhelmed by all the colors, sounds, and sensations. There seem to be hundreds of shades of green in the trees and on the ground; dark spruce bark and black and white birch bark are following each other as she enters deeper into the forest. When she looks around, she can see the various shapes and bright colors of blooming summer flowers: yellow, white, purple and blue. She feels the warmth of sunshine on her skin and the softness of moss under her feet. She hears the orchestra of singing birds, chirping of crickets and whistling of a light breeze.

When Sofia comes near the blueberry places, she starts to look around more closely. She has been told where the best places are, and as a favor to her grandmother, she wants to pick the berries by herself this year. Sofia used to visit the forest as a child with her grandmother, but now the elderly woman has become seriously ill and Sofia is dedicated to doing her best to help her. However, the trees have grown, paths have changed a little bit, and new bushes hide rocks and old stumps that fascinated her as a child. But finally Sofia comes to the bank of a small forest pond; she knows this

place. The dark water sparkles in the sun. Sofia reaches out her hand to feel the temperature of water, and she is happy to find out that it is refreshingly cool. She sprinkles her face with the water since she is feeling a little tired of the walk on a hot day. After the refreshment Sofia leans toward the pool to see a mirror reflection of her face on the surface of the water. She still remembers that it was great fun as a child.

Now the reflection in the water is different from the one she remembers. All of a sudden, Sofia feels herself somehow unreal, as if it was not herself at all in the reflection but someone else's face instead. And further, as if she was not there by the pond at all: although she can clearly see the forest landscape in all its detail, it feels almost as if she could see also the figure of a young woman leaning towards a pond but the woman is someone else. To Sofia, it does not feel like really being there: although she still can see and hear everything, it feels as if she is outside of herself, only watching the scene like a movie. But the episode is very brief, and soon Sofia realizes that it is her own mirror image looking back in the reflection. And, she thinks, maybe she has an even fuller impression of herself now than as the child who did not contemplate and who knew so little. Now, the roundness of her face and other childish features have vanished and slight wrinkles appear instead. However, the eyes are still bright and the same—she is the same Sofia still.

Suddenly, maybe incited by the short episode of anomalous feelings of not being herself, Sofia starts self-reflection and contemplates the similarities and differences between herself as a child and now. She views her whole life as a narrative, thinking about the episodes that have affected her and about the decisions she has made. She thinks about her experiences, values, and thoughts. She deliberates whether she has been the same person through all of her life and what kind of person she would like to be in the future. If she comes back to the blueberry fields and the pond one day as a grandmother herself, will she have accomplished what she wished for in life? Can she be proud of the decisions she has made? Sofia smiles, thinking that life may not have been what she thought as a young child, but it has turned out to be much

more complex; still, she can tell a coherent life story and be content with the course of her actions. And after all, contemplating might not even matter that much but instead she can just enjoy this bright sunny day, the tranquility of nature, of being right there and then. Feeling equanimity with herself again, Sofia takes a deep breath and directs her eyes to the blueberries. Concentrating on discerning the berries in the lush forest, Sofia forgets about her self-reflection. But what kind of journey into her self-consciousness did Sofia actually take?

1. Introduction

1.1. Selfhood and self-consciousness

Selfhood is deeply connected to our subjective experience and 'I' is the protagonist of everyone's life. 'I' or a self is present in every moment we live through, experiencing joy and sorrow, developing and making decisions. Yet we get easily puzzled when we ask ourselves who this 'I' actually is: Who is the thinker of my thoughts and the dreamer of my dreams? Who directs my life and is responsible for my deeds? Although 'I' has an essential role in our everyday life, it is philosophically very challenging to formulate answers to these questions, and there are notable conceptual confusions involved in the research concerning selfhood.

Basically, 'self' or 'selfhood' is an umbrella term that includes different features of self and self-consciousness.[1] Because of the many facets of self, numerous concepts are needed in grasping it, and it is crucial in each discussion to be able to map which facet is the subject of interest. The conceptual confusions related to the self arise roughly for (at least) two reasons: the complexity of the phenomenon and the multitude of accounts of self. The complexity of self can be seen in that selfhood is involved in multiple dimensions of action and thinking. The capacity for self-consciousness and self-deliberation is a key factor in purposeful human action. Yet, selfhood is present already at the basic level of experience: self is present in everything one does. Although I am not concentrating on myself all the time, the self is intertwined in my every thought and experience. Whatever I think about, I cannot say that those thoughts are completely independent or detached from my self; even though my thoughts are not focused on myself, they occur in my stream of consciousness, and thus they are related to me essentially. One cannot get rid of herself, but the self is pertinent in mind and action, and thus, selfhood has a central role in our life. Second, because of the many facets of selfhood, different theories of self can be restricted to distinct

---

[1] For an overview of conceptions of self, see, e.g., Gallagher, 2011; Siderits et al., 2010.

facets. As different theories operate on their own conceptions and frameworks, it can be tricky to see whether the conceptions are conflicting or complementary. Even when they are complementary, it remains tricky to integrate the conceptions together in the study of self. In this dissertation, I elaborate the conception of the complex structure of selfhood and specify concepts that can be applied across several theories of self.

This introduction chapter consists of three parts. The first part (1.1.) introduces the topic of selfhood by drawing a very general picture of the study of self in philosophy and empirical sciences. In order to bring various viewpoints together, I will present Gallagher's pattern theory of self as a common ground for structuring the discussion of self. The second part (1.2) poses the research question and gives an overview of the contents. The research question concerns the connections between two forms of self-consciousness, minimal and reflective. The overview discloses how the research question will be answered. The third part of the introduction (1.3.) presents the research methods that will be used. This part explicates the basic ideas of neurophilosophy, which include the use of conceptual analysis, cooperation between philosophers and empirical scientists, and the use of concrete example cases in the inquiry of self and mind.

1.1.1. Questions of self in the philosophy of mind

The nature of self has intrigued people in and out of season, in all cultures. In the following, the history of self in western philosophy is described very briefly in order to give a general idea of the significance of the self.[2] Since antiquity, philosophers have encouraged the seeker to "know thyself." In Ancient Greece, Socrates equated his essential self with his psyche, or soul, and since antiquity the concepts of self and soul have been connected together

---

[2] This brief description of the history of the concept of self follows mainly the interpretation of Barresi & Martin (2011). Since selfhood has been an extensive theme through the history of philosophy, this rough description cannot include all interesting aspects in the study of self. For more detailed descriptions, see, e.g., Barresi & Martin, 2011; Martin & Barresi, 2006; Seigel, 2005; Solomon, 1988; Sorabji, 2006; Thiel, 2011. For ideas of self in eastern philosophy, see, e.g., Albahari, 2006; Ganeri, 2012; Siderits et al., 2010.

(Barresi & Martin, 2011). But in the seventeenth century, Descartes started to use the word "mind" as an alternative to the word "soul." According to Descartes, the self is the mind, which is an unextended substance whose essence is thinking. Thus, Descartes entered upon a new epoch in the philosophy of self, and his insights on the fundamental characteristics of 'I' became one of the cornerstones of western philosophy. "Cogito, ergo sum–I think, therefore I am" is a certain truth that one cannot doubt. Descartes argued that although one can be skeptical about all other things, she cannot question the existence of her consciousness since the occurrence of her thinking proves the thinking exists and thus also that she, as the thinker, exists. Although one can doubt the existence of the outer world or whether her sense experience gives a trustworthy picture of it, she cannot doubt that she exists and is having such experiences. That is, the existence of self is something fundamental in our lives.

After Descartes, selfhood continued to be connected to essential features of mind. Locke argued that the persistence of selves and persons[3] could be understood in terms of the unifying role of consciousness and also connected selfhood to the topic of personal identity[4]. A little later, Hume argued that the idea of a substantial, persisting self is an illusion and analyzed the rise and functional role of this illusion. In the eighteenth century, Kant constructed a synthesis upon previous rationalists' and empiricists' accounts, including self. According to Kant, there are two senses of self: phenomenal self can be studied empirically, but of noumenal self we cannot know anything else but its existence. However, this

---

[3] Locke equated 'selves' and 'persons' and, as a terminological note, 'self' and 'person' are often used synonymously nowadays too (e.g., by Barresi & Martin, 2011). However, for some authors the terms have different referents (e.g., Frankfurt, 1988). In this dissertation, these terminological choices are not touched upon in detail; the focus is primarily on 'self-consciousness' and thus rather on 'self' than 'person', although MSC is connected 'first-person perspective.' However, 'self' can be considered to be a more general or wider concept, whereas 'person' more often refers to beings with RSC; also infants and (at least some) animals arguably have MSC and thus are 'selves,' but in order to be a 'person' one needs more developed cognitive skills.

[4] Locke emphasized the significance of reflective self-consciousness for personhood and is a classic in the discussions of personal identity that were also popular in twentieth-century philosophy of self, see, e.g., Parfit, 1971; Whiting, 2002.

latter self had a fundamental importance for Kant, since Kant considered it as a precondition of experience without which one could not have experiences. In Kant's view, self is a logical agent or subject of experience that combines the elements of thought together; without self, one could not think at all. In practical philosophy, Kant considered self as a source of autonomous agency that is needed for moral responsibility. Kant's account of self had a great influence on successive philosophies of self too.[5]

An overarching idea in the philosophies of self had been an aspiration to define a unified self; however, this tendency turned into more scattered accounts after the Second World War. The self became a target of unintegrated scientific specialization as well as philosophical critiques of deconstruction and postmodernism. As Barresi & Martin (2011, 51) describe, "the self, which began the century looking unified—the master of its own house—ended it looking fragmented—a byproduct of social and psychological conditions." There no longer was a concerted attempt to unify self-theories with one another, but by the end of the century, practically every discipline that studied human nature (including biology, sociology, anthropology, political science, and neuroscience) wanted to define self from its own perspective. This provided progress with respect to limited questions. However, these questions were increasingly separated and their answers incommensurable.

At the beginning of the twenty-first century, an interest towards self has been growing again, and indeed there is no shortage of concepts of self that one can examine. The following list compiled by Strawson (1999, 100) shows how thinkers have tried to capture the self by dividing it into different facets:

> the cognitive self, the conceptual self, the contextualized self, the core self, the dialogic self, the ecological self, the embodied self, the emergent self, the empirical self, the existential self, the extended self, the fictional self, the fullgrown self, the interpersonal self, the material self, the

---

[5] Including German idealism and phenomenology, see, e.g., Martin & Barresi, 2006; Solomon, 1988.

narrative self, the philosophical self, the physical self, the private self, the representational self, the rock bottom essential self, the semiotic self, the social self, the transparent self, and the verbal self (cf. Butterworth, 1998; Cole, 1997; Dennett, 1991; Gallagher & Marcel, 1999; Gazzaniga, 1998; Gibson, 1994; James, 1890; Legerstee, 1998; Neisser, 1994; Pickering, 1999; Sheets-Johnstone, 1999; Stern, 1985)

With these kinds of lists, one might be tempted to ask which one of the concepts is the real or proper self. However, arguably a sounder strategy is to ask how these different conceptions connect together; a unified self is many-sided, and its sides need to be recognized in order to achieve a full understanding of it. Recently, there have been attempts to find a multidimensional characterization of self that is able to combine different accounts (this kind of attempt is described in more detail in 1.1.3).

1.1.2. Questions of self in empirical sciences

In addition to philosophical interest, selfhood has been widely studied within empirical sciences. For instance, psychology and psychiatry have been interested in selfhood as an important feature of mind, and these sciences have also produced noteworthy theories of self. The theories of self have played so significant a role in these sciences that they cannot be characterized even briefly in this dissertation. Thus, the descriptions below do not aim to be exhaustive but only to bring out how selfhood is an important research subject through mind sciences.[6]

Early in the field of psychology, William James (1890) proposed a sophisticated scientific philosophy of the self. According to James, the experience of self is primary, but self can be regarded both as an object and as a subject (a 'me' and an 'I'). The self as an object can be divided into the material, the social, and the spiritual self, each of which can be still further divided. About a hundred years later, another famous psychological theory on the many facets of self was

---

[6] In addition to the below-mentioned theories of psychology, self has been significant, for instance, in developmental psychology (Rochat, 2001; Stern, 1985) and psychoanalysis (Freud, 1961).

proposed by cognitive psychologist Ulric Neisser. Neisser (1988) suggested distinctions between the ecological, interpersonal, private, extended and conceptual self. In addition, the concept of self has appeared and been investigated within psychology in different forms, or as Barresi and Martin (2011, 51) put it, in "many hyphenated roles," including self-image, self-esteem, self-modeling, and self-control.

In psychiatry, the disintegration of self has been argued to underlie mental disorders. As Minkowski (1997, 114, cited in Zahavi 2014, 5) stated, "The madness ... does not originate in disorders of judgment, perception or will, but in a disturbance of the innermost structure of the self." Recently, the role of self in the origins of mental disorders has been emphasized, especially by phenomenologically oriented psychiatrists (see Sec. 6.1.1.). For instance, in schizophrenia studies, altered self-experience has been considered a key factor (e.g., Fuchs, 2015; Sass & Parnas, 2003; Zhao et al., 2013).

In the twentieth century, selfhood has also been a research subject for new sciences, including cognitive science[7] and neuroscience[8]. Cognitive science can be seen as the interdisciplinary study of mind and intelligence (e.g., Bermúdez, 2014; Thagard, 2013). Since selfhood has a major role in the functions and operations of mind, it is an interest in cognitive science. In addition, empirical neuroscience has produced new knowledge about the structures and functioning of the brain, including processes that are significant in the emergence of self-consciousness (see, e.g., Searle, 2005). Since the 1990s, neuroscientists and cognitive scientists have integrated their methodologies into cognitive neuroscience, which studies both cognitive processes and the brain processes underlying them (e.g., Bechtel et al., 2001).

---

[7] The intellectual origins of cognitive science lay in the mid-1950s in the movement to develop theories of mind based on complex representations and computational processes. Cognition was understood as the processing of information and mind as a computer-like machine that performs complex operations on a great amount of information. Nowadays, enactivism presents an alternative way to understand cognition; see Sec. 2.1.1.

[8] The term 'neuroscience' was introduced in the 1960s for a collaborative inquiry into the nervous system by scientists trained in a number of specialties, including psychology, neuroanatomy, neurochemistry, and neurophysiology (Bechtel et al., 2001).

Altogether, the interest in studies on self has increased within many disciplines in the last decades. From the side of philosophy, one reason for this growing interest seems to be related to changes in our general world view. Over millennia, the concept self was often coupled with the concept of soul or mind (e.g. by Plato, Descartes and Locke); self was seen as something different from a body and as a seat of thinking and morality. However, during the twentieth century, the common picture of the world has become more and more material, which has led to deliberating the definition of self in a new manner. Nowadays, it is more common to think that mind and body, or mental and physical, are not independent of each other but closely connected, since the mind emerges from bodily and neural processes (see 1.3, Chapters 2 and 4). Modern experimental methods, such as brain imaging techniques, enable much more detailed examination of physiology than ever before. However, mere metrics in brain imaging studies cannot reveal anything about mind by themselves; these results need interpretation and theoretical background in order to increase our understanding of the mind and self.

Indeed, since the discussion on self is growing in multiple fields, conceptual clarity in a multidisciplinary approach to self is needed more than ever. However, researchers from different fields have not always communicated with each other but have often been restricted to their own specific viewpoints. Nevertheless, it seems that cooperation could benefit all the disciplines: by learning from each other, they could together develop a comprehensive conception of self. Recently, steps for building this kind of bigger picture of self and mutual exchange of knowledge have taken place, and this dissertation also advances this multidimensional approach. In order to delineate common ground for different conceptions, the pattern theory of self will be considered next.

1.1.3. The pattern theory of self

Shaun Gallagher has developed a pattern theory of self "as a useful way to organize an interdisciplinary approach to discussions of what constitutes a self" (Gallagher 2013, 1; Gallagher & Daly, 2018;

see also Newen, 2018).[9] As illustrated by the examples above, self has been understood in various ways in different contexts, disciplines and eras. The pattern theory of self is helpful since it can clarify various notions of self as compatible or commensurable instead of creating conflict between them. According to the pattern theory, 'self' is a cluster concept with several characteristic features, a particular pattern of which constitutes an individual self.

Gallagher (2013; Gallagher & Daly, 2018) poses a tentative list of the aspects that are involved in specific self-patterns. Gallagher sees these aspects as variables that can take different values and weights, and a combination of these variables constitute a self. However, a particular self may lack a particular aspect and still be considered a self. Gallagher expresses the following list of aspects that are included in the pattern theory of self (Gallagher & Daly, 2018, 4; developed from Gallagher, 2013, 3-4):

> 1) *Minimal embodied aspects*: Core biological, ecological and interoceptive factors, allowing the system to distinguish between itself and what is not itself – extremely basic to all kinds of animal behavior.
>
> 2) *Minimal experiential aspects*: First-person, pre-reflective, conscious experience, reflecting the self/non-self distinction, manifest in various sensory-motor modalities (kinesthesia, proprioception, touch, vision, etc.) – including a sense of ownership (the "mineness" of one's experience) and a sense of agency for one's actions (Gallagher, 2000, 2012; Rochat, 2011).
>
> 3) *Affective aspects*: Affect/emotion/temperament, ranging from bodily affects to what may be a typical affective or emotion pattern (Newen et al., 2015).

---

[9]Pattern theory can be seen as an advancement of Gallagher's prior distinction between minimal and narrative self (2000). Gallagher has further elaborated his ideas of self in books (2006, 2017) and in a large number of papers (e.g., 2011, 2012, , 2017). In addition, Gallagher has edited *The Oxford Handbook of the Self* (2011).

4) *Behavioral aspects:* Behaviors and actions make us who we are—behavioral habits reflect, and perhaps actually constitute, our character. This is a classic view that goes back at least to Aristotle.

5) *Intersubjective aspects*: Human are born with a capacity for attuning to inter-subjective existence, which develops into a social self-consciousness – a self-for-others (Mead, 1913), manifested behaviorally in mirror self-recognition (Gallup et al., 2011) and the neuronal mirror system (Gallese, 2014).

6) *Psychological/cognitive aspects*: Traditional theories of the self focus on these factors, which may range from explicit self-consciousness to a conceptual understanding of self as self, to personality traits of which one may not be self-conscious at all – psychological continuity and the importance of memory are highlighted in the literature on personal identity (e.g., Shoemaker, 2011).

7) *Reflective aspects*: The ability to reflect on one's experiences and actions—closely related to the notions of autonomy and moral personhood, including the capacity to reflect and form second-order volitions about one's desires (Frankfurt, 1982; Taylor, 1989).

8) *Narrative aspects*: Our self-interpretations have a narrative structure. Although some theorists make the strong claim that narratives are constitutive for selves (Schechtman, 2007, 2011), for the pattern theory of self one can lose the ability to construct a self and still remain a self to the extent that other elements of the pattern remain in place. On some views it is important that narratives are generated by the brain, which leads some to consider narratives mostly as fictions (Gazzaniga, 1998) and selves as abstract "centers of narrative gravity" (Dennett, 1991, 1992).

9) *Extended/situated aspects*: Including the possibilities presented by physical pieces of property, and various

things that we own (James, 1890). Not only may we identify with our material belongings, the technologies we use, our professions, and the institutions we work in, but we are dynamically related to the action possibilities they afford.

10) *Normative aspects*: Ranging across possibilities presented by the kind of family structure and situation in which we grew up to cultural and normative practices, involving physical and mental health, gender, race, and economic status, that define our way of living.

Different theories of self have focused on different aspects, and the pattern theory provides a way to endorse the complexity of self. However, as such the pattern theory is a preliminary proposal that does not in itself tell how different aspects are joined together and what kind of relations prevail among them; more detailed accounts need to be developed. In this dissertation, I will apply the pattern theory in elaborating the dimensions of self-consciousness and their connections.

1.1.4. Self-consciousness
An advantageous way to approach self is to examine 'self-consciousness' since self-consciousness is the way by which we are familiar with our selfhood. In other words, self-consciousness gives a direct route to grasping selfhood. In addition, self-consciousness is one dimension that a theory of self has to explain: a comprehensive account of the self needs to take note of the characteristic ways in which we experience ourselves.

Moreover, it has been argued (e.g., by Galen Strawson, 2000) that the metaphysical investigation of the self is subordinate to the phenomenological investigation. If one wishes to answer the metaphysical questions concerning the self, one first needs to know what the self is supposed to be, and in order to establish this, one should investigate self-experiences in self-consciousness. In other words, research of self-consciousness assists in broader metaphysical studies of self: a clear picture of self-consciousness enables a better understanding of entities that can be called selves.

Following the idea that self-consciousness is a central feature in accounting for self, this dissertation focuses on self-consciousness.

The focus has the advance of narrowing the subject matter by excluding more metaphysical considerations of 'the self' that would only make the inquiry too lengthy and broad for a dissertation. However, the concept of self-consciousness is notoriously ambiguous. Roughly, self-consciousness means being conscious of one's self in some way, yet various kinds of matters are referred to by it, and this rough characterization leaves unspecified the manner in which one is conscious of one's self. 'Self-consciousness' can refer to self-directed thoughts or it can be seen as a feature of all conscious episodes. Thus, in order to wholly understand the phenomenon and to structure the discussion, the important first step is to recognize that self-consciousness comes in many varieties (e.g., Metzinger, 2013; Zahavi, 2010). In other words, talking only about 'self-consciousness' might lead to confusion and controversy if it is left unspecified which form of self-consciousness is of interest. A clear conceptual map is crucial in order to untangle the dimensions of self-consciousness and to grasp its different manifestations. Clear concepts are further required in the formulation of a solid theory of self-consciousness, which clarifies what it is like to experience self and what role these experiences play in human lives.

1.1.5. The distinction between minimal and reflective self-consciousness

The most general conceptual distinction is made between two different forms of self-consciousness: *reflective self-consciousness,* that is, explicit self-directed thoughts, and *minimal self-consciousness*, which takes place in the structures of consciousness. These forms can also be seen as levels; in this hierarchical sense, minimal self-consciousness is the fundamental lower level that is required for the higher-level reflective self-consciousness to occur. The conceptual distinction is central in this dissertation and briefly elaborated below.

Minimal self-consciousness refers to the most basic and fundamental form of self-consciousness. In this sense, self-consciousness is an implicit but necessary feature of consciousness, and 'self' is simply connected to the *subjectivity* of experience. This means that minimal self-consciousness *is present whenever I am living through an experience*. It can be called "pre-reflective"

because it is a consciousness that we have regardless of whether we do any reflecting on our experience: it is an implicit and first-order awareness. In minimal self-consciousness I am not thinking about or focusing on myself but, for instance, on perceiving the outer world; still, however, there is a subtle awareness of myself in that state: I am aware of myself as that perception's *owner* or *subject*, and this holds true for all of my experiences. Furthermore, minimal self-consciousness does not refer to any experiential content, to any specific *what*. Rather, it refers to the distinct *how* of experience, that is, to the first-personal presence of experience.

The notion of minimal self-consciousness has been called "thin" because it might seem rather technical to speak about "consciousness of oneself" on such a low level of self-consciousness, simply as a built-in feature of experience (e.g., Zahavi, 2010, 2014, 2017). Indeed, the notion is so thin that not all theorists are ready to call it self-consciousness at all but want to preserve the term for cognitively higher, explicit, reflective self-consciousness (e.g., Baker 2012 refers to this experiential level only as 'consciousness' or as 'rudimentary first-person perspective'). Yet, as we will see, although minimal self-consciousness is thin, it has an essential role in our lives since it is intertwined in every experience.

In turn, reflective self-consciousness is a "thicker" concept because self has a more obvious role in it. Reflective self-consciousness refers to the capacity to take oneself as the *object* of one's reasoning and to think of oneself as oneself, and thus it coincides with how we use 'self-consciousness' in everyday language (in English). This form of self-consciousness is cognitively more demanding and connected to language, introspection, higher-order thinking and narrativity. Thus, more aspects of self are involved in this form; it is thicker or "packed" with more layers of self. The essential role that reflective self-consciousness has in our lives is that it enables sophisticated control of action and moral deliberation.

Although the distinction is widely made and generally accepted, the two sides of the distinction are classified in very many ways. Some of these ways deriving from different research traditions are compiled on Table 1.1. below. While the main idea seems to be rather the same in all of these theories, the exact

definitions surely are slightly different. In addition, it is common to distinguish various sub-components or factors within both forms of self-consciousness. In any case, the table illustrates that the general distinction between the two forms is crucial in the study of self-consciousness and that it is widely recognized. The distinction is similar also to distinctions made in cognitive science, and it is empirically significant, since the two forms of self-consciousness are based on different kinds of neural processes (see Chapter 4).

Table 1.1. Notions for the distinction between minimal and reflective self-consciousness

Author(s)	Minimal self-consciousness	Reflective self-consciousness
Baker (2012)	consciousness; rudimentary 1PP	self-consciousness; robust 1PP
Bayne & Pacherie (2007)	agentive experiences	agentive judgments
Bermudez (2001)	nonconceptual self-consciousness	conceptual self-consciousness
Damasio (1999)	core self	autobiographical self
Fasching (2008)	the self-presence of experiencing itself	the self-identification with certain configurations of what one experiences
Frankfurt (1987)	reflexive self-awareness	self; person
Gallagher (2000)	minimal self	narrative self
Gallagher & Zahavi (2007)	pre-reflective self-consciousness	reflective self-consciousness
Goldman (1970)	non-reflective self-awareness	reflective self-awareness
Grünbaum & Zahavi (2013)	implicit self-awareness	explicit self-awareness
Kriegel (2004)	intransitive self-consciousness	transitive self-consciousness
Legrand (2005)	pre-reflexive self-consciousness	reflexive consciousness of the self
Legrand & Ruby (2009)	self-specific processing	self-related representations
Kahneman (2011)	experiencing self	remembering self
Mackenzie (2014)	embodied agent	narrative self
Metzinger (2013)	minimal phenomenal selfhood	cognitive agency
Musholt (2013)	being a self	being aware of being a self
Sartre (1956)	non-objectifying self-acquaintance; self	objectifying self-awareness; ego
Strawson (2015)	self-intimation; reflexive self-awareness	reflective self-awareness
Zahavi (2010)	experiential core self	reflective self
Zahavi & Kriegel (2016)	for-me-ness	more sophisticated forms of self-consciousness

I propose that this broad distinction into two forms of self-consciousness is a useful preliminary step on the way to a more fine-grained analysis in which the pattern theory of self can be one useful tool. In addition to the relevance of acknowledging the difference between the two forms, the distinction is a beneficial entry point because it holds up even if exact definitions of the two forms vary on their details slightly. Thus, elaborations of the distinction can generate discussion between theories using different terms.

1.2. Overview of the contents

1.2.1. Research question and objectives

Both minimal and reflective self-consciousness have been long examined; however, the majority of studies have focused only on one or the other form. Indeed, it is surprising that the interconnectedness of these two forms is so little studied. Nevertheless, it is crucial to investigate the links between these forms in order to understand selfhood in its entirety. Many researchers (including Bayne & Pacherie, 2007; Gallagher, 2013; Musholt, 2013; Zahavi, 2010a) have also explicitly pointed to the relation between minimal and reflective self-consciousness as being in need of more study. Drawing on this need, the research question of the dissertation is: *how are the two forms of self-consciousness connected to each other*? In examining this question, I will conduct a systematic study of the connections between minimal and reflective self-consciousness, which has been lacking from current studies of self-consciousness (Part I). In addition, I will show the significance of these connections in practice by analyzing different cases of self-consciousness (Part II).

The objective of this dissertation is to produce conceptual clarifications and novel knowledge about selfhood. The aimed clarifications concern the following matters:

- The relation between self-consciousness and consciousness: is self-consciousness a sub-feature of

consciousness, or are consciousness and self-consciousness always intertwined? (Chapters 2 and 5)
- The structure of self-consciousness and the factors that affect it: are 'minimal' and 'reflective' sufficient descriptions for capturing the structure of self-consciousness, or can both of them be further divided into precise sub-components? (Chapters 2 and 3)
- Which one of the two forms is more fundamental? What kind of bottom-up and top-down relations prevail between them? (Chapters 2-5)
- The unity of self-consciousness: how do the forms of self-consciousness function together? (Chapters 4-8)
- How are different conceptions of self-consciousness related to each other? (All chapters)

These conceptual clarifications are neurophilosophical in the first place; however, together they form a multilevel model of self-consciousness that can be applied also in other fields of philosophy. In addition, precisely formulated concepts are useful for empirical and clinical application: they have an important role for instance in the development of empirical theories and interpretation of experimental results. Thus, the general aim of this dissertation is to promote multidisciplinary research on self.

1.2.2. Phenomenological viewpoint

A couple of theoretical remarks about the conceptual framework are needed before starting the examination. First, the conceptual starting point of this dissertation lays in phenomenological approach to self-consciousness, here represented especially by Dan Zahavi's work (also in collaboration with Gallagher, e.g. in Gallagher & Zahavi, 2007).[10] Generally, 'phenomenology' as a philosophical tradition refers to the detailed study of experience and its structures (Zahavi, 1999, 2005b, 2010b). Since this dissertation concentrates on self-*consciousness*, it is necessary to

---

[10] Zahavi's philosophy is so much endorsed here that many times I simply substitute Zahavi's notions of 'pre-reflective self-consciousness,' 'experiential self,' and 'for-me-ness' for 'minimal self-consciousness.'

take consciousness, that is, experience or phenomenology, seriously. Because of this major role of experience in self-consciousness, the phenomenological tradition is a logical and promising starting point and offers an elaborated conception of self-consciousness.

Zahavi's account is chosen here because he has done extensive work in formulating a sophisticated multidimensional account of self that is highly relevant for the current multidisciplinary discussion of self-consciousness.[11] Zahavi argues most clearly for the idea and significance of the minimal form of self-consciousness. Zahavi's notion of 'pre-reflective self-consciousness' or 'experiential self' is useful, since it is precise enough to be compared with other notions. At the same time, the notion is not too restricted to phenomenological tradition but has potential to be connected to other conceptions — which Zahavi also encourages (e.g., Zahavi, 2010b, 2014). In short, the Zahavian basis for a conception of self-consciousness means that the experiential dimension of self is ascribed to be primary: it is the fundamental aspect of self that also grounds self-reflection. However, distinguishing only the minimal and reflective forms is not fine-grained enough to capture all the subtleties of self-consciousness.

In order to carry out a more fine-grained analysis, I use Gallagher's pattern theory of self as a heuristic tool in sorting out different conceptions and manifestations of self-consciousness. According to Gallagher (2013, 4), "One benefit of the pattern theory of self is that we can more clearly understand various interpretations of self as compatible or commensurable instead of thinking them in opposition." I apply this idea to the notions of self-consciousness; some notions of self-consciousness emphasize its psychological/cognitive aspects whereas others focus more on experiential aspects. Still, both are speaking about self-

---

[11] Zahavi has done impressive work on the connections between self-consciousness, selfhood and consciousness for more than 20 years. His account is presented in three monographs (Zahavi, 1999, 2005b, 2014), a number of papers (e.g., 2005a, 2018) and chapters in edited books (e.g., Zahavi, 2010a, 2010c, 2017, 2020). In addition, Zahavi has worked as an editor of books *Exploring the Self* (2000) and *Self, No Self?* (2010, together with Siderits & Thompson).

consciousness with good reason. These kinds of different notions do not contradict each other; they help in understanding self-consciousness as a whole. Thus, the research question can easily be approached in terms of the pattern theory, and it seems to be exactly the kind of question of self that the pattern theory is beneficial in dealing with. However, I do not commit to pattern theory as such or consider it the final way of drawing distinctions within self, but the theory is clear and illustrative enough for the purposes of this dissertation. That is, the pattern theory gives a way to combine many facets of self, which is needed in forming an exhaustive view of self-consciousness.

1.2.3. Structure of the dissertation

For the purposes of providing an analysis of the connections between minimal and reflective self-consciousness and indicating the significance of these connections, this dissertation consists of two parts. The first part lays the conceptual basis for describing self-consciousness, and the second part applies this conceptual account of self-consciousness to example cases. In other words, the storyline of the dissertation proceeds from general to specific. First, I present the research field and the theoretical framework (Chapter 1). Then, I define the main concepts (Chapters 2 and 3) and begin to look at how they are connected together (Chapters 4 and 5). After forming the general conceptual picture, I apply it to specific examples (Chapters 6-8). In the following, the structure of the dissertation is briefly outlined.

Chapter 2 focuses on the definition of minimal self-consciousness. I consider the embodied and experiential nature of minimal self. I argue that in the minimalist reading, minimal self-consciousness incorporates only the experiential aspect of self, which is always present in experience and has a special status in the pattern theory. However, in the robust reading, minimal self-consciousness also involves other "pre-reflective" aspects such as affective, intersubjective and situated aspects. That is, 'minimal self-consciousness' in this dissertation is hierarchically the low level of self-consciousness that is always involved in experience but can also involve variance.

Chapter 3 focuses on reflective self-consciousness. First, I characterize reflective self-consciousness with remarks about the unique epistemic and motivational features of self-conscious thoughts. Then, I argue that this remark cannot alone capture the complexity of self-reflection, but an account of self-consciousness can be deepened by recognizing several features. I discuss the distinction between deliberative and theoretical stances, voluntariness of self-reflection, and identification with self.

Chapter 4 gives a look at the neural realization of the mechanisms underlying self-consciousness and broadens the concepts from earlier chapters to a neurophilosophical framework. I propose that the distinction between minimal and reflective self-consciousness can be associated with the distinction between self-specific and self-related neural processes. Differences in the mechanisms of these processes support the ideas of minimal and reflective self-consciousness and their subfeatures. On the other hand, neural studies show the evident interaction between different features of self.

Chapter 5 focuses on the connections between minimal and reflective self-consciousness, arguing that these connections have a major role for the whole of self-consciousness. First, I demonstrate the basic hierarchy in self-consciousness: minimal self-consciousness is a necessary condition for reflective self-consciousness. Second, I elaborate the interlinks of minimal and reflective self-consciousness and study their interaction by examining their bottom-up and top-down connections and looping. In addition, I propose that balance in self-consciousness is related to mental well-being.

Part two of the dissertation applies the ideas developed in the first part to concrete cases of self-consciousness. The consideration of these cases shows that the connections between minimal and reflective self-consciousness are not only theoretically interesting but also useful in practice in understanding instances of self-consciousness.

Chapters 6 and 7 focus on Cotard syndrome, in which patients claim that they do not have self, mind, or body parts, and even say they are dead. I argue that even this kind of experience does not

undermine theories that consider self-consciousness as a necessary feature of experience. Instead, the syndrome can be seen as a window to the structure of self-consciousness. The examination of previous explanations shows that the syndrome involves dysfunctions in several aspects of self. However, I argue that the previous explanations have failed to give a full account of Cotard syndrome since they have ignored the connections between minimal and reflective self-consciousness. My analysis of Cotard syndrome indicates that both forms of self-consciousness involve several subfeatures and that their interconnections are crucial. These findings are relevant for both theories of self-consciousness and empirical study of self.

In Chapter 8, I conduct a comparative analysis of altered self-consciousness in depersonalization and meditation. Depersonalization can be considered a milder form of Cotard syndrome, which involves alienation from one's self. Meditation, instead, is an altered state that aims at withdrawal from typical self-consciousness and is linked with improvements in mental well-being and cognitive skills. Both meditation and depersonalization are defined in terms of de-identification from self, but I argue that self-experiences in the two states involve remarkable dissimilarities. Overall, the findings in the examination of altered self-consciousness strengthen the multidimensional picture of self and the significance of the connections within self-consciousness proposed in Part I.

1.3. Methods and neurophilosophy

The methodological framework of neurophilosophy connects traditional philosophical conceptual analysis and philosophy of mind with empirical research. Generally, *neurophilosophy* is a two-way endeavor: on one hand, it concerns application of neuroscientific concepts to traditional philosophical questions and, on the other hand, it applies conceptual analysis and philosophical knowledge to settle foundational issues within the neurosciences (see, e.g., Bechtel et al., 2001; Bickle, 2009; Mandik, 2007, 2009).

As a rather new branch in philosophy, neurophilosophy is still developing but a general picture of it is outlined below.

1.3.1. Conceptual analysis

Conceptual analysis is one of the key methods of philosophy. An analysis is a process of isolating or working back to what is more fundamental; it is a means to explain or reconstruct something that initially is taken as given (Beaney, 2014). In other words, the aim of analysis is to get to basics: conceptual analysis aims to determine the basic meaning of concepts by considering what constituents the concepts involve. Typically, it has been accepted that conceptual analysis should yield definitions, traditionally by specifying necessary and sufficient conditions of a concept. Even though defining these conditions can be tricky, the ideal of specifying them is still worthwhile.[12] As Beaney (2014) remarks: "The specification of necessary and sufficient conditions may no longer be seen as the primary aim of conceptual analysis, especially in the case of philosophical concepts such as 'knowledge,' which are fiercely contested; but consideration of such conditions remains a useful tool in the analytic philosopher's toolbag." 'Self' can be seen as this kind of contested concept, and one idea of the pattern theory of self is to formulate a theory that can account for the variety and flexibility of self without being restricted to traditional necessary and sufficient conditions. However, it still is advisable to formulate these kinds of defining conditions that indicate the status of constituents of a concept, thus enabling a clear formulation of the concept. Conceptual analysis has been practiced especially in the tradition of analytic philosophy. Nowadays, conceptual analysis is also used in neurophilosophy, the background of which lies in analytic philosophy.

---

[12] An example of conceptual analysis determining necessary and sufficient conditions is the so-called classical definition of knowledge as true justified belief. S knows that p iff 1) p is true, 2) S believes that p and 3) S is justified in believing that p. This classical definition was refuted by Gettier's (1963) famous counterexamples against the sufficiency of conditions 1 to 3. Gettier cases inspired philosophers to elaborate the analysis and to formulate refined conditions for knowledge. For disputes about conceptual analysis, see, e.g., Beaney, 2014.

1.3.2. Experience and the brain

Before introducing neurophilosophical ideas in more detail, a few words about the relation between consciousness and the brain are needed in order to better understand the special characteristics of consciousness. By taking into account the differences between mind and brain and the problem of identifying the two, it is easier to avoid easy fallacies or oversimplifications of the matter and to see how different disciplines can contribute to each other.

The essence of consciousness, along with self-consciousness, is its subjective character; there is something it is like for a subject to undergo experiences (e.g., Nagel 1974; the subjective character is also considered in Secs. 2.1. and 5.1.). Nagel's (1974) idea of this "what-it-is-likeness" has been central in modern philosophy of mind. That is, consciousness in general is characterized by a *first-person perspective*: we are conscious of ourselves in first-personal manner and immediately familiar with our experiences. Importantly, my first-person perspective encompasses my experiences as they are uniquely for me. In contrast, the results of empirical studies describe (self-)consciousness from a third-person perspective and aim to generate objective conclusions that can be generalized across groups of people. Already this indicates that first-personal self-consciousness and third-personal neuroscientific description of it simply are not the same. Straightforwardly, too, we do not experience our consciousness as any kind neural processing (but as lively experiences), and hence it is obvious that experience and underlying neural processing are not equivalent.

Due to its subjective character, the problem of consciousness is unique for scientific research. Nagel (1974) argues, and phenomenologists like Zahavi agree, that the subjective character of experience is not captured by reductive analyses; neither it can be explained by functional states or intentional states. Nagel (1974) famously uses the experience of a bat as an illustration of the divergence of subjective and objective conceptions of experience. Since bats are mammals, it is reasonable to assume that they have experiences as humans do. However, the sensory system of bats is very different from ours; for instance, bats have echolocation in addition to vision. Because of these differences, the experiences of

bats are beyond our capacity to conceive; their subjective experience of the world is not like anything we can experience or imagine. Of course, I can try to imagine what it would feel to look like a bat or behave like a bat, but when I try to imagine what it is really like to be a bat, my imagining fails. Without modifications in my biological structure, my extrapolation of bats' experience always remains incomplete. Thus, even if subjective experience can be studied by empirical sciences successfully, it never can be reduced to, or described exhaustively by, an objective scientific explanation.

Chalmers (1996, 2007) has approached the same point by distinguishing easy problems of consciousness from a hard problem. *Easy problems* refer to our abilities, for instance, to discriminate stimuli, report information, monitor internal states, or control behavior. These problems are answered by explaining certain behavioral or cognitive functions, and there is good reason to believe that eventually the explanations of the mechanisms for these functions can be given in neurobiological or computational terms. *The hard problem* of consciousness, instead, is more difficult. It is the problem of subjective experience: that there is something it is like to have mental states and be oneself. A solution to the hard problem cannot be given by referring to functions or structures. Even if we had an explanation of all the relevant functions of consciousness, the further question still remains: why is the performance of these functions accompanied by experience?[13]

Further, the fundamental difference between first-personal experience and third-personal science has been described as an

---

[13] It is worth remarking that often the idea of the special subjective character of consciousness is used as an argument against materialism. Solving the hard problem requires explaining the relation between mental and physical, and since materialist explanations that appeal only the physical are unable to do that, they must be rejected. In this dissertation, this metaphysical question of the nature of consciousness is not touched upon. However, it can be remarked that, according to many philosophers (including Zahavi and Gallagher), a purely reductive or eliminativist account of consciousness is insufficient since it dismisses the subjective character of consciousness. In addition, it can be noted that Chalmers also uses the term 'qualia' in referring to the subjective phenomenal character of consciousness, but, for simplicity and brevity, the discussion on qualia is not dealt with in this dissertation.

explanatory gap that prevails between the two (Levine, 1983). Since experience is subjective, it is in its entirety in principle out of reach of any objective scientific description. In other words, the two are separated by an explanatory gap: "we don't have any way of determining exactly which psycho-physical identity statements are true" (Levine, 1983, 354). In addition, it is good to notice that some kind of a gap or difference always prevails between an experience and any description of it, even when it comes one's own experience. Subjective consciousness is immediate; one is *directly* familiar or acquainted with her experience, whereas a description of experience is always *indirect*, mediated by language or illustration. Further, descriptions of past experiences are accessed by memory and thus also dependent on our abilities to describe and remember.[14] Providing descriptions is important for analysis and understanding, but they never capture experience as such. However, it can be argued that the explanatory gap can be bridged or narrowed: experience itself and a scientific explanation of it are not identical, but scientific research on consciousness is needed since it can teach about the emergence and structures of experience.

The difference between consciousness and brain research is also conceptualized by distinguishing between personal and sub-personal levels of explanations (see, e.g., Musholt, 2013). Personal-level explanations refer to the subject's experience (or phenomenology) and conscious mental states. Sub-personal-level explanations instead provide information about the physiological or computational enabling conditions of personal-level phenomena. Sub-personal-level explanations can involve reference to unconscious mental states, but the mental states are not but experienced at the sub-personal level.[15] Talk about self-

---

[14] The difference between an immediate experience and a later report of it, and the related problems of indirectness in the study of consciousness, can also be discussed in terms of Kahneman's (2011) 'experiencing self' and 'remembering self', or Sandman's (2017) notion of 'E².'

[15] For mental states that are not conscious, for instance, subliminal perception and peripheral vision show that perception occurs without awareness of it. Also emotions, wants and

consciousness clearly is talk about personal-level phenomena, whereas results of brain-imaging studies describe the sub-personal level. It is important not to confuse these levels and to avoid the mistake of ascribing properties of only one level to the other. For instance, conscious mental states can be properly ascribed only to the person, not to parts of the person, such as the brain (or areas within the brain; Bennett & Hacker, 2003). In short, self-consciousness is experienced by a person, whereas neural processing occurs on a sub-personal level.

Since phenomenology is primarily interested in first-personal experience, Zahavi (2010b, 9) also highlights its difference from sub-personal neural processes: "it is very important to emphasize that the discovery of a significant complexity on the subpersonal level ... cannot by itself force us to refine or revise our phenomenological description." Neuroscientific results can only motivate further research, but the sub-personal and personal levels cannot be considered straightforwardly similar. Zahavi sees that phenomenology should be informed by the best available scientific knowledge and encourages cooperation between phenomenology and empirical science. At the same time, Zahavi (2010b, 14) emphasizes that "a fruitful cooperation between the two should not make us deny their difference... Ultimately, the only way to justify a claim concerning a complexity on the phenomenological level is by cashing it out in experiential terms."[16]

1.3.3. Neurophilosophy
Drawing from the above, it seems evident that philosophy and neuroscience have a common interest in mind and consciousness. The exact research questions and methods of philosophy and neuroscience differ, but because of their shared interests, neurophilosophers think that cooperation between the two can benefit both disciplines (e.g., Bechtel et al., 2001; Mandik, 2009).

---

thoughts can be unconscious; for instance, one can recognize that she was sad or angry only after somebody else points it out (e.g., Musholt, 2013; Rosenthal, 2005).

[16] Zahavi (2010b, 8) sees that in the end "the ultimate concerns of phenomenology are transcendental philosophical and that transcendental philosophy differs from empirical science."

Neurophilosophy[17] is "a sub-genre of naturalized philosophy – philosophy that embraces Quine's (1969) vision of philosophy as continuous with the natural sciences–" (Mandik, 2007, 418) and takes neuroscience in the primary focus. This kind of naturalized philosophy is not just an analysis of science but essentially involves a dialogue between the sciences and philosophy. The naturalized approach to understanding the mind and brain considers them as part of the natural world and recognizes the biological, evolutionary, and environmental processes that are involved in shaping them. However, a philosophical theory is not rendered neurophilosophical by just mentioning the brain. Instead, neurophilosophical appeals involve explicit and detailed use of contemporary neuroscientific literature. Furthermore, it is important to note that neurophilosophy does not claim that neuroscience could provide philosophical conclusions, but that contemporary neuroscience plays role in the premises of the arguments for those conclusions (Bechtel et al., 2001; Mandik, 2009).

In a broad sense, which is employed in this dissertation, neurophilosophy refers to the naturalistic endeavor of cooperation between philosophers and empirical scientists.[18] This involves a two-way interaction between philosophical and neuroscientific analysis. On the one hand, neuroscientific results might in some cases motivate the reexamination of philosophical analysis, even if they cannot in and of themselves force a philosophical re-description. Neuroscientific data can indicate that a certain kind of

---

[17] The term 'neurophilosophy' originates from Patricia Churchland's book *Neurophilosophy* (1986), which aimed to introduce neuroscience to philosophers and philosophy to neuroscientists. See also next footnote.

[18] Especially, the term 'neurophilosophy' (in the sense used in this this dissertation) is meant to be neutral in its commitments to metaphysics of consciousness. Churchland's (1986) book was sympathetic towards reductive materialism; however, the general endorsement of naturalism in neurophilosophy does not tie to any particular account of metaphysics of consciousness but can be connected to several views, including many forms of emergent materialism and panpsychism. Further, although 'neuro'philosophy obviously highlights neuroscience, and neuroscientific research is referred to in this dissertation, the general idea of neurophilosophy lies in naturalism and an interdisciplinary approach on studying mental phenomena.

neural processing is necessary for certain experiences, since if one simply lacks or loses that kind of neural processing, she also loses that kind of experience (e.g., cerebral achromatopsia). If neurobiological findings suggest changes in a conceptualization of the nature of certain conscious experience, then this information might contribute to what constitutes the experience in question—at least, it recommends closer examination of the experience. In certain circumstances, an appeal to sub-personal mechanisms might help us explain both the presence and breakdown of personal-level phenomena. Thus, sub-personal neural mechanisms can be considered as constitutive conditions for experience, bringing out the necessary material and functional basis of consciousness. In other words, information on its underlying mechanisms can reveal aspects of the personal level too, and in this way, empirical studies can be relevant for conceptual ones.

Further, the scientific study of consciousness provides philosophers with the opportunity to reference concrete case studies in their argumentation. Instead of traditional thought experiments (which can be disputable, since intuitions might not always meet), philosophers can analyze real-world examples. This kind of empirically informed appeal to concrete experiences and the description of the neural processing can also assist in illustrating the different aspects of consciousness or self, in this way making the abstract concepts more concrete.

On the other way around, philosophical conceptual analysis assists in bringing conceptual clarity to the explanandum in neuroscience and in making sense of and systematizing the results of empirical studies. In addition, clearly defined concepts are crucial in the elaboration of empirical theories, in making tenable inferences from results of experiments, and in developing empirical paradigms. It has recently been argued that psychology has not made much progress and is in a theory crisis (for an overview, see Eronen & Bringmann, 2021). The crisis results from difficulties in developing theories, and one reason for the difficulties is the lack of validity of psychological constructs. It is further argued that mathematical modeling is unlikely to solve the crisis, but clearly and transparently defined concepts would (Eronen & Bringmann,

2021). These remarks highlight the importance of conceptual clarity that neurophilosophy and this dissertation aim at.

Precise concepts are crucial, especially in studying complex phenomena such as self. Since selfhood comes in many varieties, the results of empirical studies remain oversimplified and lead to too-strong interpretations if they outright claimed to have found or settled the self in the brain. Without recognizing different aspects of self, empirical theories cannot generate experimental design that could properly target relevant aspects of self and avoid measuring errors. Altogether, loosely defined concepts lead to unfocused experiments, which result in inaccurate or biased conclusions, and thus, conceptual clarity is essential to the scientific study of self.

Conceptual meticulousness assists in recognizing the difference between experience and the brain and different explanatory levels in the study of self (see also Sec. 4.2). As an example, Craig uses the concept of 'sentient self' in developing a notable theory of the significance of the insula in the generation of bodily feelings for a basic sense of self (Craig, 2009, 2010). Craig (2010, 563) defines the sentient self as "the ultimate neural representation of all feelings that is engendered in network hubs in the left and right anterior insulae." However, as presented in this dissertation, a self is a person who undergoes experience, makes decisions, and acts in the world. According to this conception, a sentient self is not a neural mechanism but the neural mechanism is a necessary material (non-conscious) basis for being a sentient creature. That is, self is not any neural process and not in the brain, but a person whose self-experience is enabled by neural processes.

Consistent with these lines, Zahavi and Gallagher consider phenomenology connectable with naturalism and the cooperation between phenomenology and natural sciences beneficial (e.g., in Gallagher, 1997, 2003; Zahavi, 2010b, 2014). According to Zahavi (2010b, 9) naturalized phenomenology "engages in a meaningful and productive exchange with empirical science."[19] The

---

[19] However, Zahavi (2010b, 14) highlights that 'philosophical naturalism' should be understood as this kind of modest proposal of meaningful and productive exchange between phenomenology and empirical science. Instead, a radical proposal of philosophical

naturalization of phenomenology "wouldn't merely be a question of employing phenomenological insights in the empirical investigation of the mind. Rather, the idea would be that the influence goes both ways, that is, it would also be a question of letting phenomenology profit from—and be challenged by—empirical findings. This is why it is entirely appropriate to speak of a *mutual enlightenment* (cf. Gallagher 1997)." In this enlightenment, phenomenological insights can be employed in the empirical investigation of the mind. Phenomenology can be indispensable in the endeavor to provide a precise description of the explanandum, which is essential for the project of identifying and localizing the relevant neurobiological correlates. Further, phenomenological analyses and distinctions can elucidate basic theoretical assumptions made in cognitive science and empirical science, aiding in the development of new experimental paradigms. On the other hand, phenomenology cannot ignore concrete findings of empirical science but must be able to accommodate them, and empirical evidence can even force revisions in phenomenological analyses. Thus, although Zahavi and Gallagher are not explicit representatives of neurophilosophy, the neurophilosophical endeavor is compatible with their views on the benefits of exchange between phenomenology and empirical sciences.

Sometimes, 'neurophilosophy' is further distinguished from 'philosophy of neuroscience'. The former applies neuroscientific concepts and results to philosophical topics, usually in the philosophy of mind, whereas the latter is the philosophical examination of neuroscience. In other words, philosophy of neuroscience is a sub-discipline of the philosophy of science, while neurophilosophy brings neuroscientific theory and data to contribute to questions in the philosophy of mind concerning phenomenal consciousness. This dissertation does not explicitly

---

naturalism is misguided. This "radical proposal sees the naturalization of phenomenology ... as one that will eventually make phenomenology part of, or at least an extension of, natural science and it argues that this is something we should aim for." Generally, phenomenologists argue against "scientism" according to which all the questions in the world should and could be unambiguously answered by natural sciences.

distinguish between the two but concerns both directions of exchange between philosophy and empirical sciences. The main focus here is on 'neurophilosophy' in the above-mentioned sense; defining the concepts of self-consciousness is the major objective and empirical case studies are also applied in this examination. On the other hand, mostly in Chapter 4, conceptual knowledge of self-consciousness in philosophy is applied in order to make sense of empirical results and to give suggestions on how to approach the intricacy of self in empirical settings. This kind of application could be counted as 'philosophy of neuroscience'.

It is worth noticing that in addition to neurophilosophy, other multidisciplinary approaches to the study of consciousness have also been launched. These include neurophenomenology and front-loading phenomenology, which both explicitly draw from phenomenology. Neurophenomenology[20] is "a neuroscientific research program whose aim is to make progress on these issues associated with the explanatory gap" (Thompson et al., 2005). A central idea in neurophenomenology is the use of first-person methods or first-person data to reveal new third-person data about the physiological processes underlying consciousness. These methods involve training experimental subjects to be more sensitive to their own experiences, thereby generating new data. Without the possibly limiting effects of predetermined theoretical categories, trained subjects describe their experiences using an open-question format. Their descriptions are systematized to categories and then the categories are validated intersubjectively and used in the interpretation of neuroimaging data. On the other hand, neurophenomenology proposes that insights from, for instance, neurobiology and dynamical systems theory may assist in refining classical phenomenological analyses (see, e.g., Lutz & Thompson, 2003).

---

[20] The neurophenomenological research program was initially presented by Varela (1996) and further developed by Lutz (2002), Lutz and Thompson (2003), and Thompson (2007). Recently, for instance, Berkovich-Ohana et al. (2020) argue for the importance neurophenomenology (in meditation research) and give a review of neurophenomenological studies.

Front-loading phenomenology[21] instead uses phenomenology from the other direction: the idea is not to focus on the training of experimental subjects but to start with the experimental design. Here, insights developed in phenomenological research are used to inform the way experiments are set up. For instance, experimenters have used the phenomenological conceptual distinction between the sense of agency and the sense of ownership (Gallagher 2000, described in Sec. 2.4.1.). The distinction has been utilized in experiments that attempt to distinguish the neural correlates of self-agency in contrast to other-agency. For example, the study of Farrer and Frith (2002) showed that the sense of agency for one's actions correlates with the anterior insula activation bilaterally. In contrast, the sense that a perceived action is caused by someone else was connected to activation in the right inferior parietal cortex.

Although the theoretical roots of this dissertation generally lie in Zahavi's phenomenological approach to self, neurophilosophy is used as its method instead of the phenomenology-rooted variations since neurophilosophy provides a wider and more neutral framework. Neurophenomenology and front-loading phenomenology explicitly draw on phenomenological tradition as well as an enactivist conception of cognition. Instead, neurophilosophy can be considered a more neutral endeavor that is not committed to particular philosophical theories or claims. Thus, the scope of broadly defined neurophilosophy is more extensive; it does not explicitly involve the use of first-person methods or use phenomenological categories in design of experiments, but it can utilize and analyze the results and methods of the great number of empirical experiments already carried out. Thus, neurophilosophical research and results might also be more easily and faster produced and applied in further research that is not committed to neurophenomenology. This does not mean that neurophilosophy and neurophenomenology are rivals but rather

---

[21] Originally proposed by Gallagher (2003). For a recent discussion of front-loading phenomenology and other phenomenologically inspired empirical methods, see (Martiny et al., 2021).

complementary endeavors to create discussion between philosophy and empirical sciences. In other words, the methodology of this dissertation does not disagree with ideas of neurophenomenology. However, I neither want to argue for neurophenomenological ideas nor to consider the possible problems that they involve here; I want to keep the treatment neutral and focused on the concepts of self-consciousness. Further, this methodological choice aims to create an interesting viewpoint on Zahavi's research: a Zahavian conception of self-consciousness will be applied to empirical example cases, in which it can show its strength or instead reveal possible deficiencies in its explanatory power. This way, the Zahavian conception can be considered in and reconciled with scientific theoretical frameworks that exceed phenomenology.

In summary, this dissertation endorses a multidimensional conception of self and employs a combination of a neurophilosophical method and a phenomenological conceptual basis. This combination ascribes experience to a prominent role in a theory of self-consciousness and entails a non-reductionist version of neurophilosophy (see, e.g., Zahavi, 2014; Gallagher, 2013). In addition to recognition of the difference between personal and sub-personal levels of explanation, the phenomenological viewpoint highlights that the relation between these levels is not a simple reduction of the former to the latter. This was considered in several points in this chapter; the first-personal character of self-consciousness cannot be fully captured in terms of purely functional or neural description, even though neural processes are considered constitutive for self-consciousness.[22] This argument against reductionism is rather straightforward from the phenomenological standpoint that considers experience primary (e.g., Zahavi, 2005b, 2010b, 2014). In addition, Gallagher (2013, 6) explicitly proposes that the pattern theory is a non-reductionist

---

[22] In addition, the multidimensional conception of self does not exclude the possibility that there could be other constitutive elements in addition to neural mechanisms. For instance, the enactivist approach proposes this kind of constitutive role also to the whole body and interaction with the environment (see, Sec. 2.1.1.).

theory of self: "The patterns at stake in a pattern theory of self are not reducible to neuronal patterns, or patterns of brain activation." At the same time, the pattern theory aims to provide understanding about self that is empirically plausible and to dissociate from a "traditional" theory, which considers a self as a substantial (soul-like) entity.

Noteworthily, the idea of non-reductionism can be defended also independently of phenomenology or the pattern theory of self. As a mind-body theory, reductive materialism can be defined as a "theory that says that consciousness exists but it consists of only ordinary neurophysiological processes and therefore it can be exhaustively described in purely neurophysiological terms" (Revonsuo, 2010, 301). This reductive view has been widely criticized since it seems to ignore subjective psychological reality (see, e.g. Revonsuo, 2006, 2010). Instead of an identity relation between consciousness and neurophysiology, a theory of emergent materialism considers their relation more intricate. According to emergent materialism, the brain is a holistic system of complex organization and consciousness emerges from brain activities as a special type or a higher level of brain activity. The conception of emergence highlights that consciousness is a new type of phenomenon that has new kind of properties; The higher level phenomenon cannot be reduced to traditional neurophysiology since the lower level parts of the neurophysiological system do not possess the higher level features. This kind of conception has the advantage that it recognizes the usefulness of a psychological level of description. Further, the idea understanding of complex phenomenon in terms of different explanatory levels and their dynamics applies well to the project of clarifying multidimensional self-consciousness.[23] This project is conducted in this dissertation

---

[23] To recap, this dissertation does not explicitly take any metaphysical stance towards consciousness but aims to stay rather neutral. Entailed by the conceptual basis of the dissertation, the reductionist materialism is considered insufficient account for (self-)consciousness. However, the critique of reductionism does not imply dualism, which considers mental and physical as different substances. Instead, neurophilosophy highlights the crucial role that brain has for consciousness and thus, seems to promote a monistic view. The hard problem of consciousness of is beyond the scope of this dissertation; the position

through phenomenological neurophilosophy, and the relation between experience and neural processes is elaborated on especially in the Chapter 4.

1.4. Summary

Self plays a principal role in one's life and has been a central theme in the philosophy of mind. Throughout the history of philosophy, self has been equated with mind or soul, and philosophers strived for defining a unified self. However, in the twentieth century, the study of self disintegrated into many sciences and narrow notions that grasped only some sense of self. Recently, there have been new multidisciplinary aspirations to integrate the different aspects of self together within a pattern theory of self.

A useful approach to self is to study self-consciousness. Self-consciousness is the way by which we take up with the self, and only by forming a picture of self-consciousness can we answer the more metaphysical questions about the self. Self-consciousness has been carefully studied in the phenomenological tradition, and the detailed phenomenological conception of self is used and applied in this dissertation. According to the phenomenological conception, the essence of self is minimal self-consciousness that refers to the subjective character of experience. Beside this fundamental form of self-consciousness, reflective self-consciousness is the other major form of self-consciousness, referring to the capacity for explicit self-directed thoughts. However, it is still unclear what relations prevail between the two forms of self-consciousness. In order to clarify these relations, this dissertation examines how the two forms of self-consciousness are connected. The dissertation consists of two parts: the first lays conceptual foundations by defining the concepts of minimal and reflective self-consciousness, examining their neural mechanisms, and analyzing their connections. The second part applies the

---

employed here roughly represents emergent physicalism but is also compatible with panpsychism, and leaves the door open for the discussion between different metaphysical theories.

defined concepts to concrete cases of altered states of consciousness: Cotard syndrome, depersonalization, and meditation. These states elicit the features of typical self-consciousness too, giving further support to ideas of the structure of self-consciousness presented in the first part.

The research method here is neurophilosophy, which aims at cooperation between philosophers and (neuro)scientists. Progress in neuroscience is interesting for philosophers because it can teach about the conditions for experience and because nowadays, it is advisable that philosophy of mind be empirically informed. Scientific knowledge of neural mechanisms does not in itself offer solutions to philosophical problems; however, it can provide premises that assist in arriving at right conclusions. On the other hand, the conceptual and theoretical analysis by philosophers can assist in the formulation of empirical theories by offering conceptual clarity and help in developing paradigms for empirical research. Thus, cooperation between philosophers and empirical scientists is mutually beneficial in studying the mind and self. Although the explanatory gap exists, it can be narrowed or at least partly bridged.

2. Shades of minimal self-consciousness

The aim of this chapter is to clarify the concept of minimal self-consciousness (MSC in brief). I start the chapter by considering a debate about the concept of minimal self; roughly, it can be interpreted in either an experiential or embodied sense. The notion of MSC endorsed in this dissertation highlights the experiential sense, which defines MSC as a subjective character of consciousness. However, it is also useful to see the connections between experiential and embodied senses, and I describe their combination with the notions of first-person perspective and subjective perspective. Then, I elaborate the sufficient and necessary features of MSC. The idea is that being an embodied being is necessary but not sufficient for MSC, since there can be embodied systems that lack self-consciousness. In addition, there can be minimally self-conscious experiences that lack embodied features. Concerning the pattern theory, I propose that in a robust sense, MSC involves at least embodied, experiential and affective aspects of self. Instead, in a minimalist sense, MSC refers to subjectivity of experience, which highlights the significance of the experiential aspect in the pattern theory.

Determining the hierarchically lowest level of a phenomenon is a necessary first step for an extensive theory (see, e.g., Metzinger, 2013; Zahavi, 2005). The notion of minimal self is crucial in theories of self since it refers to the most fundamental form of selfhood and determines the minimal requirements for being a self. However, the notion is ambiguous, and there are at least two ways to understand it: as experiential or embodied. On the one hand, the basic form of self is often understood in terms of *experiential subjectivity* or pre-reflective self-consciousness (see, e.g., Strawson, 2000; Zahavi, 2005). According to this view, minimal self is essentially connected to consciousness: subjectivity is seen as structural part of consciousness, and in order to be a self, one must be a conscious being. In short, being a (minimal) self is simply a matter of having a subjective point of view. On the other hand, the minimal conditions for being a self can be understood in terms of

embodied identity that marks a concrete spatial distinction between self and environment (Barandiaran et al., 2009; Legrand & Ruby, 2009). However, if selfhood is understood just in terms of spatial or perspectival bodily identity, it seems that being a self does not require consciousness, and this seems to be in tension with the experiential conception of minimal self. [24]

According to the conception of minimal self in this dissertation, being a self is to have MSC or to be a subject of experience, and being the subject of experience is to have a *first-person perspective* on the world. In more detail, it should be noticed that the meaning of '1PP' importantly encompasses two matters. A 1PP is a perspective: it is a view on the environment from a particular spatial and temporal orientation. It basically refers to the ability to distinguish oneself from the environment and, at the same time, emphasizes the self's embodiment and situatedness in the environment. But more importantly, the perspective is essentially first-personal; the orientation is subjectively experienced from the subject's own point of view. It is given in a unique way to the subject herself. The latter point highlights that although people can be at the same place at the same time, each of them has a different subjective point of view, and their experience can vary accordingly. For instance, two persons in the same place can witness a sudden thunderstorm and share the spatial and temporal perspective of being in the storm, but each have a different subjective point of view on it. One might be scared and terrified while the other enjoys the forces of nature. This shows that a mere perspective cannot define self-consciousness, but rather subjectivity is the decisive feature of MSC.

2.1. Minimal self-consciousness as embodied perspective

Although MSC should not be understood merely as a perspective, the notion of 1PP is useful since it assists in seeing how MSC is essentially embodied. The experiential and embodied features of

---

[24] Cf. also the debate between psychological and biological/physical approaches in the discussion concerning personal identity, see, e.g., Whiting 2002.

self are complementary and tied together; embodiment is necessary for experience to occur, and one experiences through the body. In order to give a full account of MSC, I will consider the embodied reading of MSC in this subchapter, and then complete it with the experiential reading in the rest of this chapter.

The phenomenological conception of MSC highlights that subjectivity is embodied: "As perceivers and agents we are embedded and embodied agents. All perception and action involve a component of bodily self-experience" (Gallagher & Zahavi, 2007, 142). In more detail, phenomenologists argue that the lived body (Leib) is experienced pre-reflectively; it is the background of experiences (see, e.g., Colombetti, 2007; Merleau-Ponty, 1962; Thompson & Zahavi, 2007). In other words, MSC involves a form of non-objective bodily self-awareness, which is "the zero-point" of experiences (Thompson & Zahavi 2007). For example, when I am writing my ideas down in a notebook, my awareness is mainly focused on my thoughts and the effort of finding the right words to express them. However, at the same time, I am pre-reflectively aware of my hand holding a pencil as part of my body, and generally of my body through which the writing and experiencing occurs. Thus, perceptual experiences and action can be considered in part as experiences of my body. In order to clarify this role the body plays in experience, phenomenologists use the distinction between two senses for the concept of 'body' (found in the German language): 'Leib' refers to the lived, feeling, and expressive body, whereas 'Körper' refers to the body as a physical object examined in third-person observation and scientific investigation (Colombetti, 2007; Thompson, 2007; Thompson & Zahavi, 2007).[25]

---

[25] Body can be taken as the object of self-consciousness in both of these senses (see, e.g., Colombetti & Ratcliffe, 2012; Legrand & Ravn, 2009; Shusterman, 2008). The living body (Leib) or a bodily activity can be the object of awareness, such as when an individual focuses on her breathing or touches her arm, they are not *objectified*, meaning "experienced as mere things or processes" (Colombetti & Ratcliffe, 2012, 146). Instead, examples of the lived body (Körper) as the object of awareness include the measurement of one's height, examining the size of one's muscles in the mirror, or observing the color of one's eyes. In these cases, the body is present as physical, "thing-like" features that are examined in the same way as other outer objects.

Embodiment connects self and world. As, for instance, Bermúdez (2007, 463) formulates, perception of the external world has an irreducible first-person component: perception is essentially perspectival and egocentric. Self-consciousness is a contrastive notion: a subject is aware of herself relative to and as distinct from environment. Self as an embodied system can make a distinction between itself and environment (non-self). Further, it is important to notice that the embodiment is not passive but active in exploring the environment; perception and action are fundamentally interwoven (Gibson, 1977). Below the character of embodied 1PP is considered in more detail.

As a first-person *perspective*, MSC first refers to a spatiotemporal self-location (for a definition of MSC that highlights location; see, e.g., Blanke & Metzinger, 2009). The richness of embodied 1PP can be noticed in a simple example. For instance, when I am walking on a seashore at sunset, I can see the cliffs, waves and forest. All these things have a certain location in relation to my body. The body is the source of the perspective or the "center" around which outer objects are placed. Also other senses give spatial information, such as auditory sense telling in which direction the waves are and where seabirds fly. Further, by means of proprioception, I can sense my movements and positions of my body in keeping my balance in walking; I need to adjust my steps to the perceived shape of the rocky shore. Or when I look at a bird in flight in the sky, I adjust my eye movements according to the motion of the bird. Thus, 1PP involves spatial embodiment and even a simple movement in space requires the combination of perception and action. The features of the body also shape experience, for instance, the senses we have determine which kinds of experiences we can have. For instance, humans have developed color vision but we cannot perceive ultraviolet light (as birds can); neither do we have the sense of echolocation that bats have.

Further, all experiences occur in time. MSC refers to "immediate" experience, that is, to the presence of experience right now. In other words, temporality of 1PP does not refer in the first place to 'time' in the sense that at the same time that I am walking at the seashore I would be aware of what year or day of the week

it is, or how this moment is linked to my activities last week or to my future plans. Rather, it refers to time as a constituent or structure of consciousness in which experience manifests itself; while my glance turns from the open sea to the shoreline and further to the forest and then back to the open sea again, all these objects of consciousness have smoothly followed each other in time, connected to previous and subsequent experiences. [26] Subjectivity is a feature of ongoing experience, and thus time as the temporal feature of 1PP is included in MSC.

2.1.1. Enactive agent

In recent philosophy of mind, the embodiment has also been understood in a more extensive sense than the perspective described above. An approach that explicitly points out the more extensive sense is here briefly considered in terms of enactivism, which argues that self-consciousness emerges as embedded in the whole body and in interaction with the environment. Neurophilosophy and enactivism share the idea of multidisciplinary study of mind, and thus, it is mutually useful to be familiarized with each other's ongoing research. This is especially relevant in order to gain a full understanding of self, since selfhood is a major research topic in enactivism too.[27] Enactivism emphasizes the embodied nature of mind and the linkage between self and consciousness. However, the enactivist ideas of embodiment describe a self as an autonomous cognitive system, and it is not exactly clear how the enactivist description is connected to the experiential reading of MSC that is endorsed in this dissertation. In any case, the self-experience presupposes

---

[26] Time-consciousness itself is a major feature of mind and, as such, a debated topic in the philosophy of mind, but it is beyond the scope of this dissertation. However, it is worth noticing that the phenomenological tradition, which has emphasized the role of MSC, has yielded sophisticated accounts of time-consciousness, too, and also studied how MSC and time-consciousness are necessarily connected. See, e.g., Zahavi (2014); Gallagher & Zahavi (2007). For a conception that emphasizes time in MSC, see Kiverstein (2010).

[27] However, the enactivist approach is only briefly outlined here, and this dissertation does not argue for it nor present detailed criticism of it. Enactivism earns involvement in this discussion here since it offers an extensive viewpoint on self, but it is not necessary to consider it in more detail in order to carry out the inquiries of this dissertation.

embodiment, and the enactivist considerations can assist in showing how being a self is necessary for consciousness even if one would not commit to claims of enactivism in a strong sense.

Generally, enactivism is presented as a paradigm for cognitive science that provides an alternative to traditional computational cognitive science (see Stewart et al., 2010; Varela et al., 1991).[28] Instead of relying on the traditional computer metaphor of mind, enactivism argues for the biological basis[29] of cognition and starts the analysis of mind from the organization of a living system. In enactivism, cognition is not conceptualized in terms of computational information processing and internal representations, but as essentially embodied activity, which is based on the autonomous identity of the cognitive system and its sensorimotor interaction with the environment. In brief, enactivism argues that mind is bound to body and body is bound to environment; cognition and consciousness arise as a compound of these elements.

The enactive approach endeavors to describe and unify several related ideas under one heading (below combined from Colombetti & Thompson, 2008; Di Paolo & Thompson, 2014; Thompson & Stapleton, 2009; Thompson et al., 2005). The first idea is that "living beings are autonomous agents that actively generate and maintain their identities and thereby define their own cognitive domains" (Colombetti & Thompson, 2008, 55). This idea introduces the key concept of an autonomous system,[30] a system

---

[28] Generally, enactivism or the enactive approach originated in the book *The Embodied Mind* (Varela, Thompson & Rosch, 1991). Its background is e.g., in the biological theory of autopoiesis (Maturana & Varela, 1980) and phenomenological philosophy (e.g., Merleau-Ponty, 1962). Nowadays, there are different versions of enactivism (for an overview see, e.g., Kiverstein & Clark, 2009; Ward et al., 2017).

[29] The biological basis of cognition is one central idea of enactivism (originally presented by Maturana & Varela 1980) and often expressed in terms of deep continuity between mind and life. This continuity is described, e.g., by Thompson (2007, 128): "Life and mind share a set of basic organizational properties, and the organizational properties distinctive of mind are an enriched version of those fundamental to life." However, the emphasis of biology varies in different versions of enactivism.

[30] Enactivists define an autonomous system as "an operationally-closed and precarious system" (Di Paulo & Thompson, 2008, 69). Operational closure means a specific kind of organization in which all constituent processes of the system are conditioned by some other

that is literally autonomous in being self-enabling by sustaining itself in time partially due to the activity of its own constituent processes. This kind of autonomous system does not process preexisting information "out there" but instead brings forth or enacts information that is significant from its own perspective in continuous reciprocal interactions with its environment.

The second idea is that the nervous system is an autonomous system, which "actively generates and maintains its own coherent and meaningful patterns of activity according to its operation as a circular and reentrant sensorimotor network of interacting neurons" (Colombetti & Thompson, 2008, 55-56). In addition to autonomy, this kind of complex system is cognitive, which entails that the interaction between the cognitive system and environment is essentially asymmetrical: the system is the active source of interaction. That is, the system is an agent that regulates the coupling and establishes a specific perspective from which the coupling with the world acquires normative status. With this status, the system creates its own norms, which can range from basic vital requirements (the self-maintenance of the agent's biological infrastructure) to complex psychological and cultural norms.[31]

The importance of interaction between an autonomous system and environment entails the third idea of enactivism, namely, that cognition is a form of embodied action (Colombetti & Thompson, 2008, 56). Cognitive structures emerge from sensorimotor patterns

---

process in the system, i.e., the enabling conditions of the constituent processes are tied together. Precariousness means that a process of the operationally closed system runs down or stops if it is deprived of the enabling network of relations that constitute the system.

[31] This capacity for asymmetrical interaction with the environment is typically ascribed to the condition of adaptivity, defined as "the ability to regulate the operationally closed processes in relation to conditions registered as improving or deteriorating, viable or unviable." Further, the adaptive and active regulation is called sense-making, which is the other key concept in enactivism (see, e.g., Di Paolo & Thompson, 2014; Thompson & Stapleton, 2009). While a cognitive system evaluates its surroundings by its own norms, the surroundings acquire meaning for the system; the system makes sense of its environment. Sense making is the interactional and relational side of autonomy, and the whole enactivist concept of cognition has been defined as "sense-making in interaction: the regulation of coupling with respect to norms established by the self-constituted identity that gives rise to such regulation in order to converse itself" (Di Paolo, 2009). However, these concepts of autonomous adaptivity and sense-making are complex and cannot (and need not) be considered here more closely.

in perception and action. The interaction between cognitive agent and environment modulates the formation of neural patterns—and in turn, neural activity informs sensorimotor interaction. The embodied agent is a self-organized autonomous system that is linked with the environment by its action and creates meaning.

The fourth idea is that experience is "central to any understanding of the mind and needs to be investigated in a careful, phenomenological manner" (Colombetti & Thompson, 2008, 56). That is, the enactive approach asserts that cognition and consciousness are linked together; thus, cognitive science and phenomenology should be pursued as complementary and mutually informing endeavors. In other words, enactivism takes cognition to be embodied in both a structural and a phenomenological sense. The structural embodiment refers to the idea that neural, bodily, and environmental processes are subsumed in cognition.[32] The phenomenal embodiment, in turn, highlights cognition as a subjective experience that involves bodily self-awareness and interaction with the environment. To recap, the enactive approach argues that the human mind "is not reducible to structures inside the head" Colombetti & Thompson, 2008, 56).

---

[32] This is related also to the idea of "extended mind," according to which factors of the external environment can play a necessary, constitutive role in cognition (Clark & Chalmers, 1998; Clark, 2003). According to the original extended mind thesis (Clark & Chalmers, 1998), if a part of the world functions in the same way as a part of a cognitive process in the head, then the part of the world is part of the cognitive process. In more enactivist terms: a cognitive system may incorporate external elements that "function transparently in the body's sense-making interactions with the environment" (Thompson & Stapleton, 2009). Nowadays, the ideas of embodied and extended mind are many times connected together in the "4E"-framework of cognition (Es for embodied, embedded, extended, enacted; see, e.g., Menary, 2010a, 2010b; Newen et al., 2018; Ward & Stapleton, 2012).

A principal critique of the extended mind thesis invokes the causal-constitutive fallacy (Adams & Aizawa, 2001, 2010; 2010), pointing out that external elements can have causal influences on cognition without being constitutive parts of cognition. This discussion might be applied also to the other Es of the 4E framework, e.g., it could be debated whether interaction with environment is constitutive for cognition, or merely causal. However, that debate cannot be settled here, but the embodied basis of self is acknowledged to be significant in any case. Neural and bodily processes especially are taken to be constitutive for (self-)consciousness in the strong sense of providing a necessary material basis for it, interaction with the environment is considered significant at least in the causal sense.

Instead, an experience is created by the interaction of the brain, the body, and the world.

Self is a significant theme in enactivism since the key concept of an autonomous system can be associated with the concept of self. That is, self is the active agent that maintains and generates itself, forms norms for itself in regulating its interaction with the environment, and (at least in humans) can also have complex forms of self-reflection (for one sophisticated enactivist notion of self, see Thompson, 2014). A noteworthy theoretical tool for describing self in enactivists and scientific terms is the concept of a *self-specifying system*. This concept is directed at self as an agent or subject of experience, brings the enactivist idea of an autonomous system to the general discussion of being a self, and seems to come close to the notion of minimal self.[33]

2.1.2. Self-specific perspective

The motivation to develop the notion of 'self-specificity' was the insight by Legrand and Ruby (2009) and Christoff et al. (2011) that empirical research on self has been defective in targeting self only partially (see Sec. 4.1.2.). The 'self-specific' is defined as a "feature that is 1) exclusive: characterizes oneself and no one else, and 2) non-contingent: changing or losing it entails changing or losing the distinction between self and non-self" (Christoff et al., 2011, 104). Accordingly, a self-specifying feature specifies the self as subject and agent by implementing a functional self–non-self-distinction. Thus, it can be noted that as such the notion aims to capture a functional self–non-self-distinction and can be interpreted to target minimal self only in the sense of being merely an embodied system, not being self-conscious in the experiential sense.

The notion of self-specificity aims to capture the sense of self that is present in experience even when the self is not taken as the object or content of consciousness. Legrand and Ruby (2009, 272)

---

[33] The term 'self-specifying system' is used in enactivist literature e.g., by Thompson (2014) who refers to the studies of Legrand & Ruby (2009) and Christoff et al. (2011) that are discussed in the next section and also in Sec. 4.1.2. Thompson (2014, 326) defines a self-specifying system as "a collection of processes that mutually specify each other and thereby constitute the system as a self-perpetuating whole in relation to the wider environment."

underline that the conception of the self-as-content in reflective self-consciousness is only partial, since any representation of oneself is not self-specific. First, many general contents of self-consciousness can be attributed either to the self or to others, and thus, they do not describe the self exclusively. For example, this is the case for personality traits or actions: both oneself and others can be perfectionistic, and both oneself and others can lift a hand on request. Since these types of the contents of consciousness are potentially owned both by oneself and by others, they do not meet the criterion of exclusivity and cannot be considered intrinsically self-specific. Second, some contents of self-consciousness are only contingently related to oneself. This includes, for instance, one's unique facial features: they would not allow the specification of the self as such, since the self–non-self distinction can be made even if these contents change. One obviously does not cease to be oneself by merely changing one's facial features, and thus, such contents fail to meet the criterion of noncontingency. Legrand and Ruby note that even somatosensory contents characterize the self only contingently; although these contents are exclusively related to the self, the distinction between a given self and non-self does not collapse as soon as they change, are lost, or are misattributed.[34]

Legrand and Ruby (2009) further clarify that, instead of being intrinsically self-related, a perceptual content is functionally self-related. A content can be processed as self-related, but "contents as such cannot constitute the self since they presuppose a self-specific process determining a functional distinction between self and non-self" (Legrand and Ruby, 2009, 273). That is, the primary self-specific differentiation only allows the secondary differentiation between self-related and non-self-related contents. Thus, it would be a mistake to equate the self with self-related contents: a focus in non-self-specific contents disregards the processes that make a particular content self-related in the first place.

Legrand and Ruby (2009, 274) emphasize that self-related contents presuppose more basic processes and find perception

---

[34] For instance, the self–non-self distinction remains relevant for schizophrenic patients even if they misattribute their intentional actions to others.

particularly relevant. Perception is always related to the self through being grounded in the perspective of the perceiving subject, even when perception does not involve a representation of the self (an idea that stems from phenomenological philosophy [Legrand, 2007; Zahavi, 2005b] and it is also present in enactivism [Thompson, 2007, 2014]). Furthermore, the perspective of the subject can be characterized by acknowledging the difference between contents and the perspective of perception. The content determines what is perceived, whereas the perspective determines who perceives the content, how, and from where. This implies that perception encompasses more than the perceived contents. Crucially, the perceiving self remains present throughout the contents of perceptions, which are processed through different sensory modalities and can be misrepresented. In fact, "a perspective grounds every perception and representation held by any given subject" (Legrand and Ruby, 2009, 274). For instance, the simple experience of tasting a strawberry involves a content (a strawberry instead of a lemon), a mode of presentation (tasting instead of seeing the strawberry), and a perspective (one's experience of tasting the strawberry). The specific perspective makes the tasting of the strawberry one's own perception; that is, the tasting is experienced from the individual's perspective.

Thus, the subjective perspective is self-specific; it fulfills both criteria for self-specificity (Legrand and Ruby, 2009, 274). First, the condition of being exclusive to a self is met: a perspective discerns the self from non-self. Although two people can perceive the sweet taste of strawberry, their perceptions are assessed from particular subjective perspectives that differ systematically, and neither perception can be reduced to the other. Second, a perspective characterizes a self noncontingently: a change of perspective implies a change of a self. One cannot have representations that would be grounded in another person's perspective. Surely, one can consider a perspective of another subject and hence adopt a third-person perspective, such as when thinking about what can be seen from another side of a room or what kind of feelings another person is experiencing. However, one inevitably does this thinking from one's own perspective, and one cannot entertain a perspective

that would be completely detached from one's own. Legrand and Ruby (2009, 274) expressed this as follows: "My perceptions and experiences are anchored in my perspective, and by virtue of this, they are mine rather than someone else's or nobody's."

Overall, Legrand and Ruby (2009) argue that being a self involves a self-specific perspective. This perspective is a fundamental self-specifying process: self as the subject of representation is present in all representations, even in those that do not have self as their object. The self–non-self distinction is constituted by the perspective and it is primary to the distinction between self- and others-related contents. The latter distinction is secondary since it is determined by the perspective.

The description of self-specificity assists in giving a detailed picture of minimal self, since it indicates how self is present in all perception and action and how minimal self is needed for cognition. However, since self-specificity by definition is a functional description, it might be limited to the embodied reading of minimal self only, not necessary involving self-consciousness (see also Sec. 4.2.3.). That is, there might be self-specifying systems that are not conscious, and thus, self-specificity is not sufficient for self-consciousness. Yet, the characterization of self-specific perspective has a clear resemblance to the notion of minimal self-consciousness as a constitutive feature of experience. The notion of self-specific perspective targets the most basic form of self that is present whenever there is experience, is ascribed to the how of experience instead of the what, and distinguishes one subjectivity from another. Legrand and Ruby's terminological choices, such as 'subjective' perspective, also highlight that the perspective involves subjectivity as something-it-is-like-(for-me-)ness. The scientific study of self-specific processes is further discussed in Secs. 4.1.2 and 4.2.3.

In the rest of this dissertation, I will use a concept of *subjective perspective* in describing MSC. This use could be considered an experience involving elaboration of Legrand and Ruby's (2009) notion of self-specificity, and in this elaborated sense the subjectivity is considered primary. In other words, this concept explicitly connects embodied and experiential readings of self. Following Zahavi (and others in the next section), it takes the

experiential or subjective dimension to be necessary for self and recognizes that typical self-experience involves embodied and perspectival features of the living body (Leib). However, it is notable that that the conception here holds the *subjective* as a necessary feature of minimal self-consciousness, whereas the *perspective* is not a necessary feature of self-consciousness, and minimal self-consciousness should not be understood only in terms of perspective. This point is elaborated in the next section and discussed again in Sec. 2.4. in terms of minimal and robust reading of MSC.

2.2. Minimal self-consciousness as integral feature of experience

According to the experiential reading of minimal self, the lowest level of selfhood is defined in terms of minimal self-consciousness. That is, minimal self is connected to consciousness. Many times, the idea is firstly described by distinguishing MSC from more cognitively demanding reflective self-consciousness: MSC does not refer to reflection of oneself but is an integral part of experience. Thus, MSC is self-consciousness in a specific thin sense characterized, for instance, in the following quotations:

> "When we are thinking about x, the mind is focused on x, not on our thinking of x. Nevertheless, the process of thinking about x carries with it a nonreflective self-awareness." (Goldman, 1970, 96)

> "a consciousness of oneself as an immediate subject of experience, unextended in time" (Gallagher, 2000, 15)

> "This self-consciousness we ought to consider not as a new consciousness, but as the only mode of existence which is possible for a consciousness of something." (Sartre, 1956, liv)

> "minimal self-consciousness, namely, the feeling that your consciousness belongs to you, that you are the subject of your awareness" (Thompson, 2014, 63)

"At this level, the self is the subject of consciousness, experienced as the subjectivity of consciousness...pre-reflexive self-consciousness is fundamental, in the sense that it is the foundation of any other form of consciousness." (Legrand, 2005, 17-18)

"It is impossible to think or experience something consciously without thinking or experiencing it self-consciously" (Kriegel, 2004, 200)

These characterizations highlight that MSC is always present in immediate experience. The term 'minimal self-consciousness' is used here since it describes well that the referent form of self-consciousness is thin in being minimal. Although this dissertation draws heavily on Zahavi's ideas on self, instead of using 'pre-reflective self-consciousness' or any other of Zahavi's notions, the term 'minimal self-consciousness' is preferred in order to employ a more neutral notion that is not too tied to phenomenology but can be easily used among representatives of other traditions too. In addition, the preference for 'minimal' derives from Gallagher's (2000) notion of 'minimal self' that refers to self at the low level but still involves several features (see Sec. 2.4.1.).

2.2.1. Subjective character of consciousness

One way to encapsulate the idea of MSC is to state that *all conscious states' phenomenal character involves minimal self-consciousness as an experiential constituent*. This formulation is from Zahavi and Kriegel's (2016) paper "For-me-ness: What it is and what it is not," and it is viewed below, point by point, in order to clarify what is meant by MSC. It is noteworthy that Zahavi and Kriegel's (2016) notion of for-me-ness is a combination or shared idea of different philosophical traditions and as such applies to the ethos of multidisciplinary cooperation in this dissertation.[35]

---

[35] Whereas Zahavi's notion of 'pre-reflective self-consciousness' of 'experiential self' is deeply-rooted in the phenomenological tradition, Kriegel's notion of 'intransitive self-consciousness' represents modern analytic philosophy. Although Zahavi and Kriegel (2016) remark that their theories differ from each other in details, they still usefully frame a common ground for various notions of minimal self-consciousness. Zahavi (2014, 16–17) describes the

- Minimal self-consciousness pertains in the first instance not to *what* is experienced but to *how* it is experienced.

MSC refers to the subjectivity of consciousness, instead of objects of consciousness. Minimal self-consciousness is not a quality like purple, sweet or soft. Although it is sometimes referred as 'ownership', it is important to note that it does not mean 'owning' in a similar way to how we possess external objects such as cars, clothes, or apartments. Minimal self-consciousness does not refer to a specific experiential content, to a specific *what*; instead it refers to the *how* of experience, to the first-personal presence of experience. That is, minimal self-consciousness refers to the fact that experiences I am living through are given differently to me than to anybody else.

- Minimal self-consciousness is an invariant dimension of all phenomenal character.

We have a great diversity of experiences, but they all are characterized by the same fundamental first-personal self-givenness. All experiences fundamentally involve a dimension of mineness or for-me-ness. For instance, if I compare my current perception of a computer screen, my remembering of my last holiday on a sunny beach, and my anticipating of the music concert on Saturday, they all have in common that they are my experiences, experienced by me, occurring in my stream of consciousness. These experiences involve 'invariant' MSC in the sense that MSC is present in all of

---

differences between his and Kriegel's accounts by pointing out that Kriegel explicitly rejects the notion of a non-objectifying self-consciousness, which Zahavi endorses. Instead, Kriegel argues for a (one-level) self-representationalist theory. However, Kriegel admits that self-representation per se is insufficient for subjectivity. Thus, Kriegel answers to counter-examples that stem from the existence of self-referring sentences and the possibility of functionally equivalent zombies. Kriegel ends up proposing the view that only non-derivative, specific, essential self-representation is sufficient for consciousness (Kriegel, 2009, 162). When articulating these requirements for relevant self-representation, Kriegel proposes that the subject's epistemic relations to her conscious states are special (Kriegel, 2009, 107-8) and admits that the relevant type of self-representation constitutes a very unusual form of object-awareness. Because Kriegel needs to make all these qualifications, Zahavi notes that one might wonder whether the difference between this kind of highly unusual form of object-awareness and non-objectifying form of awareness is all that substantial after all.

them. However, MSC is not 'invariant' in the sense that it could not involve any changes (see Sec. 2.4.).

The point entails that MSC is a constitutive feature of consciousness. As described in Section 1.3.2., the mainstream way to describe consciousness nowadays is to invoke the subjective character of experience (e.g., Block, 1995; Nagel, 1974). I am familiar with my own experiences from a 1PP, whereas the experiences of others, as well as scientific descriptions of my experiences, are assessed from a 3PP. The idea is that my experiences immediately have a certain feeling within them: a mental state is conscious iff there is something it is like for the subject to be in it. Since conscious states are characterized by a subjective mode of experience, it can be said that all experiences are implicitly characterized by a certain for-me-ness or mineness—that is MSC.

That is, the idea of MSC elaborates the notion of phenomenal consciousness by arguing that there is more to experience than its contents. In Kriegel's (2006, 2009; see also Levine, 2001) words, the *phenomenal character of consciousness* can be construed as a combination of two characters: qualitative and subjective character. A phenomenally conscious state's *qualitative character* (blueness of the sky, taste of coffee) refers to the content of consciousness and makes it *the* phenomenally conscious state it is. Instead, a phenomenally conscious state's *subjective character* (for-me-ness, minimal self-consciousness) makes it phenomenally conscious *in the first place*. This way the subjective character remains invariant across changing qualitative characters. Kriegel (2006) notes that the same kind of idea that he terms 'qualitative character' has been argued by Rosenthal (1991) in favor of distinguishing consciousness from what Rosenthal calls 'sensory quality'. Rosenthal (2010, 2005) also highlights that unconscious mental states can have sensory quality; for instance, we make perceptual discriminations independent of consciousness. According to Rosenthal (2005, 139): "Sensory qualities will occur even when sensory states are not conscious. But when the states with sensory qualities are conscious, there will be something it is like to be in those states, and sensory qualities will be properties in virtue of which that is so." In Krieger's

terms, mere content or qualitative character does not render a mental state conscious, but subjective character is decisive for phenomenal consciousness.[36] This is explicated in the next point.

- Minimal self-consciousness distinguishes conscious experiences that present something to someone from non-conscious representation (e.g., blindsight[37]) of the same objects.

A mental state that lacks minimal self-consciousness is a non-conscious state. This indicates the fundamental role that minimal self-consciousness has for phenomenal consciousness: it turns a non-conscious mental state into a conscious one. Non-conscious mental states can take place *in* me, and one can "host" them in an impersonal sort of way, without being aware of them. Instead, conscious states do not just take place in me, but they are something like *for* me. In more detail, Zahavi and Kriegel (2016; Kriegel 2009; Zahavi 2014) call this a non-deflationary interpretation of MSC. According to a deflationary interpretation, for-me-ness is a non-experiential aspect of mental life and refers simply to the occurrence of experience in someone (a 'me'). By contrast, a non-deflationary interpretation considers for-me-ness as an experiential aspect of mental life. According to this view, the claim that an experience is "for me" says something more than the claim that an experience occurs "in me." It is to state a

---

[36] According to Kriegel (2006), Block (who launched the concept of phenomenal consciousness in 1995) would admit that phenomenal consciousness involves this kind of for-me-ness. However, it should be noted that the distinction between qualitative and subjective characters is not a mainstream conception of phenomenal character in consciousness studies. Instead, either 'qualitative character' or 'subjectivity' alone has been used in referring to the phenomenal character of consciousness (see, e.g., Van Gulick, 2014). More typically than drawing the distinction (made by Kriegel, Levine, Rosenthal), the qualitative character (or 'qualia') is simply considered subjective, and it is said that unconscious mental states are not qualitative. For instance, in studying self-consciousness and the unity of consciousness, Bayne (2004, 221) sees that "personal-level accounts invoke features of the content or character of conscious states to explain why they are (or are not) unified." However, the notion of MSC argues that it is insufficient to describe (self-)consciousness only in terms of contents of consciousness. See also Chapter 5 here.

[37] Blindsight refers to cases in which cortically blind people (their blindness results from lesions in the primary visual cortex) respond to visual stimuli that they do not consciously see (e.g., Weiskrantz, 1986).

phenomenological fact in addition to a metaphysical fact. Without minimal self-consciousness, there could be no experience of qualitative characters.

- Minimal self-consciousness is not just a geometrical feature of perceptual experience.

The geometrical feature refers to a mere visuospatial perspective of reality[38], whereas MSC essentially involves subjectivity. This subjectivity can be viewed in an epistemic asymmetry: experiences are characterized by a subjective presence that makes them first-personal to the experiencer but inaccessible in the same way to a plurality of subjects. The epistemic asymmetry is grounded in the ontology of experience: My experiences are present to me in a way that is in principle unavailable to others.[39] Although I and someone else would perceive something from the exact same geometrical perspective, our experiences would still be different since we have distinct subjectivities or we are different selves. And even if it turned out that it was possible to know what someone else is thinking while she is in a brain scan, or even to watch the contents of her thoughts on a screen, we still would not be able to have her 1PP with her ongoing subjective feelings. We would only have access to the contents and the mere geometrical features of her experience. Or the geometrical features could be associated with the contents of experience, referring to the qualitative character of consciousness instead of subjective character.

---

[38] For instance, Blanke & Metzinger (2009) describe a weak first-person perspective as a "purely geometrical feature" of the visuospatial presentation of reality. Zahavi & Kriegel (2016) argue that this kind of notion has nothing to do with subjectivity or for-me-ness and that it would be better not to use the label of 'first-person perspective' for this geometrical feature.

[39] The primacy of subjective 1PP and its uniqueness for one subject does not mean that the mental life of others would be fully incomprehensible or out of our reach. Rather, the phenomenologists highlight our capacities for empathy and intersubjectivity; we can access other's mental states by "direct perception," further understand them in many forms of communication, and also others shape our self-understanding (see, e.g., Zahavi, 2014). However, the fact that we are intersubjectively attuned does not exclude the fact that we cannot feel anyone else's subjectivity as our own. This exactly is what is meant by the subjectivity of experience (as put in Chapter 1, appealing, e.g., to Nagel, 1974).

That is, talk about a plain 'perspective' only in a geometrical sense does not entail consciousness. We can say that a non-conscious robot equipped with a video camera records a view from a certain perspective—but without experiencing anything. Or we can say that implicitly perceived visual information of the outer world manifests itself in a perspective of a subject—but the subject is not conscious of the information. However, when it comes to self-consciousness, it is essential that the perspective is first-personal. For instance, in visual perception, we are not only receiving and registering information about the outer world (like robots might do), but there is something it is like to perceive the view.

- Minimal self-consciousness is not a detachable self quale, it cannot occur on its own.

The point seems to be meant to highlight the previous points that in the first place, MSC does not refer to contents of consciousness, albeit it is intertwined with them (and rather the same point was also made using the terms subjective and qualitative character of consciousness). However, Zahavi and Kriegel (2016) make a further claim, according to which MSC does not occur without contents of consciousness. Yet, this particular claim might also be disputed, for instance, by meditation studies that consider the meditative state as a state in which the typical subject-object distinction disappears. In meditation, a subject ceases to be occupied with the objects of consciousness in order to become aware of the self-presence, that is, the subjectivity of experiencing itself (Fasching, 2008). From this point of view, it can be proposed that it is possible to experience "pure subjectivity" without perspectival features of experience or specific contents of consciousness (see Chapter 8).

- Minimal self-consciousness is a minimum point of self-awareness.
- Minimal self-consciousness does not necessarily involve a capacity to think of oneself as oneself, be aware of one's states as one's own states, or any such cognitively demanding capacities (Zahavi & Kriegel, 2016, 51).

MSC is present throughout of the spectrum of experiences, including experiences that lack reflective self-consciousness. Since minimal self-consciousness is considered to be a part of consciousness, it enables one to also attribute consciousness to creatures that lack language and higher cognitive capacities. Hence it is natural to see, for instance, infants and animals as conscious beings.

- Minimal self-consciousness is the categorical basis of our capacity for first-person thought, which explains why we can usually (if in possession of the requisite conceptual skills) report on our experiences immediately and effortlessly (Zahavi & Kriegel 2016, 51).

As the previous point exhibited, MSC is the most fundamental form of self-consciousness: it can occur in the absence of reflective self-consciousness. However, the same does not hold for the other way around: reflective self-consciousness always includes minimal self-consciousness and is based on it. Minimal self-consciousness is first-order consciousness, whereas reflective self-consciousness is second-order consciousness in which one "observes" the first order mental states. Thus, minimal self-consciousness is a necessary prerequisite for the reflective type to occur. And further, every time one is reflectively self-conscious of herself, she is also minimally self-conscious at the same time (more about the connection between minimal and reflective self-consciousness in Chapter 5).

2.3. Comparison to Metzinger's alternative conception of minimal self-consciousness

In order to clarify the phenomenological concept of MSC more, I briefly compare it with an alternative conception of minimal phenomenal selfhood (MPS in brief) presented by Thomas Metzinger (e.g., Metzinger, 2003, 2013, 2020; Blanke & Metzinger, 2009).[40] The comparison here is rough but applies to the illustrative

---

[40] Zahavi has defended the concept of MSC against many kinds of critiques in a number of publications (e.g., Zahavi, 2010c, 2014, 2017a, 2018), and for the sake of brevity I do not

function. Gallagher and Daly (2018) summarize Metzinger's concept of self as "a mere phenomenal image produced by neural representations," and it can be contrasted with the phenomenological (and enactivist) concept of self in many ways. Below, I consider the ideas of perspective, content of a self-model, no-self view, selfless experience and primacy of experience versus brains in the account of self.

Metzinger describes MSC with the concept of MPS, that is, the simplest form of self-consciousness. Metzinger (2013) describes MPS as "the central enabling condition for having a subjective, consciously experienced 1PP" and defines it as "Self-identification as *transparent spatiotemporal self-location*," which at least seems to present a spatiotemporal definition of MSC that phenomenologists deny (Sec. 2.2.). I consider this first, since the analysis can reveal that this dispute about experience between notions of MSC and MPS is more terminological than substantial. Below, I consider Jennifer Windt's (2010, 2015) notion of MSC, since Metzinger endorses it and Windt is more thorough in her analysis.

Windt (2010, 213) defines MPS in brief as "an immersive and partially transparent spatiotemporal reference frame." According to Windt, experiences are organized around a sense of spatiotemporal self-location that involves the experience of occupying a space, the experience of "now", and the experience of duration. Windt (2010, 4) remarks that there is not a distinction between a spatiotemporal 1PP and a sense of spatiotemporal self-location: "both refer to the phenomenological property of being located at (and relative to) a certain point in space at a certain point in time." Thus, Windt notices, a minimal form of self-experience does not require sense of agency, a multisensory experience of "owning" a body, nor a visual 1PP. MPS is constituted simply by the sense of immersion or of (unstable) location in a spatiotemporal frame of reference (Windt, 2010, 304–6). In other words, Windt considers spatiotemporal situatedness to be necessary for being a self (2010, 2015).

---

consider that wider discussion in more detail here. For Zahavi's answer and critique of so-called anonymity theories of consciousness, see Sec. 5.2.1.

Windt sometimes (e.g., 2010, 304) says that she uses the term '1PP' in the sense of a purely spatiotemporal 1PP. However, according to Zahavi's notions, MSC is *not* a mere geometrical feature of a model of reality but it is essentially subjective; the perspective is experienced from the subject's own point of view. Nevertheless, a closer look seems to reveal that actually Windt's conception is rather compatible with Zahavi's ideas: in ascribing spatiotemporal situatedness to be necessary for selfhood, Windt takes this situatedness to be experienced. That is, Windt explicitly acknowledges the deep link between conscious experience and phenomenal selfhood. As she (2015, 561) writes: "The ISTH [The Immersive Spatiotemporal Hallucination model of dreaming that Windt advocates] model tries to do justice to the central intuition that conscious experience ... is tied to the presence of at least a minimal phenomenal self...." Windt also explicitly accepts Zahavi's idea that the self is the very subjectivity of experience. Further, Windt considers that Zahavi, Strawson and Metzinger share the idea of essentially phenomenal self: "Despite the differences between these three authors, they converge on a single point: conscious experience and at least a minimal form of phenomenal selfhood or subjectivity are inextricably linked..." (2015, 558–60). Clues for the subjectivity of minimal selfhood are already obvious in Windt's characterizations of MPS with terms that involve experience. For instance, in "a sense of spatiotemporal self-location" or "the sense of immersion or of (unstable) location," Windt uses the term "sense," which can be considered to include experience. Windt's point is exactly that the spatiotemporal self-location is experienced; 1PP feels like something to someone and the other way around, when there is something it is like, there is also a 1PP. This perspective is not just a geometrical one, but a subjective self-location and as such the requirement for minimal selfhood.

Thus, according to this kind reading of MPS, it seems rather compatible with the Zahavian notion of MSC. The essence of minimal self-consciousness is subjectivity, and the subjectivity manifests itself in a spatial and temporal frame of experience. It is remarkable that although these philosophers come from different

research traditions (Zahavi from phenomenology and Metzinger and Windt from a more empirically oriented analytic philosophy) and use different notions, they share the idea of minimal self as experience-involving. This is worth noticing since it gives the opportunity to connect different research lines and gives reasons for the concept of MSC from multiple traditions.

The second dispute between the notions of MSC and MPS concerns the emphasis of the phenomenological primacy of experience of self versus representations and models of self. A clear difference between Zahavi and Metzinger is that Metzinger prefers terminology of self-models while Zahavi typically refers to the (self-)experience itself. According to Metzinger (2013), MPS includes "a conscious self-representation that is not experienced *as* a representation." However, it is not immediately clear what Metzinger exactly means with the term "self-representation." Yet, representation might be understood to refer to a content of representation, and this sense is contradictory with the Zahavian notion of minimal self-consciousness that is understood as a manner of experiencing, which is present in experience not as a content but in experiencing itself, regardless of the content of consciousness.

The third, maybe most obvious and discussed, dispute between Zahavi and Metzinger is the no-self doctrine that Metzinger advocates and Zahavi denies (e.g., Zahavi 2005, 2014; see also Legrand, 2005). Metzinger argues that whenever one thinks about herself, she finds "models," that is, images and memories in which she is the protagonist. However, the existence of these self-models does not entail the existence of a self as an endurable soul-like entity. Thus, Metzinger points out that self, understood as this kind of entity, has an illusory character. From this Metzinger draws the conclusion that self does not exist. By contrast, Zahavi argues for the existence of a real experiential and embodied self. Although self is not a content of self-models, self does exist as an embodied subject of experience that is capable of self-consciousness and does have models of itself.

The fourth dispute concerns selfless experience that is unfeasible according to the phenomenological notion of MSC, whereas Metzinger seems to think it possible. This is not clear in

the notion of MPS, but in his recent work Metzinger (2020) also uses the notion of 'minimal phenomenal experience' (MPE in brief) in describing the specific phenomenology of "pure consciousness" in meditation (see Sec. 8.1.1.1.). According to Metzinger (2020), MPE has no personal-level self-as-subject but is non-egoic self-modeling, that is, selfless and not tied to an individual first-person perspective. Thus, Metzinger seems to claim that there can be experiences without MSC. However, according to the concept of MSC, selfhood is always present in experience, and minimal phenomenal experience would be minimal self-conscious experience at the same time. The point in the concept of MSC is that it is not exhaustively described as geometrical perspective but should be understood as subjectivity of experience. That is, the concept of MSC implies that there cannot be "selfless"[41] experience in this minimal sense.

The fifth dispute between Zahavi's and Metzinger's views of minimal self-consciousness concerns the primacy of experience. Both consider a multidisciplinary perspective important in the study of self, but Metzinger emphasizes the empirical science more strongly. In other words, it seems that Metzinger does not see self-experience as primary but takes the empirical neural level to be decisive in understanding selfhood. Contrary to Zahavi, Metzinger might consider neural realization of self more pivotal. For instance, Metzinger (2013, 2) states, "I believe that the 1PP can be naturalized, because it is ... a highly specific representational *format* creating an internal mode of presenting knowledge and information in the brain of a conscious organism...." This characterization points out a useful empirical approach to self, but, according to the experiential notion of MSC, the empirical side should not be taken as the only or prime sense of self. That is,

---

[41] However, it can be noted that although the concept of MSC implies that that there cannot be experience that lacks minimal selfhood, the meditative experience can be characterized "selfless" in the sense that it is unselfish: it does not seek one's own gain and is not motivated by concern for oneself but for others (Fasching 2008). Further, the meditative state might be called "egoless," i.e., it is consciousness without a self-identification with a certain self-model or direct interest for the self. The notion of selflessness is discussed in eastern philosophies and meditation research; see Chapter 8 and, e.g., Siderits et al., 2010.

although it is beneficial to look at the neural correlates of self-consciousness, self is not "in" the brain, and the study of self starts from self-experience, acknowledging also the embodied and embedded features of self. Altogether, selfhood is not comprehensively described referring only to geometrical perspective or contents of self-models and their neural components.

2.4. Elaborating the notion of minimal self-consciousness

After comparing several views on MSC and illustrating the phenomenological conception of it, I want to elaborate the concept of MSC by carrying through novel clarifications for it in this subchapter. One point in Zahavi and Kriegel's (2016) description of MSC (Sec. 2.2.1.) was that "minimal self-consciousness is an invariant dimension of all phenomenal character." This point is meant to distinguish MSC from contents of consciousness and to highlight the for-me side of experience. However, I want to emphasize that the point does not imply that MSC would be fully invariant, but it can manifest itself in numerous ways. MSC refers to subjective features of experience that are not the content of consciousness but manifested in the manner of experiencing. For instance, the same content of consciousness or the whole world can feel very different depending on the subject's affective background—for instance, whether one is undergoing in her life an intense episode of grief or falling in love. Both cases involve bodily changes and changes in the way one faces the environment. In the first case, one's movements can be heavy, the world can feel distant and she can have difficulties focusing on everyday communication. In the latter case, one's steps are light, the environment can look more beautiful than before, and she is eager to communicate with her partner. Although MSC is the most elementary form of self-consciousness, it is not an invariant feature of experience but can be manifested in numerous ways.

In recognizing the variation in MSC, Zahavi draws a distinction between two different phenomenological claims on MSC: a minimalist one and a more robust one (this formulation is from

Zahavi, 2014, 41; for the variance of MSC, see also, e.g., Zahavi, 1999). According to the *minimalist reading*, the MSC "simply refer[s] to the subjectivity of experience, to the fact that the experiences are pre-reflectively self-conscious and thereby present in a distinctly subjective manner" that is not available to anybody else. Zahavi considers that this feature is always preserved in experience. However, Zahavi points out that according to a slightly more *robust reading*, the MSC can refer to "a sense of endorsement and self-familiarity, to the quality of 'warmth and intimacy' that William James claimed characterizes our own present thoughts (James 1890: 239)." Zahavi acknowledges that MSC in more robust sense can be disturbed and perhaps even be completely absent.

2.4.1. Minimal self-consciousness in the pattern theory of self

Below, I elaborate the differences between minimal and robust manifestations of MSC by means of pattern theory. This is useful in order to grasp the subtleties of our self-experiences and also to clarify the concept of MSC. Gallagher (2013) points out that an advantage of the pattern theory is that it helps us to see how the various aspects of self may be related in important ways. Related to the minimalist notion, I propose that the notion of MSC endorsed in this dissertation entails that the minimal experiential aspects of self have a special significance in a pattern of self. The experiential aspect is a necessary feature of self, whereas the other aspects are not. Related to the more robust notion, I propose that MSC involves also embodied and affective aspects, intersubjective, behavioral and extended/situated aspects, and it is intertwined with more sophisticated language-involving aspects. In other words, the experiential aspect is intertwined with all the others since it occurs together with those in self-consciousness. However, strictly minimalist self-consciousness encompasses only minimal experiential aspects—and only a part of experiential aspects.

In terms of pattern theory of self, it seems evident that whenever one talk about 'self-consciousness,' minimal experiential aspects are involved. Minimal experiential aspects are the core of minimal self-consciousness since they refer to the most fundamental form of self-consciousness. According to Gallagher (2013, 4) minimal

experiential aspects "contribute to an experiential and embodied sense of ownership (the 'mineness' of one's experience, as well as of one's body and movement), and a sense of agency for one's actions" (Gallagher, 2000, 2012; Rochat, 2011). This characterization reveals that experiential aspects involve at least two closely related features of self-consciousness: the sense of ownership and sense of agency. According to Gallagher (2000, 15), the *sense of agency* refers to "the sense that I am the one who is causing or generating an action" and *the sense of ownership* refers to "the sense that I am the one who is undergoing an experience." In normal experience, these senses coincide, but there are also experiences that lack the sense of agency. For instance, if somebody pushes me in crowd, I can have a sense that I am moving but without a sense of causing or controlling the movement. Thus, the sense of ownership is the most fundamental form of self-consciousness; it is the feature of the experiential aspect of self that cannot lack in experience. This notion of sense of ownership can be equated with the notion of subjectivity or subjective character that is used in this dissertation (in Sec. 2.2). The subjectivity is the minimalist form of MSC that is always present in self-consciousness, while other features of self can vary and even be missing.

However, I want to point out that in more robust reading, MSC involves a number of aspects of self. MSC refers to the manner of experiencing, instead of the content of experience, and this manner can have many shades, which I briefly consider next (and that are also considered in Chapters 6-8). First, as the discussion above showed, the experiential and embodied aspects are closely connected. This is a basic assumption in the neurophilosophical approach to conscious mind (see Chapter 1) and is emphasized in recent theories of embodied mind (and enactivism). According to Gallagher (Gallagher & Daly 2018, 15), embodied aspects are "core biological, ecological and interoceptive factors, allowing the system to distinguish between itself and what is not itself—extremely basic to all kinds of animal behavior." However, it is not fully clear how Gallagher sees these aspects, and for clarifying remarks, their role is elaborated below.

The role of embodiment is so significant that without embodied aspects of self, there could not be experiential ones. That is, embodied aspects are a necessary for experiential ones. This claim can be understood in two senses: as a claim about self-consciousness involving bodily features or as a claim about a (non-conscious) material basis of self-consciousness. The first sense was considered above (in Sec. 2.1) in terms of the living body; embodied aspects often are part of experience since the body is involved in perceiving and action, that is, in the "ecological" interaction with the environment. In other words, the body frames experience even when it is not the object of experience. The interoceptive factors, which Gallagher mentions, are also always present in experiencing. Interoception refers to the sense of the internal physiological condition of the body (Seth, 2013; Sec. 4.1.2.), and some kind of tacit inner monitoring of one's bodily states is involved and necessary for experiencing.

However, the pattern theory can be interpreted in the sense that connects embodied aspects to sub-personal processes. In this sense, it is important to recognize the significance of embodied aspects, but if they refer only to the function of sustaining self–non-self-distinction or interoception, which we often do not pay attention to, they seem to be unconscious or non-conscious. Gallagher (2013) is not explicit about how he understands embodied aspects, but following the definitions in the pattern theory, Gallagher seem to refer to this sense of non-conscious bodily processes. Since Gallagher (Gallagher & Daly 2018, 4) describes minimal experiential aspects as "first-person, pre-reflective, conscious experience, reflecting the self/non-self distinction," it seems that embodied aspects themselves are not conscious, but only experiential aspects involve consciousness about the embodiment. This dissertation follows this idea and considers embodied aspects mostly in neural terms. In other words, minimal embodied aspects are seen as the subpersonal basis for personal-level experiential aspects of self and are significant as such even if they are not experienced. In this sense, for instance, 'bodily self-consciousness' involves experiential aspects connected with the embodied ones.

Minimal experiential aspects of self typically occur together with other aspects. Affective aspects, at least, are typically included in minimal self-consciousness: Even when one is not focusing on herself but on the environment, she feels herself as a sentient bodily agent with emotions (e.g., Colombetti, 2007, 2011). The role of affectivity is many times linked to embodiment, which highlights it as a basic feature of self-experience (e.g., Colombetti, 2011; Seth, 2013). Behavioral aspects also seem to easily fall into MSC according to the idea of embodiment that emphasizes action as a basic feature of self (Sec. 2.1). Gallagher's (Gallagher & Daly 2018, 4) characterization of these aspects is very broad ("Behaviors and actions make us who we are") and could be interpreted to involve elements from both minimal and reflective self-consciousness.[42]

Further, the intersubjective aspects enter the picture when we are engaged with other people. For instance, in direct perception, I can feel the mental state of my communication partner only from her motion and facial expressions without any reasoning. In addition, intersubjective aspects are crucial in the development of self-consciousness. It has been argued that even before full-fledged reflective self-consciousness, infants are pre-reflectively aware of themselves not as objects of others' awareness but as co-subjects of a co-attended experience (e.g., Ciaunica, 2015; Ciaunica & Crucianelli, 2019). Further, one learns language only in communication with others, and interaction with other people leads one to grasp various ways in which she can examine herself (e.g., Baker, 2012; Zahavi, 2014).

Furthermore, MSC is involved in using reflective self-consciousness. Although one is not exercising her reflective self-consciousness all the time, the psychological/cognitive aspects enrich the scope of self-experience immensely. We have a rich and complex cognitive phenomenology in which MSC can manifest itself in tremendously different ways. One is not only living through immediate perception of this moment but can also ponder her

---

[42] These aspects might be associated with philosophy of action, but that broad field of philosophy is beyond the scope of this dissertation. For philosophy of action, see, e.g., Paul (2020); Wilson & Shpall (2016).

many feelings and thoughts, accessing herself in memories and making plans for her future. Thus, self also turns into a narrative agent that tells a life story into which her ongoing self-experience converges. The narrativity and interpretation of events is easily, even with a little reflection, involved in experience. And MSC is also involved in experiences that are not focused on one's self.

Extended/situated and normative aspects (Gallagher & Daly, 2018, 4) also seem to enter into our everyday experience easily. Extended/situated aspects refer to the way we may identify with, for instance, our material belongings, the technologies we use, and our professions. These kinds of elements can be associated with embodied and experiential aspects; the classic example is a blind person's cane, which the person does not experience as an external thing but rather as a part of her body (Merleau-Ponty, 1962).[43] Normative aspects involve, for instance, the kind of family structure in which we grew up and cultural and normative practices. Also these kinds of factors seem to be present in experience without reflection; for instance, cultural norms affect the way one feels and reacts in a situation. Thus, the experiential aspect indeed seems to be interwoven with all the other aspects in self-consciousness.

In order to keep this dissertation focused enough, in the following chapters, I limit the consideration of MSC to include experiential, embodied and affective aspects of self. These aspects are very basic ones, and thus in the core of MSC (which is recognized in both embodied and experiential readings of minimal self). They are present in experiences that lack other aspects, for instance, in situations that do not involve action, other people, higher-order reflection or use of external elements. In addition, the definitions of some other aspects seem to be rather vague; for instance, extended/situated and normative aspects involve a number of factors but it is not clear how those factors are related (for instance, it seems somewhat different to identify with a material device and with a profession). Thus, I want to provide a picture of basic MSC, and that picture can be elaborated in future studies to

---

[43] The extended aspects might be discussed in terms of "extended self," which would be an interesting research topic, but cannot be elaborated here. For "extended mind," see fn.32.

incorporate more aspects of self. I do not deny that the other aspects are significant for self-consciousness, but for simplicity they are excluded from the study of self-consciousness here.

2.4.2. Necessary and sufficient features

To recap, in minimalist reading, MSC refers to the subjective character or subjectivity of consciousness, which is part of the experiential aspect. In robust reading, MSC refers to a subjective perspective, which also involves bodily and affective features in experience. I want to point out that, in terms of pattern theory, this conception gives a special status for the experiential aspect. The experiential aspect is necessary in a way in which other aspects are not for self-consciousness, consciousness and self. Because of this necessary status, only experientiality comprises the minimalist reading.

First, this conception of MSC implies that the experiential aspect is, by definition, necessary for self-consciousness. Other aspects may alternate, but the experiential aspect is always present in self-consciousness. Simply, subjectivity is necessary and sufficient for MSC. There is MSC iff there is subjectivity. Subjectivity is necessary for self-consciousness; there cannot be self-consciousness without subjectivity. In addition, subjectivity is sufficient for MSC: if there is subjectivity, there is self-consciousness, no other features of self are required. Thus, the experiential aspect is special in the pattern of self-consciousness because it can occur without the other aspects but not the other way around. Thus, MSC is the most fundamental form of self-consciousness that grounds and is involved in all other forms.

Second, MSC is a structural feature of consciousness, which reveals the intimate connection between consciousness and selfhood. The phenomenal character of consciousness involves two characters (as presented in Sec. 2.2.1.): qualitative character refers to contents of consciousness, whereas subjective character refers to subjectivity of consciousness. MSC is the subjective character, which makes a mental state conscious in the first place: there cannot be contents of consciousness without MSC. This means that the minimum of self-consciousness is at the same time the minimum of

consciousness, and that MSC distinguishes conscious and non-conscious mental states. MSC as the subjective character of consciousness is a necessary and sufficient feature of a conscious mental state.

Third, the notion of MSC entails that the experiential aspect is necessary for being a self. That is, MSC is also special in the sense that it is a feature a self cannot lack. The emphasis of experiential MSC (in Zahavi's characterizations) highlights the interlinks between selfhood, self-consciousness and consciousness. MSC determines a subject of experience, and given that a subject of experience is a self, MSC determines a self. Mere embodiment (that, e.g., a robot also has) is not enough for being a self; subjectivity is needed. In other words, consciousness emerges only in sophisticated enough cognitive systems, and only these kinds of systems can be considered as selves. Thus, the notion of MSC defines a particular kind of cognitive system: a system that has a particular subjective perspective and is a self.

Despite this essentiality for self and consciousness, the phenomenon of MSC is hard to capture. It is a difficult research object since it always is intertwined with the contents of consciousness and other aspects of (self-)consciousness. Regarding the content, that is, the qualitative character, of consciousness, MSC cannot be described as a part of it because, by definition, it is not an object of consciousness. Instead, MSC is involved in representing all objects of consciousness as the subjective character of consciousness. Regarding the other aspects of self-consciousness, as proposed above, minimal self-consciousness occurs together with them.

This is a notable methodological point: MSC cannot be captured by introspection on its own, since whenever one turns attention to herself, she already is exercising her reflective self-consciousness. Basically, MSC is a tacit feature of experience, a kind of background in which all experiences take place but which cannot be examined in isolation from the experience. Also empirically, determining the neural profile for minimal self-consciousness is challenging. Since MSC refers to the ubiquitous subjectivity of consciousness, there is not a contrast case in which

one conscious mental state lacks MSC and the other has it. Instead, the empirical research has to focus on the variance of MSC between conscious mental states. Overall, the tacit or fleeting character of MSC does not make it less significant but asks for methodological innovations and ingenuity. The above conducted conceptual analysis is required in order to define MSC precisely, and clear concepts are also needed in the development of imaginative paradigms for empirical research on MSC (which will be considered in terms of neural activation in Chapter 4 and in terms of altered states of consciousness in Chapters 6-8).

2.5. Illustration of layers of self-consciousness in dreaming

It has been presented above that MSC is the most elementary form of self-consciousness; however, one might still wonder how minimal the MSC can get and how it manifests itself in different instances. Attempts to illustrate minimal self-consciousness in a simple form have been carried out by studying altered states of consciousness (more about this method in Sec. 6.1.). Here I briefly consider a particular altered state of consciousness, dreaming, that can illuminate layers of minimal self-consciousness (Metzinger, 2013; Revonsuo, 2005; Thompson, 2014; Windt, 2015).[44] Generally, dreaming refers to subjective experience during sleep and is interesting for philosophers because dreams can be used as an instrument that guides to a deeper understanding of consciousness, self-consciousness, and subjectivity (Metzinger, 2013; Revonsuo, 2006; Thompson, 2014; Windt, 2015). [45] Metzinger (2013) argues that dreaming can significantly assist in answering to the problem of how to isolate minimal phenomenal selfhood (MPS) particularly. According to Metzinger (2013, 8), dreaming is the best global contrast condition for isolating MPS; dream research is able to reveal the strictly necessary features of self-consciousness. Thus,

---

[44] For a more detailed version of this section, see, Haanila (2021).
[45] In more detail, dreaming has been defined in terms of simulation (Revonsuo, 2005; 2006), hallucination (Windt, 2010; 2015), and imagination (Thompson, 2014).

dream research can make a decisive contribution to the philosophical project to conceptually describe layers of human self-consciousness. Since self-consciousness is present in dreaming in peculiar forms that lack some features of normal waking consciousness, different aspects of self can stand out more easily.

In general, considering dream experiences is relevant for the study of selfhood since some kind of dream self is present in the great majority of dreams. Roughly, a 'dream self' is the protagonist of the dream with whom the dreamer identifies herself (see, e.g., Revonsuo, 2005; Thompson 2014; Windt 2015). Many times the dream self resembles the waking self, at least in respect of some features, but not necessarily. Further, typically the dream self differs from the waking self at least in its (meta)cognitive skills; the central characteristics of the dream self is its lack of the full mental capabilities of the waking self. The dream self typically suffers from a lack of rationality and deliberation; it does not mind the discrepancies and incoherencies in its surroundings and own actions. Further, the dream self is amnestic; even if some of the waking self's memories were accessible, not all of them are. For instance, the dream self might have a nice conversation with a relative who passed away long ago without noticing it as strange in any way. All in all, a lack of self-reflection is the characteristic feature of the dream self.

In other words, it is clear that in dreams one can have a sense of self but lack reflective self-consciousness. This indicates that psychological-cognitive and narrative aspects are not necessary for self-experience. It is also remarkable that the sense of agency can diminish in dreams; that is, one can have dreams of being a mere passive observer without active participation in the dream events. These kinds of dreams can also involve lack of affectivity (although many times dreams involve intense emotions; see Sikka, 2020) characterized by calmness in watching the dream scene. Thus, dream experiences offer a way to get to a peculiarly minimal self-consciousness.

In addition, the dream research shows that there can be self-consciousness without consciousness of the body. Metzinger (2013) and Windt (2010, 2015) have used this example of the most

minimal self-consciousness. Metzinger (2013) supports Windt's (2010) definition of MPS as "transparent self-location in a spatiotemporal frame of reference" and praises the definition as advantageous for covering highly atypical experiences such as bodiless dreams. Bodiless dreams are a rare but well-known phenomenon in which a dreamer identifies with an extensionless point in perceptual space. Metzinger explains that in these cases the dreamer has an "abstract self-representation"; that is, "the currently conscious self-model does not contain any perceptual or spatially extended features of bodily content: There is no visual, auditory, olfactory, gustatory, tactile, proprioceptive, vestibular, nociceptive, thermal, or interoceptive information represented on the level of bodily self-consciousness." Metzinger (2013, 7) refers to this kind of experience as "bodiless subjectivity, i.e., states in which body representation is absolutely minimal, but in which a stable sense of selfhood and an 'asomatic 1PP' can be found."[46]

Thus, bodiless dreams illustrate a stripped-down case of (self-)consciousness: they lack many central features of normal waking consciousness, yet they are subjective conscious states that involve subjective perspectives. This indicates that MSC, and the experiential aspect in particular, indeed is the most fundamental level of self-consciousness. It can occur in the absence of other features of self-consciousness but not the other way around. Experience can lack even a representation of a body that typically is involved. Further, dreams in general can involve contents and features that lack the richness of waking consciousness; environment can undergo incoherent changes, and the dream self may suffer from errors in reasoning that she does not, however, mind. Thus, dream research can contribute to dissociating different levels of self-consciousness by showing that the subjective perspective is present even in highly atypical experiences and most

---

[46] According to Metzinger (2013), in addition to dreams, there are also two other sources of bodiless subjectivity: the scientific observation of OBEs (i.e., Out-of-Body Experience) and meditation research. However, Metzinger notices, both asomatic OBEs and "pure consciousness" experiences in meditators are rare phenomena and thus, more difficult ways to investigate MSC. Nevertheless, these altered states of consciousness seem promising ways to examine MSC; more about meditation later in Chapter 8.

simple ones, such as bodiless dreams. This strengthens the conclusion that MSC determines the most minimal form of selfhood since it prevails even when other aspects of self-consciousness are missing.

2.6. Summary

This chapter has advanced a concept of MSC as a constitutive feature of immediate experience: MSC refers to a subjective perspective or, in the end, to the subjectivity of experience. The concept of MSC determines the minimum of self-consciousness and is needed in grounding more sophisticated forms of selfhood higher in the cognitive hierarchy. However, the notion of minimal self can be understood in two ways. The embodied reading highlights the spatiotemporal perspective and how we experience through the body. This reading can also be connected to enactivist approach to mind, according to which self is constituted by bodily processes and interaction with the environment, or is can be seen as a rather functional description of a cognitive system. Self is a self-specifying cognitive system that can make a concrete distinction between itself and the environment. Yet, this kind of embodied reading, which defines self in a functional way, risks missing out consciousness. Instead, the experiential reading of self-consciousness takes experience as the starting point, which is reasonable since the whole notion of 'self-consciousness' would be poorly defined without reference to experience. MSC refers to the first-personal presence (how) of experience, not to a content (what) of consciousness. This entails that MSC is a necessary constituent of experience.

As a tacit and ubiquitous feature of consciousness, MSC is difficult to discern. However, the analysis in this chapter has been able to elucidate its shades. In minimalist reading, MSC refers to the subjectivity and self as the subject of experience. In robust reading, MSC involves all the features of experience that concern the manner of experiencing instead of the contents of consciousness. Many times, these features involve the sense of agency and embodied, perspectival and affective factors in

experience. However, subjectivity alone is the necessary feature of MSC. In terms of the pattern theory of self, this means that the experiential aspect is special in a way that other aspects are not; it is interwoven with all other aspects in (self-)consciousness, but it can also occur without the other aspects. Altogether, MSC is fundamental for consciousness and being a self, and more sophisticated forms of self-consciousness require and involve it. When one starts to scrape off the layers of experience to figure out what there is in the core of consciousness, she will eventually find and be left with the experiential aspects of self.

3. Shades of reflective self-consciousness

Reflective self-consciousness provides us with rich opportunities to make reflective judgments about our mental states that are directed to ourselves and to modify them. Reflective self-consciousness (RSC in brief) can be defined as a capability to take oneself as the *object of one's thoughts* and to think of oneself *as oneself*. By means of RSC, one can focus her attention on herself and evaluate and direct her action. RSC is a necessary condition for moral responsibility and self-critical deliberation, and for that reason, many traditional theories of self have concentrated on it. Although the notion of RSC is more straightforward and less controversial than the concept of MSC, it too involves multiple aspects of self, and it is not clear-cut. Therefore, the notion of RSC also needs elaboration.

In this chapter, the notion of RSC is firstly described generally and then specified through various distinctions within self-reflection. The most general distinction within RSC is drawn between thinking of oneself in third-person and first-person; first-personal thoughts are unique since they have special semantic, epistemic and motivational features. I propose that these remarks should be deepened by more fine-grained notions of self-reflection. I complement the account of self-reflection with the distinction between a deliberative and a theoretical stance toward oneself. Further, I point out the difference between voluntary and involuntary self-conscious thinking. Lastly, I consider the mode of identification in self-reflection. These three additional distinctions are not competing but rather describe different dimensions of RSC and thus can elaborate the picture of self-reflection.

Generally, this dissertation concentrates on human RSC. However, the (simply defined) capacity of taking oneself as the object of thoughts and of recognizing oneself as oneself can also be found in some other species. In developmental psychology and comparative psychology, the capacity of (reflective) self-consciousness has often been operationalized by the highly influential paradigm of the mirror self-recognition task (Gallup,

1970, 1977, 1982; Rochat & Zahavi, 2011) In this task, a subject's forehead is marked with a red stain in such a way that the subject does not notice the marking. Then the subject is placed before a mirror and if she tries to wipe out the relevant stain (the one on her forehead, not the one in the mirror), she recognizes herself and is in possession of RSC. Infants "pass" the mirror test at the age of 18 months on average (e.g., Amsterdam, 1972). In addition, some non-human species pass the test, including chimpanzees, orangutans (Gallup, 1982), bottlenose dolphins (Reiss & Marino, 2001), and Asian elephants (Plotnik et al., 2006). The mirror tests have also been criticized since it is not clear what the relation between self-consciousness and mirror self-recognition exactly is (see, e.g., Zahavi, 2014, 199-205). For instance, mirror tests do not take the varieties of self-consciousness, including MSC, into account and cannot be considered as the main operationalization of the spectrum of self-consciousness. Furthermore, success in mirror tasks have been noticed to involve variation demonstrating an element of social or cultural context (e.g., Broesch et al., 2011). Nevertheless, the mirror self-recognition tests illustrate that RSC includes the capacity to recognize one's picture in a mirror; this recognition is a complex skill, taking time to develop, and rare in animal kingdom.

Yet, human RSC goes way beyond mere mirror tests and is unique among species because of the complex cognition and linguistic skills we have. We not only recognize our bodily features as our own in the mirror reflection but also form concepts, beliefs, inferences, and narratives about ourselves; our self-understanding is enriched and structured by the expressions of language. This kind of sophisticated RSC is necessary for being a moral agent which makes it an essential feature of humanity. One is responsible for her action only if she has the capacity to evaluate and direct it, [47] and this responsibility is enabled by RSC (see, e.g., Zahavi, 2014). In addition, reflective self-consciousness has been connected

---

[47] For this reason, e.g., very young children and people suffering from severe dementia are not considered to be responsible for their actions. On moral responsibility in general, see, e.g., Talbert, 2019.

to memory, tendency to see one's life as a narrative, and a long-term personal identity in Lockean sense (e.g., Baker, 2012; Gallagher & Zahavi, 2007; Gallagher, 2000; Kriegel, 2004).

Altogether, RSC makes possible certain ways of thinking and acting that are frequently considered characteristically human. As Bermúdez (2007) formulates, RSC "plays a distinctive role within the cognitive economy": without the capacity to be aware of one's own thoughts, one would be unable to engage in many intellectual activities and deliberate action (see also, e.g., Moran, 2001). Zahavi also acknowledges (2014, 50):

> Ultimately ... I don't think we should make do with the thin notion of experiential self. This notion, although fundamental, has some clear limitations, and it should be supplemented by thicker notions that do justice to other important aspects of self. More specifically, our account of human selfhood will remain inadequate as long as we fail to consider the self that forms plans, makes promises, and accepts responsibilities, the self that is defined and shaped by its values ideas, goals, convictions and decisions.

These thicker senses of self are enabled by RSC.

3.1. The more full-blown notion of self-consciousness

The general notion of RSC as the capability to take oneself as the object of one's thoughts is grounded in both phenomenology and analytic philosophy. In the phenomenological approach, Zahavi and Gallagher (2007, 61) define RSC as "an explicit, conceptual, and objectifying awareness that takes a lower-order consciousness as its attentional theme." In analytic philosophy, for instance, Kriegel (2003) formulates that a subject is in possession of RSC when she "is conscious of her thought that p or conscious of her perception of x as her thought or perception."

Reflective self-consciousness has been considered so significant that some theories define the whole notion of 'self-consciousness' as self-reflection. These theories do not ascribe MSC self-

consciousness at all but refrain from the term 'self-consciousness' except at a cognitively higher level (Baker, 2012; Musholt, 2013; Revonsuo, 2006; Taylor, 1989). However, according to the multidimensional approach to self (advanced in Chapters 1 and 2), the first-personal character of MSC grounds RSC. MSC provides an experiential basis for any subsequent self-ascription, reflective self-identification, and self-conscious thought; reflection grasps experience, which was minimally self-conscious already before grasping.

In any case, the difference between minimal and reflective self-consciousness is clear. Unlike the minimal form, reflective self-consciousness is an introspective, rare, voluntary and effortful type of self-consciousness (Kriegel, 2003). In contrast to pre-reflective MSC, RSC is reflective by definition; one concentrates on her mental (or bodily) states and takes herself as the object of consciousness by introspection. Further, while MSC is a necessary constitutive feature of all experiences that is always present in consciousness, RSC is rare and voluntary[48] in that a great deal of one's cognition is not about oneself but focused on environment, other people, etc.; only in special cases does one decide to focus on herself especially. Furthermore, as a built-in feature of consciousness, MSC does not need a special effort for occurring in experience, but RSC is effortful. It requires concentrated cognition in which the subject directs and maintains her focus on herself. RSC also includes a significantly wider time perspective; while time as a feature of MSC refers to immediate experience, RSC enables one to reflect upon her past and plan her future.

Analytic philosophers many times highlight the uniqueness of knowledge one has of herself through RSC (e.g., Gertler, 2011). In the exemplar case, a person wakes up in a dark and quiet room trying to figure out what has happened to her. However, the figuring out is tough because she suffers from amnesia; she has no idea where she is or how she has got there; and even memories of

---

[48] The voluntariness of RSC can also be disputed; see Sec. 3.4. However, generally RSC is characterized by voluntariness, which can be considered a distinctive feature of RSC compared to MSC.

her past life are vague. Since the room is dark, she has no information of her current surroundings either. However, she is able to contemplate her condition and reflect on her scope for action by RSC. She starts to feel her way by touch in the room and thinks that when she finds a wall, maybe she could find a light switch as well and can get light into the room. She also feels faint with headache, and when she is running her hands over her head, she can localize the pain in her temple. Luckily, the headache is easing; maybe her memory too will return soon. Has she been knocked down, kidnapped, and thrown into some suspicious place? Or has she fainted and hit her head in her own bedroom? In any case, the exemplar indicates that a person can have a wholly functioning RSC although she lacks some other aspects of self. At the lower level of minimal self-consciousness, she has the subjective experience of waking up; even without visual or auditory perception, she localizes herself in 1PP, she has bodily feelings, and she can explore her surroundings by touching. At the level of higher cognitive skills, she can think about herself (as herself) and reason what to do without a more extensive knowledge of herself or her environment.

3.2. First-person versus third-person: uniqueness of self-conscious thoughts

A distinctive feature of human reflective self-consciousness is essentially connected to language.[49] Hence, one natural route to study self-consciousness is to examine the use of the first-person pronoun 'I' by which self-consciousness manifests itself. Analytic philosophers have been puzzled and particularly interested in the specific semantic, epistemic and motivational peculiarities of *self-conscious thoughts*, that is, thoughts referring to oneself by the use of the first-person concept 'I.' These peculiarities distinguish self-consciousness from consciousness of things other than oneself. In addition, importantly, these peculiarities hold only when one is

---

[49] That is why it is also labeled a 'conceptual self-consciousness', e.g., by Bermúdez 2001, 2007.

referred to in first person. Thus, the most common distinction within RSC is made between first- and third-personal considerations of oneself. Below (some of) these peculiarities are briefly presented, bringing out the specific perspective that first-personal RSC has in contrast to a third-personal stance toward oneself.[50]

Many times, the discussions on special characteristics of (conceptual) self-consciousness start by noticing that a particular class of first-personal statements are *immune to the error through misidentification relative to the first-person pronoun* (generally shortened to IEM).[51] Shoemaker (1968) introduced the concept in investigating statements in which 'I' is used "as subject." According to Shoemaker, this is the use that philosophers have found puzzling and that can reveal the significance and uniqueness of self-consciousness. Thus, Shoemaker points out that there are two different uses of the word 'I' (or 'my'): "the use as object" and "the use as subject."[52] Examples of the first include such sentences as "My arm is moving," "I am bleeding," and "I have grown six inches," whereas examples of the second include: "I see so and so," "I think it will rain," and "I feel pain." Shoemaker considers that the use of 'I' as subject is more fundamental, for it has necessary features that the use as object has not. The subject-use is connected

---

[50] The whole discussion of 'I' cannot be covered here but the aim is to give a general idea of the intricacy of RSC and a brief introduction to the topic. More detailed overviews and discussions on these topics can be found, e.g., in Brook & DeVidi (2001); Gertler (2011); Smith (2020). In addition, the theme has been discussed in eastern philosophies (using different concepts); see, e.g., Siderits et al. (2011).

[51] Castañeda (1966, 1967) made a closely related point in studying third-person sentences like "The editor believes that he* is F." Later, Matthews (1991) extended this discussion to sentences like "I think that I* am F." Following Castañeda, for the cases in which a person attributes a first-person reference to herself, 'I' has been marked with an asterisk in the literature. The 'I*' expresses a self-concept by which a person conceives of herself as herself and cannot mistakenly believe that she is referring to someone other than herself (Baker, 2012).

[52] Wittgenstein (1958, 66-7) presented this distinction between two different uses of 'I' and is often mentioned as the initiator of the discussion on 'I' in analytic philosophy. James (1890) made a parallel distinction between 'I' and 'me'—'I' reflecting the self as a subject of experience and 'me' the self as an object of experience. The distinction can also be seen in Kant, who anticipated analytic philosophy's ideas of self-consciousness (see, e.g., Brook, 2001).

with a subjective perspective on oneself, which sets it apart from any other way of referring to oneself.

The use as object cases of 'I' involves recognition of a particular person, and these cases carry a possibility of error. For example, in a case when one sees a number of arms in the mirror and with a quick glance thinks to herself "I have a nice arm," she may have mistakenly taken the nicest arm to be hers when it in fact is someone else's. That is, she can be right in considering the arm is nice but mistaken about who has the arm. Or if one is wrestling, finding herself in a tangle of bodies and seeing a bleeding leg and saying "I am bleeding," it can happen that what she says is false even though she is correct in that a certain person is bleeding. Because her body is tangled up with the body of the other wrestler, she is mistaken about whose the bleeding leg is and in thinking the bleeding person to be herself (Shoemaker, 1968; Wittgenstein, 1958).

Instead, when 'I' is in use as subject, there is no question of recognizing a person: these sentences are immune to error due to misrecognition of a person; that is, they are IEM. It would not make much sense for the subject to ask who the subject of believing or perceiving is in these sentences. For instance, if one says "I feel pain," she is immediately aware of the pain, and it makes no sense for her to wonder whether the pain that she is aware of is hers. Or in saying "I see a canary," one can be mistaken in taking what she sees to be a canary (it might be some other bird), or in an extreme case of hallucination, she can be mistaken that there is a bird at all to see. But she cannot misidentify herself as the person she knows to have the feeling of seeing a canary (Shoemaker, 1968).

However, IEM applies only to a particular class of self-conscious thoughts. Firstly, IEM applies to the thoughts that are directed to one's mind, as opposed to one's body. As in the above examples, the thoughts about one's body involved 'I' only in its object use, and one can be mistaken whose body is in question.[53] Secondly, the particular class of self-conscious thoughts that are

---

[53] This idea has been used in arguing that the referent of 'I' is not body, which can be misidentified, but thinking and mental life instead, see, e.g., Anscombe (1975).

IEM is relative to certain mental grounds (e.g., introspection); a thought is not IEM by virtue of its content alone. For instance, the judgment that one is jealous is IEM when it is grounded in introspection. However, the judgment is not IEM when it is grounded in the overheard testimony of one's analyst (since it is possible that one has misinterpreted her analyst's words, wrongly taking his use of "Smith" in "Smith is jealous" to refer to her).[54] This particular class of self-conscious thought is referred to also as those made on the basis of information "from the inside"; in contrast, judgments made "from outside" are based on exteroceptive perception or testimony (Grünbaum & Zahavi, 2013; McGinn, 1983). However, in addition to the class of thoughts that are IEM, self-conscious thoughts have even more essential features that characterize all of them.

Namely, more fundamentally, all self-conscious thoughts are guaranteed success in referring to the subject of the thinking, and this entails special epistemic and motivational properties. Insofar as I am thinking of myself with an 'I'-concept, I know that I am referring to myself and there is no possibility of reference failure ("as Descartes's 'cogito argument' brings out," Shoemaker 1968, 557). It is impossible to think an I-thought and not refer to oneself. This *immunity to failure of reference* is connected also to *reference without identification*: when a person refers to herself by an 'I'-concept, she is aware of herself as the subject of thinking without knowing any identifying properties of the subject. Further, she knows that "she is referring to herself, and this knowledge makes whatever she is ascribing to herself stand out as something which is significant to her" (Grünbaum & Zahavi, 2013, 223; Perry, 1979; Shoemaker, 1968). These specific epistemic and motivational features of self-conscious thoughts are elaborated below.

First, there are epistemic differences between 1P and 3P references to oneself: a person cannot think an 'I'-thought without knowing and understanding that she is thinking about herself (Castañeda, 1967; Grünbaum & Zahavi, 2013; Perry, 1977;

---

[54] The precise scope of first-person judgments that are IEM is a contested matter; see, e.g., Smith (2020).

Shoemaker, 1968). This is based on immediate familiarity with her own experience and does not involve any identificatory step. Identification goes together with the possibility of misidentification, but there is no such possibility for self-conscious thoughts; my self is accessible to me in a way in which it is not to others. By contrast, when one refers to herself in the 3P mode—that is, by a name, a definite description, or a perceptual demonstrative—there are more identificatory steps and always the danger that "she is in fact referring to someone else...or that she does not realize that she is in fact referring to herself" (Grünbaum & Zahavi, 2013, 224).

The possible errors of 3P self-reference can be seen, for instance, in Perry's (1977) thought experiment of amnesiac Rudolf Lingens, who is lost in the Stanford Library. There, Lingens reads, among other things, a biography about himself without realizing that the main character is he himself; a name or even a very detailed description does not give self-knowledge to Lingens, and he does not understand that he is referring to himself when he thinks about the life of Lingens. On the other hand, one can fail to refer to herself in 3P. For instance, if Lingens happened to see only relatively short people in the library, he might refer to himself with a definite description "the tallest person in this library"; however, it might turn out that also somebody taller is there riffling through the books, and the description does actually not hold true for Lingens.

Secondly, 1P self-reference includes specific motivational features that matter to the subject's emotional feelings, practical reasoning, and action (Grünbaum & Zahavi, 2013, 224; Baker, 2012; Castañeda, 1966, 1967; Perry 1977, 1979, 2001; Shoemaker, 1968). In other words, first-personal thoughts are "self-locating" and thereby enable and motivate action, while non-first-personal thoughts are not linked to action in the same direct way.

A famous example is Perry's (1979) story about himself shopping in a supermarket and noticing that there is a trail of sugar on the floor. Perry ponders that someone has been sloppy and starts to seek the shopper with the torn sack in order to tell him that he is making a mess all over the supermarket. But when Perry realizes that it is actually he himself who has the torn sack, he changes his action: he stops the seeking and rearranges the torn sack in his cart.

Kaplan's (1989, 533) example makes the connection between self-conscious thinking and motivation for action even clearer. In this example, I see in a window the reflection of a woman whose pants appear to be on fire. If I believe "Her pants are on fire," pure self-interest does not necessarily motivate me to do anything about it. However, if I believe "My pants are on fire," pure self-interest will surely motivate me and I will behave very differently.

It can be noticed that the point of the unique character of immunity to failure of reference is valid also for the object-use cases of 'I'. Considering, for instance, the earlier example in which one is wrestling, sees a bleeding leg and mistakes that it is she who is bleeding. Even in this case, she cannot be mistaken in that she is the subject of thinking of "I am bleeding" and the thought also motivates her to be careful with her leg. That is, all self-conscious thoughts have distinctive epistemic and motivational features that distinguish 1P self-reference and 3P self-reference.

3.3. Deliberative versus theoretical stance: importance of deliberative stance

In the following subchapters, I propose that the character of RSC should be elaborated in order to capture the shades of self-reflection, and I examine self-conscious thoughts with three additional distinctions. I consider these three cases as different ways to classify and systematize RSC; they are not in competition with each other but can be used in parallel. I propose that these classifications can be seen as providing answers to different questions thus complementing each other. The first distinction is drawn between deliberative and theoretical stances, which literally are stances or attitudes with which one can exercise RSC. The second distinction, between voluntary and involuntary RSC, instead refers to the start of self-reflection; the process of taking self as the object of consciousness is the same, but the initiation of the process can happen in two ways. The third distinction within RSC concerns the degree of identification with self-related contents. This last-mentioned distinction extends the analysis of RSC to also cover

cases in which one does not identify with herself and in which self as the content is voluntarily dissolved away.

In analyzing self-reflection, Richard Moran (2001) distinguishes between deliberative and theoretical stances that one can take toward herself. These stances are discussed here since they assist in illustrating the complexity of RSC and deepening the classical analysis of self-conscious thoughts. The deliberative stance especially is connected to the motivational dimension of self-conscious thoughts. In addition, Moran advances the idea that the capacity for first-person awareness has a special relevance to the psychological well-being of a person. According to Moran (2001, 136), this idea has operated as a background assumption in contemporary philosophy of mind, although it has not been investigated on its own. Moran strives for clarifying the assumption by arguing that the psychic health of a person requires alternating between the stances to herself and especially viewing herself in a deliberative stance. Moran's point is important since it indicates that the study of self is interesting not only for the fun of conceptual analysis but also for increasing mental well-being. I endorse the point, and I will use it in further analyses (especially in Chapters 6-8).

Along the same lines as self-consciousness has been described in this dissertation, Moran (2001, 32–3) first characterizes specifically the first-person manner of awareness that accompanies conscious mental states. According to Moran, self-consciousness is immediate, nonobservational, and involving reference to oneself through use of the pronoun 'I', instead of some mediating description under which the person might fail to recognize herself. However, Moran sees that we need a fuller characterization of first-person awareness in order to account for the special features it entails. This is because it is possible for a person to have immediate awareness of her mental state, which meets the above conditions but is still essentially a kind of outsider's perspective on the mental state. [55]

---

[55] Moran (2001, 63) remarks that "issues concerning the special features of the first person may be as much in the area of moral psychology...as in epistemology or metaphysics." More

In order to understand self-consciousness, Moran (2001) argues that we need to recognize the difference between two stances from which one can inspect herself. *Deliberative stance* is associated with practical reasoning, and it is essential in understanding first-person self-awareness. By contrast, *theoretical stance* is rather descriptive and third-personal. In both stances, self is taken to be the object of consciousness and the pronoun 'I' is used, but deliberative stance is more linked with an action, whereas theoretical stance involves more explanatory examination of one's mind. However, one stance does not eliminate the other. They both are needed, and RSC is characterized by their interplay.

In Moran's framework, questions from the theoretical stance concerning oneself are of the sort "What do I believe?" or "What is it that I feel?" These kinds of questions are answered by *a discovery of the fact* of which one was ignorant, and theoretical inquiry terminates in *true description* of one's mental state (Moran 2001, 58–63). Theoretical stance gives merely attributional self-knowledge, which is knowledge of one's state of mind that is mediated by some identifying description. Moran (2001, 128–9) calls the perspective also a rationalizing *interpreter's stance* by which one seeks to explain her actions and attitudes. One sees her beliefs as psychological data that explain her behavior purely in its role as a psychological state, giving a more or less good indication of her belief. Capacity for this stance is important, assisting, for instance, in normative evaluation of one's action, but it is not

---

specifically, Moran's conception of self-consciousness could be counted in the philosophy of action and will, which is a relatively new approach on understanding human action (started by Frankfurt's paper (1971); see, e.g., Arpaly & Schroeder (1999). It is also good to note that generally Moran argues for "a tradition of thought, roughly Kantian, but with sources both in Locke and in early Stoicism, which aligns, or even identifies, reflective self-consciousness with rational freedom" (Moran 2001, 139). According to this tradition, rationality requires self-knowledge, which itself implies self-consciousness; rationality is conceived as practical deliberation, and self-consciousness is its necessary condition. According to Smith (2020), this approach to self-knowledge can be found also at least in Bilgrami (2006); Boyle (2009); Burge (1996); Shoemaker (1988); for a general discussion, see Gertler (2011, Ch. 6). However, that wider discussion is not within the scope of this dissertation, but the distinction between deliberative and theoretical stances is considered simply as an elaboration of the notion of RSC.

sufficient for capturing the special characteristic of self-consciousness.

Deliberative stance, instead, is a position fundamentally different from that of another person (Moran 2001, 128). Deliberative questions of oneself are of the sort "What am I to believe?" or "What shall I intend?" and, as such, a matter of determining what is true. Thus, deliberative reflection belongs to the same family of thought as practical reflection; it "does not conclude with a normative judgment about what would be best to do, but with the *formation of an actual intention to do something*" (Moran, 2001, 59). Deliberative inquiry terminates in the formation of the endorsement of an attitude. Deliberative or practical questions are answered by a *decision or commitment* of some sort (the answer is not simply a response to ignorance of some antecedent fact about oneself as within theoretical stance). In other words, within the deliberative stance, a person's reflections on her own state have a dynamic or self-transforming aspect; the person herself plays a role in formulating how she thinks and feels (Moran, 2001, 58–63). Moran considers deliberative stance as a *stance of the reasoning agent* who has first-person authority. This stance entails that the believer does not treat her belief as an opaque psychological fact. Her belief is not for her a psychological datum that could even in principle justify her "behavior purely in its role as a psychological state" (Moran, 2001, 128–131). The relation in which a person is with herself in deliberative self-consciousness is first-personal, not the same kind of relation she has with other things. As Moran (2001, 113) describes, "I do not 'move my body' as I might move a piece of equipment, nor do I relate to my attitudes as mental furniture to be arranged."

Deliberative stance is essential to human nature and plays a crucial role in making us who we are. This is also the "strong" or "thick" sense in which some philosophers define self-consciousness altogether. For instance, in Taylor's (1989) words it is essential to our nature that we are "self-interpreting animals." According to Taylor (1989), conceptions of self-consciousness that do not meet this level are either nonexistent or insignificant, since a self can only exist in a normative space. However, within a multidimensional

approach to self (presented in Chapters 1 and 2), a solid conception of self-consciousness covers and recognizes the multiple forms of selfhood, including pre-reflective minimal self-consciousness, and can still point out deliberative self-reflection at the top of the hierarchy or as the "thickest" form of self-consciousness. Moran takes this route, on one hand ascribing (minimal) self-consciousness to consciousness generally and on the other hand, considering (reflective) self-consciousness essential for first-person authority and deliberation.[56]

Since the deliberative stance is rather complex, a further illustration of it is useful. Moran (2001, 59–64) describes it as an endorsement that involves an element of shaping one's attitudes. For instance, when one sees her anger as childish, or her feelings of guilt as false, the "seeing" is not purely descriptive but deliberative reflection about how to feel. Or when a person answers questions of the sort "What am I to believe here?," she thereby comes to believe something. Or when a person answers a question of the form "Is this what I really want?" by considering what is worth wanting, she thereby comes to clarify the structure of her desires.

Moran (2001, 142–5) elaborates the idea that self-reflection involves a metaphor of distancing, "stepping back" from one's current mental activities. Moran quotes Korsgaard (1996, Moran's emphasis): "The reflective structure of human consciousness sets us a problem. Reflective distance from our impulses makes it *both possible and necessary* to decide which ones we will act on: it *forces* us to act for reasons." Moran brings out that the metaphor of "reflective distance" presents a richer image of our self-consciousness than it may seem at first sight. "Stepping back" from one's impulse obviously points to creating distance, but it also involves *observation and bringing into view* and thereby a possibility to clarify the mental state in question. Further, stepping

---

[56] Zahavi (e.g., 2014, 52) also refers to Moran in describing RSC. In addition, Zahavi names, e.g. Frankfurt (1988) as sharing the idea of the multidimensional self. Similarly to Moran, Frankfurt emphasizes authorship, i.e., a thick sense of self-consciousness, as essential for being a person but at the same time recognizes that consciousness entails a basic thin form of self-consciousness, i.e., subjectivity of experience.

back enables *separation and distinguishing* (as in the sense of understanding that "the impulse is one thing and I another"). Furthermore, the metaphor entails *confrontation and facing*, revealing that something is unavoidably in one's path and that one has to decide what to do with it. In addition, "stepping back" involves the *halting* or interrupting of a (prior ongoing) action, suspending one's motion. Performing the stepping back from an impulse means that "now the impulse doesn't dominate me." As a result of the deliberation, the impulse loses the original inferential or functional role it had in one's explicit reasoning since one can bracket it. This is a genuine difference in her state of mind, not simply an externally applied description of it but unique for deliberative first-personal reflection.

Moran (2001, 146–150) further emphasizes that the deliberative stance is unavoidable for us. As Nagel (1996) puts it, "the reflective self cannot be a mere bystander"; unlike purely theoretical inquiry, deliberative questioning involves the capacity to determine psychological facts and thus, it plays a constitutive role in the psychological facts. When one is self-consciously (in a deliberative sense) reflecting on her state of mind, she always takes a stance toward what she discovers there, whether endorsement, permission or disapproval. The capacity to think entails recognition of the possibility to think otherwise. Thus, it is unavoidable for a person to decide whether some mental presentation shall count for her as a reason or not. Although one *can* treat any mental presentation of hers as data—something which gives her an indication of her genuine belief—there is also a point where one needs to face the deliberative question of "What shall I shall believe, or want, or intend?"

Though the deliberative stance has specific first-person features, Moran (2001, 161–2) highlights that self-consciousness requires both stances; the stances are not inconsistent but complementary, answering different questions. Each stance has prestige in its own domain and thus, a clash of the perspectives is not to be resolved simply by claiming that one has dominance. On one hand, there are situations of reflection in which the thinking is theoretical: it is part of the description or explanation of one's state. On the other

hand, there are situations of reflection in which the thinking is deliberative: it is part of the determination of one's state. Yet, there are also situation in which one faces a conflict between the two styles of thinking—for example, between reflection *about* one's desire and the reflection whose *conclusion* is some desire.

In demonstrating the contrasting roles of the two stances, Moran considers Sartre's (1956) example of the akratic gambler who resolves to stop gambling. Moran sees the case as helpful in showing how each stance presents its own demands as unavoidable, entailing an answer in its specific terms. In the example, the gambler has committed to her decision to stop and avoid the gambling tables. For her the situation is not about empirical evidence determining what she will do but rather a resolution of which she is "the author and responsible for carrying through" (Moran, 2001, 79). On the other hand, she knows herself empirically; from this point of view her "resolution" is a psychological fact about her and examined in relation to her history. In this theoretical point of view, her resolution appears then as an ungrounded and inconstant thing to determine what she will in fact do (since due to her akrasia, she has failed to keep in her antecedent decisions to stop). The virtue of the theoretical stance is that it tells her to be empirically realistic about herself; without such realism she might only exercise pretense or wishful thinking. However, it cannot tell her when such "realism" is simply the appearance, which results from her avoidance of the practical question before her. The virtue of the deliberative stance is that it tells her that she is not bound by her empirical history; she needs to face the question of what she *is to do*, or she would only be in a form of evasion. However, this perspective cannot tell her when her assumption of agency is a mere sham—when there are empirical reasons that undermine her normal strength to form an intention that counts in performing action. Thus, each perspective gives a useful viewpoint without denying the truth of the other perspective. The deliberative stance gives the insight that one is not bound by her empirical history; it acknowledges the relevance of the theoretical stance and the facts in her history but denies their completeness and decisiveness.

Altogether, Moran (2001) argues that the capacity for exercising both stances of self-reflection and *flexibility between the stances* is an essential part of psychological well-being, and this idea seems to also be true in terms of various pathological cases. Moran does not discuss pathologies distinctly,[57] but I want to point out that the distinction between two stances is useful in analyzing these kinds of cases (and will conduct such analysis in more detail in Chapters 6 and 7).

3.4. Voluntary versus involuntary reflective self-consciousness

In addition to the distinction between 1P and 3P approaches to oneself, I propose that it is useful to draw a further distinction between voluntary and involuntary reflective self-consciousness in order to understand the subtleties of self-conscious thoughts. This distinction could be considered a difference in the initiation of the process of self-consciousness, while the content of consciousness can be the same (one's self). This kind of distinction is made, for instance, by Colombetti and Ratcliffe (2012) in terms of "involuntary and voluntary noematic feelings" and Seigel (2005) in terms of "reflexivity and reflectivity" (in addition, e.g., Gerrans, 2015, uses a parallel distinction between voluntary and involuntary neural activation). *Involuntary self-reflection* (i.e., reflexivity) refers to the passive kind of reflection: it is "an automatic, reflex-like exaggerated self-intimation of a first-order experience" (Colombetti & Ratcliffe, 2012, 146). Instead, *voluntary self-reflection* (i.e., reflectivity) refers to the more active attention (and taking distance) to self, the attention establishing a new relationship between consciousness and its contents. In the exemplar cases above, RSC is exercised voluntarily: one decides to contemplate and direct her mental states and actions. In a way, exactly this voluntary character of RSC makes it so important: one has the capacity not only to react to stimuli but to actively choose how to respond and

---

[57] However, Moran (2001) repeatedly uses an example of an analysand in psychotherapy in describing the differences between different stances to oneself.

deliberate how to act in various situations. As mentioned in Sec. 3.1., the very voluntariness is a feature that distinguishes RSC from MSC (Kriegel, 2003).

However, RSC can occur also involuntarily, without active orienting by the subject. In everyday examples, one is concentrated on doing something that is not focused on herself when she suddenly notices something that makes her shift her attention to herself rather involuntarily. For instance, in a case where one is chopping firewood, she is concentrating on the wood and on her movements with the ax, but when she feels an abrupt pain in her finger, she suddenly turns her attention to herself. After taking the finger as the object of her consciousness and seeing a wound to her hand, she quickly assesses the situation and acts accordingly. Or in another example, one is on her way to an appointment in the city center, quickening her steps and looking ahead. When she suddenly sees her somewhat unkempt reflection in a display window, the reflection rather involuntarily catches her attention, and she slows her pace in order to see the reflection better and to prepare to rectify her appearance. It is noteworthy that these kinds of quick shifts of attention seem also to indicate the usefulness of RSC. By means of RSC, one can quickly utilize new information that is relevant for her and orient herself anew. When the urgent need for self-consciousness is taken care of, one can shift her attention back to the tasks she considers most relevant.[58]

Yet I want to point out that *frequent* involuntary RSC is involved in and is characteristic of several mental disorders. In other words, excessive involuntary self-reflection involves psychological distress. For instance, depression is characterized by excessive negative rumination; a depressed person continuously reflects on herself and her condition. Or a patient suffering from depersonalization tends to check her anomalous feelings all the time. However, such reflection in this case does not lead to normal active self-experience or an operative deliberative stance, but the

---

[58] Also some kind of "periodic involuntary RSC" seems to present, e.g., in mind-wandering in which one is not actively focused on anything particularly (see Sec. 4.1.1. and, e.g., Christoff & Fox, 2018).

self is still felt as distant and object-like—and again taken as the object of reflection (see Sec. 7.2.).

Thus, in terms of the distinction between voluntary and involuntary RSC, voluntary RSC seems to be the variant connected to psychological well-being. Normal involuntary RSC, which is triggered by a quick change in the environment, assists one in orienting herself to the new situation, thus indicating the flexibility and functionality of self-consciousness. However, continuous involuntary RSC is involved in pathological cases in which the self is taken as the object of consciousness continuously but without feelings of being an active self that is capable of affecting her mental states. In these cases, it is exactly the involuntary exercise of RSC without the hoped-for effects that seems to be one of the reasons for the anxiety and distress she experiences.

3.5. Different modes of self-reflection: with and without identification

In addition to differences in stances and ways of initiation, RSC can be exercised in different modes. In this subchapter, I bring out the case of reflecting oneself with or without identification because it further opens the spectrum of self-consciousness. The case resembles the distinction between 1P and 3P approaches to oneself, or deliberative and theoretical stances on oneself, but is considered separately for simplicity and as a further elaborate example of RSC. Within the distinction between theoretical and deliberative stances, both stances seem to involve identification. The stances are used in operations of self-conscious thoughts which often carry some kind of identification within. The deliberative stance especially entails a strong sense of first-person, accompanied by the elements of commitment and motivation. Although the theoretical stance has a more descriptive character, it is not lacking identification altogether; after all, it involves investigation and discovery of identification.

Moreover, in many discussions of RSC in analytic philosophy,[59] the idea of self-identification had enjoyed an exceptional position as an advanced or desirable state (e.g., Frankfurt, 1987; Moran, 2001). In these discussions, identification in action has been contrasted with action that is somehow deficient, being only halfhearted or with ambivalence. In other words, often "identification" and being active have been commended in the discussion of RSC, whereas a lack of identification is associated with mental distress.

However, I want to point out that the selection of the forms of RSC also includes the capacity to direct oneself into her mental states without intentions of identification but instead consciously refraining from identification. This mode of reflection is not to opposite of "wholeheartedness" in traditional discussions but aims to bring out another shade of RSC. Instead of being deficient or inferior to identification, this *capacity* for the non-identification mode of self-consciousness is important and functional since continually heightened self-reflecting can lead to distress, which is a risk for mental well-being, at least if it has become pathological. The non-identificatory mode to oneself is purposely exercised at least in (some) meditative practice, which will be discussed in more detail in Chapter 8. However, it is notable that even the non-identificatory mode requires RSC; self-reflection is used in order to refrain from identification with self-related contents of consciousness in order to sustain a neutral stance on the flow of thoughts.

In order to put together the elaborated distinctions within RSC, we can form a four-square matrix of them (Figure 3.1). In the matrix, the use of RSC is considered in terms of voluntariness and content, and the content is self. In a paradigm case of RSC, self is taken as the object of consciousness voluntarily. One identifies with the content and operates it in theoretical or deliberative stance or alternating between them. In addition, there are cases in which the process of RSC is started involuntarily; while one is doing

---

[59] Or more precisely, in the philosophy of action and will; see fn. 55.

something else, her self suddenly catches her attention. Furthermore, there can be rare cases in which RSC is exercised in order to reach a mental state in which one does not identify with the self as the content of consciousness and withdraws from the contents. Although this kind of voluntary no-content state is not present in everyday consciousness, the state can be experienced within (at least) meditative states of consciousness (see Chapter 8). However, there cannot be a case in which RSC was initiated involuntarily and which would not have self as content—for quite clearly this would not be a case of reflective self-consciousness.
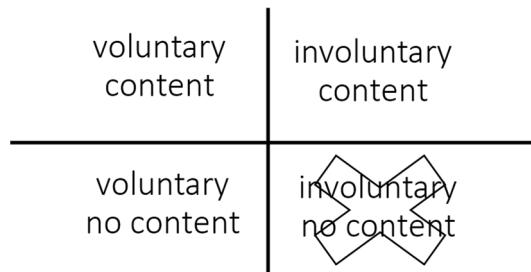
Figure 3.1. The options of reflective self-consciousness (=RSC). Often RSC is exercised voluntarily and it has content. Sometimes, RSC is initiated involuntarily and it has content. In rare cases, the initiation is voluntary and there is no contents of RSC. However, there cannot be a case of RSC that would be involuntary and have no content.

3.6. Reflective self-consciousness and the pattern theory of self

Gallagher's (2013; Gallagher & Daly, 2018) pattern theory of self does not define reflective self-consciousness or take a stand on the aspects it might be comprised of. However, Gallagher mentions explicit self-consciousness in psychological/cognitive aspects, which are certainly substantial for self-reflection. Based on the ideas presented above, I propose that in terms of the pattern theory of self, reflective self-consciousness involves at least minimal experiential, psychological/cognitive, reflective and narrative aspects of self, but social or intersubjective aspects are many times

also considered important for reflective selfhood. This number of aspects indicates that RSC is a complex form of cognition, and the consideration of the pattern theory below can illustrate the shades of self-reflection.

As presented in Chapter 2, MSC is a constituent of RSC (see also Sec. 5.1.). This means that in addition to the contents of self-reflection, RSC involves the minimally self-conscious tacit features of experience, including experiential, embodied, affective and intersubjective aspects. The intersubjective aspects are closely connected to RSC, since the capacity to have oneself as the object of consciousness involves understanding that one can be an object for others' consciousness too. The intersubjective aspects were mentioned in the discussion of MSC (in Sec. 2.4.1.), and their importance for self is clear in that they are connected to both main forms of self-consciousness. In other words, self-consciousness and intersubjectivity can be regarded as two sides of the same coin (e.g., phenomenologists have studied social aspects of self; see, e.g., Zahavi, 2014 Parts II & III). Sometimes consciousness of oneself in relation to others is associated with mirror self-recognition (discussed in the beginning of this chapter; e.g., Gallup et al., 2011). Crucially, intersubjective aspects are necessary in the development of full-blown RSC since the sense of self-for-others develops only in interaction (see, e.g., Ciaunica, 2015, 2019). This developmental importance of intersubjectivity can also be argued in terms of development of language and self-concepts; the linguistic skills that are characteristics of human RSC develop only in a language community (e.g., Baker, 2012).

The psychological/cognitive aspects of self could be considered the core of RSC, since they include the conceptual understanding of self as self and explicit self-consciousness. Traditional theories of self have focused on these aspects, including arguments for psychological continuity and the importance of memory for personal identity (see, e.g., Shoemaker, 2011). Within the distinctions drawn in this chapter, it seems that especially the idea of voluntariness of RSC can be connected to the psychological/cognitive aspects that are employed in (un)volitional thinking of other things as self also.

It is not clear how the psychological/cognitive aspects are connected to reflective aspects, which Gallagher and Daly (2018, 4) describe as capacities to reflect on one's experiences and actions. These capacities are closely related to the notions of autonomy and moral personhood and include the ability to reflect and form second-order volitions about one's desires. The need to specify the sphere of psychological/cognitive aspects can be seen behind Moran's discussion of stances to self-conscious thoughts. Thus, from the concepts employed in this chapter, the idea of deliberative stance was thickest in highlighting that self as a reasoning moral agent shapes herself and commits to her actions. In addition, the idea of identification to oneself seems to involve reflective aspects, since the traditional discussions of identification are linked to discussions of autonomy and personhood that are mentioned in the description of reflective aspects—and apparently have been a motivator to differentiate the reflective aspects from psychological/cognitive ones in the first place. On the other hand, the mode of de-identification could be connected to the psychological/cognitive aspect, since an attitude of withdrawal could also be applied to thoughts that are not focused on oneself.

However, in a broad reading, reflective aspects could involve all cases of self-conscious thoughts. This kind of reading is used in the rest of this dissertation; reflective aspects refer to the cases in which 'I' is used, or which concern the difference between 1P and 3P approaches to oneself. According to this reading, this chapter has elaborated especially the manifestation of reflective aspects in RSC. However, this distinction between psychological/cognitive and reflective aspects is rough and not defined in detail.

Narrative aspects involve the ideas that selves are inherently narrative entities and that our self-interpretations have a narrative structure (Gallagher, 2013, 4). One tends to see her life as a narrative and to consider herself as protagonist of a life story (Dennett, 1991; MacIntyre, 1981; Schechtman, 2011). The narrative aspect is taken as so significant for sophisticated selfhood that many times the level of having RSC is labeled as the 'narrative self' (e.g., Gallagher, 2000; Mackenzie, 2014). However, the degree of narrativity is debated (e.g., Schechtman, 2011, gives a nice

overview). Some philosophers emphasize that for being a self it is essential to have a coherent life story, while others deny these claims and see themselves more as "episodic" persons (see, e.g., Strawson, 2004).

The definition of RSC, which I embrace in this dissertation, does not highlight the narrative aspect. RSC simply refers to the capacity to take oneself as the object of one's cognition and to operate the contents of consciousness that are about oneself. Narrative aspects or skills, which well forth from linguistic skills generally, are important to recognize as a manifestation of RSC but a rich self-narrative is not underlined. For the narrative is not necessary for RSC, but self can be an object of consciousness in simple, particular cases without a rich narrative. In other words, when a self is taken as the object of consciousness, the content of consciousness can be a complex autobiographical narrative, or one's mirror reflection, or an effortful movement of a body part at a particular moment—all these cases are instances of RSC. In a paradigm case, RSC requires cognitive/psychological and reflective aspects that refer to conceptual capacities, general higher-order reasoning skills and the capacity to formulate and grasp narratives. However, a precise self-narrative is not necessary in reflecting oneself. This can be seen, for instance, in the example of the amnesiac person who wakes up in the dark room: the person can reflect on herself (as herself) and deliberate her action without an extensive narrative of herself. Certainly, narrative aspects are linked with cognitive-psychological ones, narratives are a sophisticated form of self-reflection, and a coherent self-narrative is related to mental well-being—but RSC can also be manifested without narrativity. Zahavi (e.g., 2014, 90) also emphasizes that human selfhood has an eliminable normative dimension, which is linked to notions such as reflection and responsibility, but the dimension does not necessarily take a narrative form and generally, a first-person perspective is a prelinguistic presupposition for any narrative practice.

After all, the simple definition of RSC, which is used here, can be seen as an advantage of the notion. The definition of RSC simply as a skill for cognition of oneself enables the investigation of

this form of selfhood without taking a stance on more metaphysical theories of self (e.g., in what sense reflective self is narrative), and the notion can be applied within different theories. For instance, RSC can be described in terms of the analyses of the uniqueness of first-person reference, and RSC can be considered as a necessary component for narrative self.

3.7. Summary

In this chapter, I defined RSC as the capacity to take oneself as the object of one's cognition and to think of oneself as oneself. This capacity involves complex linguistic and cognitive skills, and renders us moral agents. The first step in understanding the significance of RSC was to study classical analyses, which argue that self-conscious thoughts, thoughts that refer to oneself by the use of 'I', have unique epistemic and motivational features. The thinker of a self-conscious thought necessarily knows and understands that she is referring to herself, and this understanding relates the subject to her action in a special motivational way that immediately matters to her practical reasoning.

However, I argued that the mere remarking of self-conscious thinking is not sufficient analysis of the shades of RSC, but more elaborate conceptual tools are needed in order to grasp the manifold forms of self-reflection. This chapter presented three conceptual distinctions within the different manners in which one can reflect on herself; deliberately or theoretically, voluntarily or involuntarily, and with or without identification. These distinctions are not exclusionary but complementary; they can be seen as answering different questions on the exercise of RSC. The first distinction between theoretical and deliberative stances concerns the attitude or position with which one examines herself (with which stance does one employ RSC?). In theoretical stance, the self is taken as an object of description and explanation, whereas in deliberative stance, the self is taken as an agent of practical reasoning that forms commitments, decides to do something, and shapes its attitudes. The second distinction between voluntary and involuntary RSC refers more to the initiation of the self-reflection

(how did the process of RSC start?), not to the content or stance of self-reflection. The third distinction between identification and non-identification with oneself concerns the mode in which the contents of self-reflection are observed (does RSC involve identification with the self as the content?) and also includes cases in which RSC is used in order to refrain from self-identification.

In terms of the pattern theory of self, RSC involves at least experiential, psychological-cognitive, reflective, and narrative aspects. The first mentioned three aspects correspond to the capacity central in the definition of RSC, whereas narrativity seems to be a further capacity that is enabled by the cognitively high level of reflective capacities. In addition, I pointed out that RSC is important for mental well-being. Self-conscious thoughts have a practical and motivational side that is directly related to action. Deliberative stance especially has an important role for well-being since it assists in shaping attitudes and in initiating and guiding action. In addition, the capacity for voluntary and flexible shifting between the stances of self-reflection is significant for mental well-being. Without this flexibility, the capacity of RSC becomes dysfunctional and loses its characteristics to regulate and direct mental states and actions.

4. The neural mechanisms for self-consciousness

This chapter's intention is to bring neurophilosophy into the discussion of self-consciousness and to promote cooperation between philosophy of mind and empirical mind sciences. On the one hand, the value of conceptual analysis for sciences is indicated: clear concepts and theories help scientists to be clearer on their work. On the other hand, (Zahavi-style) phenomenological concepts of self-consciousness are applied in practice and discussed in the light of neuroscience, and it is considered how an appeal to the results of empirical science can be important for philosophers of mind. An empirically informed viewpoint assists in opening the concepts of self, and empirical studies can "show" features that are involved in self-consciousness. In other words, this chapter makes progress from phenomenological concepts discussed in earlier chapters to a multidisciplinary study of self, in which different forms of self-consciousness can also be tractable within empirical sciences.

I start the chapter by giving an overview of the data and the conceptualizations of self within empirical science, and applying the distinction between minimal and reflective forms of self-consciousness to empirical research. The examination of the two forms of self-consciousness can assist in forming a clear conceptual map of selfhood that is needed in carrying out detailed empirical research. The distinction between two forms of self-consciousness is significant in the multidisciplinary field which means, in terms of neuroscience, that the two forms are based on different kinds of neural processes. This implies that the tracking of the neural correlates of two types of self-experience asks for different empirical paradigms.

After the focus on empirical studies, I turn to a more philosophical analysis of self-consciousness. In order to respect the multidimensionality of self, I endorse a framework that involves multiple levels of explanation. These levels describe steps that are taken in the process of the scientific study of self-consciousness. The first step is to formulate concepts that grasp the richness of self-experience. Then the concepts are operationalized in functional

terms that allow for the identification of types of neural processes that are involved in these terms. Then the operationalized concepts can be applied in designing neuroimaging studies that examine the relevant neural activation. However, these levels of explanation can be properly discussed only (in Sec. 4.2.) after the consideration of neural studies of self (Sec. 4.1.). I consider these levels in terms of both minimal and reflective self-consciousness. I examine how the concepts of self-consciousness elaborated in earlier chapters fit to empirical studies and point out how empirical results also highlight the interconnectedness of MSC and RSC.

4.1. The neural basis of self-consciousness

In this subchapter, I examine the neural profile of self-consciousness and the focus is on empirical studies. Only after forming a picture of the wide neural profile of self can I apply more philosophical analysis in the next subchapter. Generally, self-consciousness emerges from distributed interactions among several neural networks (see, e.g., Frewen et al., 2020; Zhao et al., 2013). In recent years, probably the most-discussed neural network related to self has been the so-called cortical midline structures (CMS in brief). These midline structures comprise the anterior cingulate, medial prefrontal cortex, posterior cingulate cortex, and precuneus/retrosplenial cortex. The CMS system has been considered to be of key importance for "self" as opposed to "other" processing. In many studies, the processing of self in CMS has also been conjoined with the default-mode network (DMN in brief; see, e.g., Buckner et al., 2008; Northoff & Panksepp, 2008; Schneider et al., 2008; Wicker et al., 2003). The DMN of the brain exhibits functional activity during rest and shows deactivation during externally directed, attention-demanding tasks (for this pattern, DMN components have also been referred to as task-negative brain regions). The DMN includes the precuneus/posterior cingulate cortex, medial prefrontal cortex and bilateral temporoparietal junction. However, in addition to CMS, self-processing has been associated with regions in the so-called "mirror neuron system," including both frontoparietal (inferior frontal gyrus, pre-central

gyrus, precuneus, supramarginal gyrus, inferior parietal lobule) and limbic regions (anterior insula and anterior mesial frontal cortex), which respond equivalently to specific goal-directed actions whether they are performed by self or others (e.g., Zhao et al., 2013). In addition, inter-hemispheric connectivity seems to play an important role in self-processing (e.g., Zhao et al., 2013). Further, some studies have connected self as agent to sensorimotor networks (e.g., Ebisch & Aleman, 2016; Legrand & Ruby, 2009) or self-processing to the salience network (e.g., Sui & Gu, 2017).

It is notable that the focus of research is more and more on interactions within different neural networks—that is, on extensive patterns of neural activation—not only on single areas.[60] The involvement of the very wide neural networks in the processing of self evinces the major role that selfhood has in our cognition and experience. Since such large-scale neural processes are involved in grounding self-consciousness, it is crucial to be able to conceptually and theoretically distinguish between different aspects of self-consciousness in order to make sense of the empirical data. Currently, various concepts of self have been presented in order to work out the aspects of the neural profile of self. Since the concepts of self vary in different studies, the discussion between studies can get complicated and might lead to confusion. The confusion between notions that are used reveals that there still is conceptual work to be done in elaborating the multidisciplinary study on self.

I propose that the conceptual distinction between two forms of self-consciousness offers a clear starting point for an empirical

---

[60] Nowadays, there is an enormous amount of empirical data related to selfhood, and thus, all the relevant research cannot be discussed here. The studies considered here are limited to the spatial neuroanatomy of neural networks and do not address the neural basis of self, for example, in the temporal neuroelectrophysiological or transmitter domains. For summarizing electromagnetic and transmitter studies, see, e.g., Lou et al. (2017). According to Lou et al. (2017), electromagnetic and transmitter manipulation have demonstrated that a paralimbic network of medial prefrontal cingulate and medial parietal cingulate cortical "hubs" is instrumental in generating self-awareness. Transcranial magnetic stimulation (TMS) targeting these hubs impedes different aspects of self-awareness with a latency of 160 ms. Further, the "network is linked by ~40 Hz oscillations and regulated by dopamine. The oscillations are generated by rhythmic GABA-ergic inhibitory activity in interneurons with an extraordinarily high metabolic rate." For a review of EEG studies; see, e.g., Knyazev (2013).

analysis of self too. The distinction can assist in understanding the diversity of neural activations that are related to self by sorting out the different sub-features of self that are involved. Especially the distinction between self-specifying processes (SSP in brief) and self-related processing (SRP in brief) (Christoff et al., 2011a; Legrand & Ruby, 2009; Thompson, 2014) strongly resembles the notions of minimal and reflective self-consciousness. The distinction has a theoretical background in phenomenological philosophy (Legrand, 2006, 2007), and it is used in enactivist theories too (e.g., Thompson, 2014). The distinction is drawn explicitly in order to encompass many sides of self-experience instead of a narrow definition of self, which would simplify the self only to particular self-related contents. Indeed, the distinction invokes James's classic distinction between 'I' and 'me'; 'I' referring to experiencing oneself as (self-specific) subjective knower and agent, 'me' referring to experiencing oneself as an object of (self-related) perception or self-attribution (Christoff et al., 2011; Legrand & Ruby, 2009). RSC is clearly present when self is targeted as 'me' since self is taken to the object of one's consciousness. This can also be seen in the definition of self-related processing as "processing requiring one to evaluate or judge some feature in relation to one's perceptual image or mental concept of oneself." MSC, in turn, characterizes self as subjective agent and knower and is connected to self-specific neural processing that is defined as "any process that specifies the self as subject and agent by implementing a functional self/non-self distinction." This definition applies also to cases in which self is not the object of consciousness but present as the subject of consciousness, consistently with the idea of MSC (Sec. 2.1.2.; Christoff et al., 2011; Legrand & Ruby, 2009). For the clear similarities in the notions, the distinction between self-specifying and self-related neural processing is here considered as a tentative neural basis for minimal and reflective self-consciousness.

Generally, self-related neural processing, which is the neural basis of RSC, is connected to a subset of midline cortical regions (MPFC and precuneus/PCC). Self-related processing involves the interactions between frontal lobes that are involved in general evaluation of actions and temporal (and parietal) lobes that are

needed for memory. Instead, the neural basis of minimal self-consciousness is connected to self-specific neural processes that arise from the integration of efferent, reafferent and affective processes that characterize, for instance, sensorimotor integration (involving activation of sensorimotor cortical areas on the parietal lobes) and homeostatic regulation (involving e.g., insular activation). The connection of self-specific processes and MSC is supported by the notion that these processes are also active during conscious mental states in which the self is not the object of reflection but is involved as the subject of consciousness and agent initiating actions.

The distinction between self-specific and self-related processes also resembles some other distinctions drawn within the neural processing of selfhood. For instance, Ebisch and Aleman (2016) distinguish between intrinsic and extrinsic self-networks. Intrinsic self is associated with self-reflection and self-referential processing, but also with the concept of core self, which integrates interoceptive and exteroceptive stimuli (Damasio, 1999; Northoff & Panksepp, 2008). Intrinsic self overlaps with the neural activity in DMN, involving autobiographical processing (the posterior cingulate cortex and hippocampus) and interoceptive processing (the anterior insula). Extrinsic self, in turn, emerges from one's participation in her environment. This crucial aspect of self is grounded in bodily interactions with the external world and mediated by the sensorimotor network. The concept of extrinsic self is also explicitly related to the concept of pre-reflective minimal self that is present in immediate experience (Gallagher, 2000; Sass & Parnas, 2003). In addition, the distinction between SSP and SRP is mentioned in many studies of self, even if the studies have not further employed the distinction in their design (e.g., Frewen et al., 2020; Sui & Gu, 2017; Zhao et al., 2013).

4.1.1. RSC: self-related neural processing

4.1.1.1. Evaluation network
As proposed above, self-related processing can be considered the neural basis of RSC, and the majority of empirical studies have

been focused on these processes. In other words, most neuroimaging studies targeting self have investigated the self by using diverse self-related stimuli or tasks (see, e.g., the meta-analyses by Frewen et al., 2020; Gillihan & Farah, 2005; Legrand & Ruby, 2009; Northoff et al., 2006). Using functional magnetic resonance imaging (fMRI) and positron emission tomography (PET) techniques these studies have traced the cerebral correlates of, for example, recognizing one's own face, attributing an action to oneself, detecting one's own first name, recalling personally relevant information, or assessing one's own personality, attitudes, physical appearance, feelings, or bodily states. Legrand & Ruby (2009, 252–54) remark that although the cognitive tasks have varied in the studies, they all targeted a common and unique object of investigation: the self. However, some of these studies have been carried through without a precise definition of self. In addition, these studies employ a large variety of cognitive tasks and stimuli activating many cerebral regions, thereby also making the meta-analysis of them challenging. The great number of studies on SRP is considered below first by referring to the pioneering meta-analysis by Legrand and Ruby (2009) and then by taking a look at more recent studies.

The meta-analysis by Legrand and Ruby (2009) is remarkable because it involves important theoretical insights concerning self, in addition to summarizing results of several brain-imaging studies of self. The theoretical points include a novel conceptual formulation of the dimensions of self; this formulation assists in putting together the previous results and can be employed in future empirical paradigms. The meta-analysis by Legrand and Ruby (2009) aimed at finding regularities among the diverse results in the neural studies investigating self and revealed a wide cerebral network, which they termed the E-network. The E-network comprises the medial prefrontal cortex, precuneus/posterior cingulate gyrus, temporoparietal junction and temporal pole; it covers at least frontal and temporal parts of CMS (the term 'E-network' is in order to stress the conclusion in the paper in question, but for simplicity, the E-network can be identified with DMN, as done in Christoff et al., 2011). The E-network was

supposed to disclose the neural correlates of the self; it was repeatedly reported to be activated in neuroscientific studies of the self, which targeted to reveal a neural substrate that is systematically more activated for self than for non-self. However, Legrand and Ruby (2009, 254–8) remarked a striking overlap between neural correlates of self and others' mind representation. In other words, the meta-analysis indicates that the E-network is not unique for self-processing. Instead, regions of the E-network are sometimes more activated in self-related task and sometimes more activated in others-related tasks.

In order to make sense of this overlap, Legrand and Ruby (2009, 258–68) invoke the common denominator of all the tasks recruiting the E-network and conclude that the activity of the network represents a general cognitive ability of inferential processing that uses information recalled from memory. Legrand and Ruby call the combination of these two cognitive processes "evaluation," which is abbreviated in the term 'E-network' (where 'E' stands for evaluation). Evaluation involves generic mental operations that are used to draw conclusions on the basis of premises and rules.[61] Activation of the medial prefrontal cortex can be associated with inferential processes, such as deductive and inductive reasoning. At the same time, activations of the medial parietal cortex, temporoparietal junction, and temporal pole can be connected to memory recall, providing a premise for the inferences. In addition, the PCC and TPJ can be associated with general attentional processes (noted, e.g., in Christoff et al., 2011). Thus, describing the E-network, or some parts of it, as self-specific is unwarranted. Evaluation is used in a multitude of cognitive tasks: it is required to answer not only a question about oneself but also questions about one's sister, one's colleague, or one's cat, for example. Legrand and Ruby (2009, 266–70) conclude that activity in the different regions of the E-network is not modulated by the person targeted in the task (self versus other), but rather by the form of relevant inferences and memory recall. Furthermore,

---

[61] Legrand and Ruby (2009) add that most of the time, these inferential processes are performed implicitly.

Legrand and Ruby notice that the E-network is not preferentially activated for self and other mind representations since it is also employed in reasoning, memory, and the resting state. Thus, classical neuroscientific studies of the self involve processes of evaluation that are not self-specific.

This conclusion of Legrand and Ruby (2009) is significant because it reveals that the neural correlates that were claimed to reveal mechanisms underlying especially the self actually revealed rather generic evaluation mechanisms. Just this kind of conceptual analysis is useful for empirical research since it brings out the complexity in explanandum, shows the importance of precise definitions, and can resolve seemingly contradictory findings. The results of the neural experiments targeting the 'self' as a common object of study pointed out different brain areas simply because they had targeted different features of self. Further, these studies focused only on some forms of evaluative self-related processing of RSC but left out MSC. Altogether, the E-network is significant for RSC, but it is important to acknowledge that the network is not self-specific: although it is activated when self is attended as an object of evaluation or perception, it is not exclusively about the self and thus, does not capture all senses of selfhood.

4.1.1.2. Differences in neural activation for different forms of self-reflection

However, further studies have indicated that it is possible to differentiate activation patterns of the E-network that are related to self particularly. Although the same network operates the evaluation of self as well as evaluation of other people and abstract things, the case is not simply that the network activates exactly the same way in different cases. But activation patterns related to different objects of consciousness can be dissociated (see, e.g., Frewen et al., 2020). For instance, differences have been found between SRP and other-referential-processing, including, for instance, greater response for SRP in MMPFC (middle medial prefrontal cortex) (Araujo et al., 2013; Denny et al., 2012; Frewen et al., 2020; Qin et al., 2012).

In addition, neuroimaging research has been able to dissociate neural correlates for different forms of self-related processing. This evinces that more fine-grained notions of self are used in empirical research than in the time of the meta-analysis of Legrand and Ruby (2009). The elaborated concepts have also rendered more elaborated research frames. For instance, the extensive meta-analysis of self-referential processing by Frewen and colleagues (2020) makes several points in order to classify the diverse results of the brain imaging studies of self and to elaborate the conceptual and methodological framework for future studies. Frewen and colleagues (2020) use the general notion of 'self-referential processing', defined as "how we respond to stimuli that reference ourselves" (164) or as processing in which one attends towards herself (165). These characterizations involve the idea that the self is taken to the object of consciousness and thus, self-referential processing (SRP) can be connected to the exercise of RSC.[62]

In order to highlight the diversity in RSC and to bring out that different forms of self-reflection are related to different neural activations, I will below consider three distinctions that Frewen and colleagues (2020) draw. The first distinction within neural activity is between verbal and non-verbal self-consciousness, the second distinction adds the resting-state to the first distinction, and the third distinction is restricted to different functions of self-reflection that are processed in the medial prefrontal cortex. These distinctions are interesting since they can be connected to conceptual distinctions and shades of self-reflection discussed especially in Chapter 3 here.

Frewen and colleagues (2020) make a major distinction between SRP as occurring verbally (V-SRP) and non-verbally (NV-SRP, i.e., bodily self-consciousness), and highlight a marked division of labor within neural activation between these verbal and

---

[62] Since the terms are rather concurring, the abbreviation SRP used here can refer to either 'self-related processing' or 'self-referential processing.' Generally, the abbreviation is used here in the sense of 'self-related processing' following Legrand & Ruby (2009) and Christoff et al. (2011), but in referring to the study of Frewen et al. (2020), the abbreviation describes their notion of 'self-referential processing'. For discussion of the terms, see, e.g., Christoff et al. (2011b); Northoff (2011).

bodily modalities of SRP. The meta-analysis highlights, for instance, the study of Araujo and colleagues (2015) that distinguishes between V-SRP and NV-SRP. In these experiments, V-SRP was triggered by conditions involving questions about facts from the subject's biographical history (e.g., "Are you a student?") and about personality traits (e.g., "Does the word 'honest' describe you?"). Instead, NV-SRP was directed, for instance, by asking questions requiring interoception (emphasizing internal sensations or interior bodily self-consciousness, e.g., "Do you feel hungry?"), or exteroception (emphasizing external sensations or exterior bodily self-consciousness, e.g., "Are your legs wet?"). Answering these different questions involved specific patterns of neural activation showing that V-SRP involved greater activation in the posterior cingulate cortex (PCC), medial prefrontal cortex (MPFC), temporal poles, and posterior inferior parietal lobe. In turn, NV-SRP involved greater activation in the posterior medial superior parietal lobe, anterior temporoparietal junction, insula, frontal operculum, and middle frontal gyrus. Altogether, Frewen and colleagues (2020) propose that CMS are more active in V-SRP and the insula more active in NV-SRP.

Further, Frewen and colleagues (2020) make a three-part distinction between V-SRP, NV-SRP, and resting state and hypothesize that the resting state exemplifies a state of possible convergence between V-SRP (e.g., introspection) and NV-SRP (e.g., interoception). Frewen et al. (2020) point out that the contents of thoughts are supramodal, consisting of a balance between different kinds of stimuli. One type of stimulus is not independent of others; rather it is a matter of degree which type predominates over the others. This implies that the neural regions of interest should not be understood to act alone but as part of distributed and complex neural networks. Thus, Frewen and colleagues (2020) encourage investigation of the conjunctive activation of processes that are involved in V-SRP and NV-SRP as compared with the resting state, since it is interesting how different forms of SRP are being integrated, facilitating or hindering other types of SRP. These remarks are interesting since they can distinguish between three kinds of self-reflection and underlying neural activation. Self-

reflection does not simply activate the E-network, but different patterns of activation of the network are associated with different types of thinking of self. At the same time, the different activation patterns are related to each other; for instance, the shift between activation of V-SRP and NV-SRP is mediated by resting state activation.

In addition, Frewen and colleagues (2020) propose a specific quadripartite model of different levels or functions of SRP in terms of the neuroanatomical framework of MPFC (medial prefrontal cortex). In this framework of MPFC, the proposed four forms of SRP can be localized from inferior to superior: emotional, neutral, observing and executive self. All these instances of self involve reflective self-consciousness, but in different shades (see Sec. 4.2.2.). In this model, 'emotional self' is attributed to the VMPFC (ventral medial PFC), while 'neutral self' is attributed to the MMPFC (middle medial PFC). These both are connected to bottom-up processing, but their activations differ in the type of the stimuli: VMPFC is more engaged in salient or arousing emotional stimuli while MMPFC is activated for more neutral stimuli. Since the resting or default state involves MMPFC especially, Frewen and colleagues (2020) attribute the activation of MMPFC to comparably affectively neutral content, for the "default" state is by definition a state of low arousal and an emotionally neutral state. Frewen and colleagues (2020) consider that this "default self" underpins the first-person (i.e., egocentric) perspective, which also seems to be default state in humans. This implies also that adopting an allocentric third-person perspective must involve a deviation from the default state, thereby being a task that requires some level of top-down control.

Within the quadripartite model, Frewen and colleagues (2020) connect the top-down processing to DMPFC (dorsomedial PFC), where taking a third-person perspective is attributed to an 'observing self', and central executive control is attributed to an 'executive self'. The observing self is needed for taking a deidentified allocentric third-person perspective in SRP and emotional processing; it is involved in downregulating negative emotions, upregulating positive emotions, and in simply facilitating a kind of decentered open monitoring toward self-referential

111

stimuli. The term 'observing' self is descriptive also for its passive nature in behavioral orientation, as occurs, for instance, in mindfulness meditation. Executive self, instead, is more active in nature; it is the source of cognitive control in self-engagement with the surrounding world. Frewen and colleagues (2020) ascribe the representation of executive self to the yet more superior aspects of DMPFC (in area 8, the frontal eye fields), which is part of the central executive network together with the DLPFC (dorsolateral PFC) and needed, for instance, in coordinating and executing eye movements. Thus, Frewen and colleagues (2020) propose that the function of both of these DMPFC nodes is coupled to the subjective experience of agency but in different contexts: passive observation versus active execution.

Further, Frewen and colleagues (2020) describe how bottom-up SRP typically takes place in the background of awareness from where it sometimes is taken in the foreground of consciousness by top-down SRP. Resting state is characterized by broadly-focused, involuntarily and spontaneous processing of different self-referential contents. Frewen and colleagues (2020) distinguish between three bottom-up SRP (or more metaphorically, three attentional "spotlights" which are dim but always "on" in the theater of consciousness): interoceptive NV-SRP (e.g., heartbeat, breathing rate, connected to activation in the posterior insula), exteroceptive-proprioceptive NV-SRP (e.g., body position or location, connected to activation in the R-IPL), and introspective V-SRP (i.e., thoughts, connected to activation in L-IPL). These bottom-up processes are continuously active, performing endogenous monitoring functions. Instead, top-down SRP (i.e., attentional spotlights which are often "off," but characteristically bright when turned on, connected to DMPFC and the frontoparietal control network) involves cognitive control and can be consciously focused to whichever of these processes in order to examine or regulate it or to initiate action.

Altogether, the studies resulting in describing the E-network and SRP are needed for understanding the neural correlates of self. The evaluation mechanisms of the E-network are crucial for thinking in which one's self is the object of consciousness and which thus meets the definition of RSC. At the same time, the conceptual

and theoretical study of the SRP is significant in order to establish which neural processes are connected to various forms of self-reflection or self-evaluation. For instance, the quadripartite model is a useful elaboration of the neural profile of self since it gives an integrative and more fine-grained picture of the processes involved in SRP than is presented in most studies.

Understanding the neural mechanisms of RSC is valuable for the conceptual study of self, since it can indicate which features of self are involved in different modes of self-reflection. Detailed information of the neural base of self-consciousness can support theoretical ideas of the particular types of experience of self-reflection. On the other hand, unexpected findings of neural mechanisms of self-consciousness can invite more theoretical work in explaining them. The above examined empirical results do not involve anything radical for the conception of RSC presented in the earlier chapter but generally support the idea that self-reflection involves many shades and that it is important to be able to make fine-grained distinctions within RSC. For instance, the 'executive self' might be connected to self-conscious thought in the deliberative stance that is linked to action; this is supported by the top-down nature of the executive self and its connection to the executive neural network (more in detail in Sec. 4.2.2.).

4.1.2. MSC: self-specific neural processing

Thus far, MSC has been targeted in a minority of empirical studies. However, it would be advantageous to acknowledge MSC within empirical research, since that acknowledging can assist in forming a complete picture of selfhood and the role self plays in neural processes. MSC is a constituent of every conscious experience, and since empirical sciences aim at tracking the neural correlates of experiences, the variations in MSC clearly are a factor that could be included in empirical analysis. However, neural processes of MSC are integrated with processes producing self-related contents, and different research questions can be interested in different features in this network of processes of self. When the research interest is strictly in the specific content of consciousness, MSC is not in the focus. For instance, when the research targets the

difference between thinking of oneself and thinking of others, the interest is in the difference of the neural correlates of different contents. However, consciousness of both contents involve also MSC: SSP is a part of their neural correlates. In other empirical settings, variations in MSC can be emphasized in the research topic. For instance, studies that are interested in the affective components on the background of contents of consciousness involve MSC in particular.

Below especially the notion of 'self-specific processing' is discussed since it explicitly directs the self as subject. Self-specific processing is concerned in terms of sensorimotor processes, homeostasis and emotion regulation (Christoff et al., 2011; Legrand & Ruby, 2009). Generally, the neural basis of MSC involves bodily and affective dimensions of self that can be connected to the sensorimotor cortical network (which is also involved in generating the sense of agency; see, e.g., Ebisch & Aleman, 2016; Legrand & Ruby, 2009) and to insular structures (highlighted, e.g., in notions of 'sentient self' (Craig, 2010; 2009), 'core self' (Damasio, 1999) and 'embodied self' (Seth, 2013)).

It has been argued that the above-described self-related processing approach on the empirical study of self-experience misses the self as the subject and agent (Christoff et al., 2011; Legrand & Ruby, 2009). In order to overcome difficulties linked to the partially defined notion of self, Legrand and Ruby (2009, 272–6) formulated the most basic notion of self: the concept of 'self-specificity' distinguishes self from non-self and specifies the self as subject and agent. As already described (in Sec. 2.1.), experiencing oneself as an embodied agent requires a subjective perspective that relates the subject and objects of experience. Holding a perspective does not require an introspection or a reference to self-related contents, and thus, the concept of self-specific subjective perspective targets MSC instead of RSC.

A paradigmatic example of a self-specifying process is sensorimotor integration, which refers to "the systematic linkage of sensory and motor processes in the perception-action cycle" (Christoff et al., 2011, 105; Legrand & Ruby, 2009; Thompson, 2014). The perception–action cycle determines the perceiving

agent by relating the agent's actions and their consequences in the world. The distinction between self and non-self is implemented by the differentiation between sensory changes that arise from the environment (non-self) and that arise from one's motor actions (self). This distinction is made in the central nervous system by relating the efferent (motor) and afferent sensory (sensory) signals. The integration of efferent and reafferent information creates a sensorimotor loop. Reafference refers to afferent signals that arise as a result of the organism's own efferent processing. That is, reafference is linked to one's own action by definition. For instance, the perception of a flying bird triggers motor commands for the production of eye movements to follow its flight path (reafference is used in assessing the appropriate amount of movement). Crucially, this perceptual act is characterized not only by a given content (the bird) but also by a self-specific perspective (I am the one watching the flight of the bird). In other words, the self-specific sensorimotor perspective rests on the integration of perceptual and motor processes; the perception of the world occurs from one's agentive perspective. [63]

Given that self-specificity is anchored to sensorimotor processes, Legrand and Ruby (2009, 278–9) propose that the relevant neurophysiological mechanisms involve sensorimotor and/or motor-related cortices. This proposition is supported by empirical results that indicated that the activity of the somatosensory cortices is modulated according to the perspective that the subject takes. The somatosensory cortices (postcentral gyrus or insula) are activated more by a first-person perspective, which is in contrast to conditions that involve taking a (mental or spatial) third-person perspective (e.g., Farrer & Frith, 2002; Vogeley et al., 2001). Furthermore, increased activation in somatosensory cortices for a first-person perspective has been found outside of any sensorimotor context (e.g., using stories [Vogeley et al., 2001] or

---

[63] Legrand and Ruby (2009) observe that their proposal is more basic than, for instance, Frith's model (e.g., Frith et al., 2000) because it invokes only sensorimotor processes and not informational contents. Rather, Legrand and Ruby's view resembles, e.g., Hurley's (1998) view.

conceptual facts [Ruby & Decety, 2003, 2004] as stimuli). This increased activity in somatosensory cortices coheres with the idea that basic self-specific processing is grounded in sensorimotor integration. The significance of sensorimotor parietal processing for self-consciousness is also supported by lesion studies. For instance, Philippi and colleagues (2012) reported a patient who preserved self-awareness with wide damages to the insula, AC, and MPFC but without damage to the parietal lobes.

In addition, self-specific processes specify a self as a bodily and affective agent. Thus, empirical information of these processes coheres with idea of MSC involving embodied and affective aspects of self (presented in Chapter 2). A growing interest in these features of self can be seen, for example, in studies of insula (Craig, 2009; 2010), interoception (Seth, 2013; Tsakiris, 2017), and the salience network (e.g., Seeley et al., 2007). The empirical evidence indicates that activation of the anterior insula (typically conjoined with activation of the anterior cingulate) is associated with any and all feelings and provides affective coloring of the self (Craig, 2009, 2010; Seth, 2013).

Christoff and colleagues (2011) point out the bodily and affective dimension of self in noting that basic homeostatic regulation involves self-specifying processes. Homeostatic regulation sustains the self–non-self distinction in life preservation: an organism strives to ensure its survival through changing its (internal and external) conditions by continually coupling afferent and efferent signals (Craig, 2009; Damasio, 1999; Parvizi & Damasio, 2001; Thompson, 2007). Afferent signals convey information about organism's internal state while corresponding efferent regulatory processes keep afferent states within a tight domain of possible values (Craig, 2009; Damasio, 1999; Parvizi & Damasio, 2001). These reafferent–efferent loops are implemented in somato-autonomic adjustments; these loops run from spinal nuclei to brainstem nuclei and midbrain structures and are modulated by the hypothalamus, insula (sensory) and anterior cingulate (motor) cortices. This interoceptive homeostatic system maintains the body's integrity (self) in relating with the environment (non-self) and supports implicit feelings of the body's internal

condition in perception and action. Thus, the homeostatic system specifies self as a bodily and feeling agent entertaining an affective perspective based on the inner feelings of the body. As the concept of MSC entails, this subjective perspective is also present in RSC. This is supported, for instance, by empirical findings that both perceptual (a mirror image of the subject's face) and conceptual (narrative condition in which a participant generated self-referential words) self-focus also increase interoceptive neural processing (Ainley & Tsakiris, 2013; Frewen et al., 2020).

Further, Christoff and colleagues (2011) highlight self-specifying processes for the self as a cognitive–affective agent. This cognitively higher level of self-consciousness subsumes the self-experience of an embodied agent. Cognitive control employs sensorimotor processes, thus extending the perception–action cycle to higher levels of cognitive integration. According to Christoff and colleagues (2011, and originally conceived in Norman & Shallice, 1986; Posner & Rothbart, 1998), the cognitive control of attention is closely linked to self-regulation, which includes the self-experience of being a cognitive agent. Further, the self-experience of being a cognitive–affective agent is present in effortful control of emotion in affectively arousing situations. The case of emotion regulation is considered below, since it broadens the notion of SSP to complex cognitive processes that explicitly involve both evaluation and emotions.

Emotion regulation has often been connected to self-related processing since it involves activation in CMS (Northoff, 2005; Ochsner & Gross, 2005), but Christoff and colleagues (2011) propose that self-experience in emotion regulation is also self-specifying. Studies on emotion regulation make a distinction between two forms of the regulation: a deliberate or voluntary form, and an implicit or incidental form (Berkman & Lieberman, 2009; Gross & Thomspon, 2007; Lane, 2008). Deliberate emotion regulation involves reappraisal; in order to change one's emotional response, the meaning of a stimulus is reinterpreted (Gross & Thomspon, 2007; Ochsner & Gross, 2005). This appraisal employs the same cognitive control mechanisms that are required for attention-demanding tasks, and which neurally involve dACC and

lateral PFC regions (Ochsner & Gross, 2005). These regions subserve explicit reasoning about how one's emotional response to a situation can be modified. For instance, when one is presented with a picture of a burn victim in a hospital bed, her immediate emotional response is distress or sadness. However, she can modify these emotions by focusing on possible positive aspects, such as the victim survived, had good treatment, and is successfully progressing toward a healthier state. Maintaining these re-descriptions involves perceptual and associative-memory systems, which further signal to subcortical appraisal systems, such as the amygdala, and thus indirectly produce a modification of the original emotional response. Christoff and colleagues (2011) propose that this kind of regulatory–evaluative loop implements a functional self–non-self distinction in which the self initiates the effortful reappraisal process that targets the emotional scene (non-self). Thus, self is specified in emotion regulation as the cognitive–affective agent who aspires to reinterpret and thereby control her emotional response. I want to point out that in terms of self-consciousness, this presents a case of top-down interaction between minimal and reflective self-consciousness in which self-reflection shapes the affective features of MSC (see Sec. 5.3.2.).

In addition to regions involved in cognitive control (dACC and lateral PFC), deliberate emotion regulation recruits DMPFC (Berkman & Lieberman, 2009; Ochsner & Gross, 2005; Phillips et al., 2008), which is connected reflective self-consciousness that enables higher-level metarepresentations of one's experience. Christoff and colleagues (2011, 109) propose that DMPFC exerts "a biasing influence on emotion processes" by maintaining such emotion-specific metarepresentations and linking them to VMPFC that is involved in the original emotion process. Thus, the DMPFC forms another regulatory–evaluative loop that specifies the self as a cognitive–affective agent of effortful emotion regulation. It is interesting that also Frewen and colleagues (2020; Sec. 4.2.1.2.) ended up with a rather similar conclusion for the significant role of DMPFC in their quadripartite model. The model highlights the activation of DMPFC for agentic "I" or self-as-subject (objective and

executive self), in contrast to MMPFC (neutral self) and VMPFC (emotional self).

Consistently with the notion of SSP, MSC has also been linked with the sense of agency at a more bodily level. For instance, Ebisch and Aleman (2016) use the concept of extrinsic self in describing pre-reflective minimal self and connect the neural base of extrinsic self to the sensorimotor network, pointing out that these neural structures also support the subjective sense of agency and intentional action. This is compatible with the idea that MSC and SSP refer to self as agent, and also with Gallagher's (2000) original idea that minimal self involves a sense of ownership and a sense of agency (see Sec. 2. 4.1.).

4.2. Multiple levels of explanation

4.2.1. Phenomenology—concepts—operationalization—neural processes

In order to generate fruitful discussion and collaboration between philosophers conducting conceptual analysis and mind scientists performing empirical research, it is important to acknowledge the several steps in drawing connections between experience and neural processes. Legrand and Ruby (2009) propose a four-step methodology in order to develop the cognitive neuroscience of the self. Their proposition is presented here since it gives a detailed enough overview of the steps and the challenges they involve. The proposed methodology for delineating the self for cognitive neuroscience proceeds as follows (Legrand & Ruby 2009, 275-6):

> a) a description of the self that is conceptually and phenomenologically relevant and that allows for

> b) an operationalization of this description in functional terms, in turn allowing for

> c) the identification of the types of neurophysiological processes involved in step b and then

> d) the design of neuroimaging studies allowing the correlation between cerebral activations and step c.

119

In addition to these four steps, I want to point out that the relation between self-consciousness and the conclusions of empirical experiments involves at least two more steps. First of all, there is a step between (self-)conscious experience and any description of it, even before step a. As presented in Chapter 1 (Sec. 1.3.2.), a subjective experience is immediate or direct, whereas its description always is indirect, mediated by concepts. Thus, step a is especially important in order to construct a detailed enough description to grasp the subtleties of experience as well as possible. Conceptual analysis is useful in forming this kind of detailed description. On the other end, interpretation of empirical results comes even after step d. Of course, if all the steps have been followed, the data should rather simply indicate the correlations of interest, but it is good to keep in mind that interpretation of the results can be biased by defects in the preceding steps. Thus, also these additional steps need to be treated with care in order to assure an accurate scientific study of self.

Further, it is notable that while proposing these steps, Legrand and Ruby (2009, 276) are careful not to confuse personal and subpersonal levels of explanation (discussed in Sec. 1.3.2.). That is, what it is like to be oneself cannot be reduced to functional or neural processes. Instead, Legrand and Ruby (2009) advance a nonreductionist naturalistic approach where theoretical and experimental investigations of the self can "enrich and constrain each other" (Legrand & Ruby, 2009, 276; Legrand & Grammont, 2005; Zahavi, 2014). The idea behind such a methodology is that investigating the self requires conceptual clarity at the experiential level of description (step a, above). Simultaneously, it must be acknowledged that the self cannot be naturalized by a direct link between a purely experiential (step a) and a purely neuronal description (step d).

In proposing these methodological steps, Legrand and Ruby (2009, 279) intended to form a unitary framework, which enables us to theoretically define and empirically study multiple processes of self. In particular, Legrand and Ruby argue that investigations of the self should involve the study of the self-specific perspective and not be confined to the study of self-directed representations.

Accordingly, Legrand and Ruby (2009; see also Christoff et al., 2011) propose a paradigm shift where the neuroscientific study of the self does not determine the self in terms of self-related contents, but rather recognizes the significance of self-specific processes. Legrand and Ruby's meta-analysis reveal that standard neuroscientific conceptions of self involve self-representational processes of evaluating self-related contents. However, these contents are not self-specific and various studies of activity of the E-network have yielded apparently inconsistent data. Legrand and Ruby explain the inconsistencies by acknowledging that the E-network is recruited in general evaluation tasks that combine inferential processing with memory recall. Legrand and Ruby argue that the notion of self-specificity offers a way out from the impasse of the equation of self with self-related content, and also allows progress in investigations of the self in both theoretical and empirical fields. Self-specificity determines the perspective that relates any represented object to the representing subject. This notion characterizes the self in terms of functional processes; this functional characterization is relevant for neuroimaging investigations and phenomenologically sound at the same time.

On the other hand, I endorse that the results of empirical research can be useful for the concepts and theories of self. The clear empirical differences between MSC and RSC illustrate and support the distinction between two forms of self-consciousness. Indeed, the difference between minimal and reflective self-consciousness can be seen in different neurophysiological processes and patterns of neural activation and in distinct relevant paradigms of examination. At the same time, neuroimaging studies indicate significant co-occurrences of the neural correlates of MSC and RSC which support their interrelatedness. Thus, knowledge of the underlying mechanism of self-consciousness can assist in sorting subfeatures of self and in understanding relations between the features. Below, I analyze the methodological steps and empirical research in more detail, first in terms of both forms of self-consciousness in turn and then in their combination.

4.2.2. Evaluation network critical for self-reflection?

Above, reflective self-consciousness was connected to self-related neural processing, but the relation between the experience of self-evaluation and neural activations needs closer examination. The notion of SRP involves evident similarities to the notion of RSC, and empirical studies of SRP support the idea of RSC as a complex cognitive skill that comes in various forms. It is reasonable that self-related processing involves especially the frontal and temporal activation that are involved in executing evaluation and inferences and can be connected to complex cognitive skills present in psychological/cognitive, reflective and narrative aspects of self.

Generally, the CMS or E-network indeed seems to be necessary for the emergence of RSC. This can be simply seen, in that species without a comparable neural network do not possess RSC (at least according to our current knowledge); for instance, only a few species pass the mirror test (mentioned in Chapter 3, see, e.g., Suarez & Gallup, 1981). Empirical evidence for the significance of the E-network is clearly present in human development also; infants do not have complex RSC but only MSC (see, e.g., Block, 2009; Zelazo et al., 2007). The capacities of self-reflection develop gradually with the development of the brain, especially the E-network, including development particularly in the pre-frontal lobes and the generation of connections between different brain areas (Rochat, 2018; Zelazo et al., 2007). In addition, disorders of self (e.g., Ebisch & Aleman, 2016; Zhao et al., 2013) involve abnormalities in the activation of the E-network, which further evinces that the functioning of and tight connections within this network are crucial for the typical experience of self-reflection.

RSC involves complex cognitive skills enabling thinking of oneself in different modes, and also the neural profile of SPR is very diverse. Philosophers have elaborated several conceptual distinctions in order to capture different forms of RSC, and it could be useful to apply these concepts also in the study of the neural processing of self. That is, it is interesting to examine the similarities between philosophical-theoretical analysis of self and the interpretations of the results of brain imaging studies. Since self-reflection comes in different forms, and the neural basis is

constitutive condition for self-consciousness, the neural processes grounding different forms of self-reflection should also systematically differ from each other. The conceptual distinctions within RSC discussed in Chapter 3 included differences between considering oneself in 1P and 3P, deliberative and theoretical stance, voluntariness of self-reflection, and identification with versus withdrawal from the self as content. The original ideas concerning the uniqueness of first-personal self-reference and thought naturally did not involve any reference to neural studies but were purely philosophical in nature. However, at least some of these ideas indeed seem to be useful also in describing and understanding the results of neural studies of self.

The general theoretical distinction within RSC is drawn between approaching oneself in 1P and in 3P, and, for instance, Musholt (2013) has argued that the distinction should also be included in neural studies on self. Supporting this distinction, it has been found in empirical studies, for instance, that DMPFC and right parietal IPL/TPJ are activated when adopting a 3P observer perspective toward oneself (Dörfel et al., 2014; Frewen et al., 2020; Ruby & Decety, 2003, 2004). Instead, according to Frewen and colleagues (2020), a default egocentric perspective (default self) involves activation, for instance, in MMPFC and PCC-PRC, while a more active 1PP (executive self) involves DMPFC and frontoparietal control network. However, the concepts of 1P and 3P should be more precisely defined and operationalized in order to construct a neural model for them.

Especially interesting is that the distinction between the deliberative and theoretical stances towards oneself (Sec. 3.3.; Moran, 2001) seems to have similarities with the distinction between executive and observing selves within the quadripartite model (Frewen et al., 2020). The deliberative stance is essentially connected to action, and a parallel idea is present in the executive self that is defined in terms of active agency and initiation of action. By contrast, a theoretical stance is used in viewing oneself in a more passive and explanatory manner, more like in 3P, and this resembles the notion of observing self that emphasizes an allocentric perspective towards oneself. If the philosophical

distinction between the stances is applied, it also bears implications for the empirical level. The distinction entails different neural connectivity because the stances have different functions; deliberative is linked with action and thus has stronger and/or more complex connections with action. Whereas the function of the theoretical stance is to observe and to enable reappraisal without direct action but for the regulatory function, it is connected, for instance, to the processing of emotions. The distinction also entails that a breakdown in neural mechanisms grounding a stance should result in problems in the experience of taking that stance, too. Further, the theoretical distinction between deliberative and theoretical stances emphasizes that the stances are tightly connected and typically in "balance" with each other, and these strong connections and balance should also be present at a neural level. All these points seem to be tentatively supported by the quadripartite model; observing and executive selves are closely linked to each other but can be distinguished in the neuroanatomy of DMPFC and have at least partially different connections to other neural processes. At the same time, the distinction between stances could provide further hypotheses for future studies.

In addition, the quadripartite model seems to apply to the distinction between identification and detachment from oneself (Sec. 3.5.), revealing some of its neural basis. The observing self especially seems to be involved in non-identification mode, in which self is the content of consciousness, but just "looked at" without intention to identify or carry out actions. The similarity between the notions of non-identification mode and observing self is also supported by the reference to mindfulness meditation as an example experience of both notions.

Further, the quadripartite model is advantageous in being able to describe bottom-up versus top-down connections of self-reflection. These "directions" of self-consciousness seem also to be relevant in the distinction of voluntariness of self-reflection. Top-down self-consciousness involves voluntary focusing on self; self-related contents are actively taken to the object of attention in order to examine or regulate them or initiate action. Instead, Frewen and colleagues (2020) describe bottom-up processing as a typical state

of the DMN, characterized by involuntary and spontaneous rather neutral processing of different self-referential contents (in mind-wandering or body-wandering). In addition, emotionally loaded bottom-up processing of self occurs, for instance, in arousing situations that trigger an incidental form of emotion regulation (also mentioned by Christoff et al., 2011).

However, it is good to keep in mind the methodological steps connecting different levels of explanation; neural activation is not equivalent with personal-level experience. That is, the experience of self-reflection cannot simply be reduced to neural activation and concepts of personal-level direct experience. For instance, Moran's conceptual distinction between deliberative and theoretical stances might be used in finding proper concepts for descriptions of self-experience (step a), and these concepts can be further applied in developing functional (step b) and neuroanatomical models (step c) of thinking of self, and these models can be used in empirical experiments (step d). However, Moran's emphasis on the deliberative stance involves personal insight for the experience of rationality and taking responsibility for one's actions instead of staying a mere bystander, and these experiences cannot be captured in neuroanatomical terms.

In addition, it is good to keep in mind that the neural models of SRP are still developing and limited. For instance, the quadripartite model is still speculative and more data are needed to confirm it. In addition, the quadripartite model is restricted to MPFC, which is only one hub in the E-network, involving plenty of connections within the network and with other networks. Altogether, the empirical evidence from neural studies supports the complex structure of and variance within RSC, and the studies are constantly advancing in better reaching the subfeatures of self-reflection. However, further studies are still required to examine the subtleties of RSC and the connections between self-related processes.

4.2.3. Can empirical methods really reach the subjectivity of MSC?

Of the two forms of self-consciousness, MSC is the more challenging research subject for mind sciences. Yet, it would be a

significant shortcoming if MSC was ignored in empirical studies altogether since only by taking all aspects of self into account can the whole of self-consciousness come to be understood. MSC is a tricky object for brain-imaging studies since it is, by definition, a constituent of every conscious experience. This is an important point because the experiments of neural correlates of consciousness study the contrast between the same stimulus information when it is processed consciously versus non-consciously. Since MSC corresponds to the difference between conscious and nonconscious mental states, it has a critical role in the neural correlates of any conscious mental state. On the other hand, within conscious states, there is not a contrast condition in which MSC is present and in which it is not. There is not the same kind of all-or-nothing contrast condition to discern the correlates MSC as there is for self-related contents of consciousness which ban be present or absent in an experience. Thus, the investigation of the neural basis of MSC asks for developments in empirical paradigms. The notion of self-specificity is one useful tool that enables the inclusion of MSC in empirical research. This notion can be directed in studying, for instance, sensorimotor integration, homeostatic regulation, and emotion regulation. Since the self as the subject of experience cannot be missing in experience, the focus in empirical studies is on the changes within a self-specific system.

Generally, the data from empirical studies coheres with the idea that MSC is present in all experiences and includes embodied, experiential and affective aspects of self. Since MSC is a basic form of (self-)consciousness, it is very reasonable that self-specific processes involve basic neural operations needed in homeostasis and sensorimotor interaction with the environment. One significant structure within these operations is the insulae, which is involved in representing the body in space, interoception, generating feelings, and time consciousness. At the same time, the SSP is also present in higher cognitive functions that involve not only homeostasis or sensorimotor processing but also self as a cognitive–affective agent. The notion of MSC is useful since it gives connective conceptual tools to analyze the neural mechanisms that are present while contents and operations of consciousness vary.

Further, knowledge of neural structures and processes is consistent with the idea that MSC is a constitutive feature of experience that is also present in animals and infants who lack self-reflection. RSC develops together with the development of pre-frontal neural areas and connections. Even without more complex cognitive skills requiring pre-frontal activation, animals and infants are sensorimotor and affective agents, regulating their interaction with environment and undergoing feelings.

In developing the notion of self-specificity, Legrand and Ruby (2009) especially highlighted sensorimotor parietal processing. Legrand and Ruby (2009, 278–9) consider the possibility that sensorimotor processing could be claimed to be a necessary neural correlate of phenomenal consciousness. However, they remain cautious in the interpretation of the role somatosensory cortex activation at this point. Instead, they point out that further experiments are needed to confirm the data and to identify more carefully the neural mechanism of self-specific processes. In any case, the study of MSC has assisted in indicating the importance of sensorimotor processes in cognition that has no sensorimotor content.

Although sensorimotor integration is a paradigmatic example of self-specific process or MSC, MSC is also present when one is not moving and withdraws from perception. For instance, when one is lying down, eyes closed, in a quiet environment, MSC prevails in forms of interoception and possible mind- or body-wandering. Indeed, MSC is present until one enters a non-conscious state such as a dreamless sleep. In this most minimalist sense, the notion of MSC seems to come close to the notion of a state of consciousness (being in a conscious state versus being in a non-conscious state). The notion of a state of consciousness also fits the idea that MSC does not refer to the contents of consciousness. Altogether, MSC is present in all conscious states: as a feeling subject in basic homeostasis, bodily subject of perception, and cognitive–affective self involved in complex cognitive processes.

The operationalization of MSC for mind sciences is advantageous for making the notion of MSC clearer. The notion of a self-specifying system is an important step in converting the

concept of MSC into operationalized form that can be investigated by scientific methods. The notion is useful for empirical sciences since it entails a more careful examination of self that exceeds the notion of self-reflection. Moreover, the notion is important for practical applications—for instance, in understanding mental disorders that involve many layers of self. However, it should be noted the notion of SSP is still wide in its scope; although it manages to turn attention to self as the representing subject instead of the object of representation, it applies to numerous processes. Thus, future studies have plenty of work to do in order to specify all relevant SSPs.

Yet, a more fundamental problem in terms of SSP is that, especially in the case of MSC, it remains tricky to conflate subjectivity with neural activation. Although self can be operationalized as a subject or agent of perception and action, it still is relevant to ask if the subjectivity of MSC can ever be captured in neural terms—and the answer is still that it cannot. A first-person perspective provides immediate subjective experience while a third-person perspective and methods always are mediated by description, techniques and measuring equipment. The empirical data of the neural basis of MSC can increase the understanding of the processes underlying subjective experiences, also illustrating cases that are unfamiliar in one's experience, such as mental disorders, and thus the data are very useful in the investigation of MSC. Further, empirical research can assist in revealing the complexity of features and mechanisms that are involved in MSC, including the bodily and affective factors and their interplay.

However, the notion of self-specificity stands for a functional description of self, but MSC cannot be reduced to a mere functional description. The embodied reading of minimal self highlights the capacity to make a distinction between oneself and with the environment, instead of experiential reading that highlights consciousness in the first place and is endorsed in this dissertation (see Chapter 2). A self-specific system fulfills the condition of distinguishing itself and further specifies itself as an agent of perception and action through multisensory integration. Yet, the

question remains, how to get subjectivity from the sensorimotor functioning? Also a robot could adjust its movements in accordance with the forms in the environment, thus "specifying itself" as an agent, but we would not ascribe consciousness to the robot only for that mechanistic moving. That is, the notion of a self-specific system should be precise enough to distinguish (minimally) self-conscious systems, which have a subjective perspective, from systems that are not conscious, having only a geometrical perspective and executing only mechanistic operations.

4.2.4. The unity of MSC and RSC

To recap, the recognition of the two forms of self-consciousness and their subfeatures assists in clarifying and classifying the various neural processes involved in self-conscious states. Besides the distinction, the theories of self emphasize the unity of self-consciousness, and thus, knowledge of the relations between different aspects is also useful in completing the neural profile of self. It is rather clear that neural studies support the importance of the connections between the forms of self-consciousness in pointing out wide neural networks processing self. However, it is not easy to draw conclusions on the whole of interconnected neural processes that contribute to self. Below, the neural coupling of minimal and reflective self-consciousness is considered.

It is good to keep in mind that the distinction between minimal and reflective self-consciousness is not consistently followed in empirical conceptualizations of self. Particularly, there is still work to do in including the notion of MSC in empirical research, since the majority of studies have been confined to RSC only. Moreover, it is not always clear how the two forms are distinguished. For instance, Frewen and colleagues (2020) use the notion of SRP, which targets RSC by definition; it refers to self as the object of attention or content of representation. However, Frewen and colleagues (2020) also acknowledge the importance of the concept of self as 'I' (i.e., self as subject or perspective), considering that the uniqueness of SRP might lie primarily with this first-person perspective rather than the contents of consciousness. Still, Frewen and colleagues (2020) preserve their concept of SRP. In addition,

the notions of NV-SRP seems to direct bodily and affective processes that typically are implicitly involved in MSC. This can be seen, for instance, in the way Araujo et al. (2015) (whom Frewen et al., [2020] compliment on drawing conceptual distinctions) use the terms "autobiographical self" and "core-self," of which the first-mentioned is clearly more about complex self-reflection and the second-mentioned about embodied aspects of self. However, the case of NV-SRP is also reflective self-consciousness, since the bodily self is taken as the explicit content of consciousness. This remark seems to be present, for instance, in the distinction between intrinsic and extrinsic self-networks (presented at least by Ebisch & Aleman, 2016) that aggregates self-reflection and core self into intrinsic self, which seems to coincide with RSC. Extrinsic self, in turn, involves interactions with external world, sensorimotor processes, and the feelings of agency and ownership that are also involved in MSC when self is not held as the object of consciousness.

However, after recognizing the difference between the two, it is important to study the links between MSC and RSC. This viewpoint is crucial, since the two forms occur together, and their relations have an effect on self-experience. The significance of interaction between many aspects of self is recognized in many studies, but more at the level of mention, and it is not yet systematically targeted in experiments. One explicit emphasis on the relations between MSC and RSC is made by Ebisch and Aleman (2016, 784) who argue that the "basic structure of the self could depend on the balance between intrinsic and extrinsic self-processing." Also Frewen and colleagues (2020) find it interesting for future studies to define the neural networks that mediate SRP, resting state, and external attention in turn. Although Frewen and colleagues (2020) do not use the notion of MSC, it can be noted that MSC is present in all these conditions, and one option for future studies would be to focus on it. MSC might even be considered a mediator, or it could be included in the research question, for instance, in aiming to examine what kind of self-specific processes are involved in different conditions.

For accounting for the unity of self-consciousness, future studies have at least a twofold aim: to study the co-occurrence of two forms of self-consciousness and the interaction between them. The science of self can advance by following the methodological steps proposed above, and when the initial steps are managed with care, the steps can be taken in both directions. That is, when concepts of self-consciousness are detailed enough (step a), the concepts have been turned into functional descriptions (step b), and the relevant neural processes are identified (step c), it is possible to carry out experiments (step d) that also study the interconnections between the aspects of self. In turn, the results of these experiments can inform theories of self-consciousness.

The co-occurrence of the two forms is inevitable during RSC; every time when one is reflectively self-conscious, she is also minimally self-conscious. I want to point out that the data from brain imaging studies clearly support this co-occurrence and thus the idea that the two forms of self-consciousness are intertwined. The integrated relationship between the two forms of self-consciousness can be seen in that SRP also involves neural activation that can be linked with embodied and affective aspects of MSC. This is supported by empirical findings that in tasks of V-SRP, there is also activation in insular structures and sensorimotor cortices. Further, it is interesting that the same style of conclusions is raised in the study of both SSP and SRP. For instance, the significant role of DMPFC is concerned both in terms of a self-specifying agent performing cognitive–affective emotion regulation and in terms of top-down self-referential processes. This overlap in the results describing SSP and SRP supports the idea that MSC and RSC overlap in self-experience.

In addition to the plain co-occurrence of the two forms, it would be interesting to elaborate the nature and thickness of their interaction. Theoretically, the interaction case involves top-down and bottom-up studies, but neuroimaging research also offers unique opportunities to actually observe what kinds of complex interaction patterns occur in different cases of self-consciousness. The data from the neural activation patterns can illustrate and increase understanding of cooperation between different aspects of

self at the theoretical level too. The significance of the interaction between different aspects of self can be clearly seen, for instance, in pathological cases that show evident abnormalities in patterns of neural activation that are connected to self. For instance, Ebisch and Aleman (2016) and Zhao and colleagues (2013) argue that interaction between neural systems is of key importance in self-processing.

I agree with this idea that in psychiatric disorders, balanced and integrated connectivity between neural networks is replaced by unbalanced and disconnected patterns. For instance, schizophrenia and depression involve (partly) opposite patterns of altered neural functioning (e.g., Zhao et al. 2013). Schizophrenia is characterized by strikingly reduced connectivity between frontal and parietal parts of CMS and MNS and hyperactivity in parietal, sensorimotor parts of the network. These disconnections and the unbalance between neural networks seem to be consistent with symptoms of impaired self–other discrimination and an altered sense of agency. By contrast, depression is characterized by increased functional connectivity between frontal and parietal regions of the CMS, and by overall hyperactivity in the CMS. These patterns of neural activation go together with symptoms of excessive rumination and increased negative evaluation of self. After comparing the neural alterations in several psychiatric disorders, Zhao et al. (2013) propose a general hypothesis that self-identity disturbances "primarily result from any breakdown in integrated interactions between the frontal and posterior components of both the CMS and the MNS." Thus, the balanced interaction between different aspects of self is crucial for typical self-experience, and the study of neural disorders can assist in revealing the necessary basis of self-consciousness in general (more on pathologies of self in Chapters 6 and 7).

Overall, the philosophical theories of self-consciousness can clarify the empirical results and research by providing a conceptual framework that connects different findings of the neural processes underlying self. Within this kind of wide conceptual framework, it would be easier to see, for instance, the links between SSP and SRP, or how the notions of deliberative and theoretical stances

could be applied to many different studies. Further, the clear common conceptual framework could reduce the postulation of different kind of 'selves' and instead encourage study of the dynamics between them. For instance, instead of associating the terms of different kinds of 'selves' with the activation in MPFC, the activations of different parts of MPFC in the quadripartite model (Frewen et al. 2020) might be called emotional, neutral, observing and executive *modes* of reflective self-consciousness. This kind of terminology would cultivate the idea that self is composed of different features and interaction between them, but self is not reducible to any of these features.

4.3. An enactivist glimpse of the neural studies of self

This chapter has been concentrated on the neural profile of self-consciousness, but it is important to keep in mind that self-consciousness emerges as embedded in the whole body and in interaction with the environment. In other words, the embodied basis of selfhood includes wider processes connected with the neural ones; the brain is a crucial part of this basis, but the brain alone does not account for selfhood—the whole bodily system is needed. Further, the multidimensional reading of selfhood can also include processes that exceed the body, such as environment and culture.[64] As Gallagher (2013, 6) highlighted in the original pattern theory, the patterns of self are not reducible to patterns of brain activation, since more factors than just brain processes are involved.[65] Brain processes can reflect the way a self is constituted

---

[64] At the same time, there are interesting cases that involve self-consciousness without any bodily action or interaction with the environment but are enabled by neural activation only, such as dreaming. These cases offer a useful way to investigate self-consciousness (Sec. 2.5; see, e.g., Haanila, 2021; Revonsuo, 2006; Thompson, 2014; Windt, 2015), still the general point remains that in addition to having the brain, a self is an embodied being coupled with the environment.

[65] See also Vogeley & Gallagher 2011. Furthermore, it can be argued independently of the pattern theory of self or enactivism that all dimensions of self cannot be revealed by brain imaging methods (see, e.g. Räikkä, 2020) and thus, self cannot be reduced to purely neural mechanisms.

across different factors, but "who we are, or what self is, is more than the brain."

In other words, it can be noted that the considerations in this chapter have been "internalist" in the sense that they have been directed to the neural processes "inside the head." As mentioned in Chapter 2, this kind of approach has recently been challenged by the enactivist conception of the mind that argues that consciousness (alongside cognition) always involves bodily processes and interaction with the environment. When this idea is understood in the enactivist sense that these processes and interaction are constitutive for consciousness and self (instead of having a mere causal role), the results of neural studies are only partial. As Thompson (2014, 102) puts it, according to enactivism,

> ...the brain is always embodied, and its functioning as a support for consciousness can't be understood apart from its place in a relational system involving the rest of the body and the environment. The physical substrate of mind is this embodied, embedded, and relational network, not the brain as an isolated system.

Enactivism cannot be considered more detail in this dissertation; however, I want to point out that even if neural studies give only a partial answer to the constitution or emergence or mechanism of self-consciousness, neural studies are still significant and lend a valuable viewpoint to the structures of self. Neural studies can indicate at least some features that are present in self-conscious mental states and at least some connections prevailing between them. In addition, neural research can give important information about the breakdown of these connections in mental disorders.

Moreover, also in the enactivist framework, neural research is highly appropriate especially in the study of reflective self-consciousness that is "more" internal in nature than many other forms of cognition. If the enactivist thesis is formulated as a version that claims conscious mental states are *sometimes* constituted by environmental factors, some instances of RSC might lack these factors. For instance, perception and social cognition involve interaction with the environment by definition, whereas RSC by

definition is directed to one's mental state and thus inward in comparison to outward-directed perception or social interaction. RSC is associated with the activation of DMN that is contrasted with the task-positive network, which is activated during active engaging with the environment. Because DMN is complex, the conceptual clarity of the features and modes of self-consciousness is crucial in order to understand its function. Further, it is interesting that also the "default" of DMN can change, and these changes can be induced by intense practices of self-consciousness (in meditation; see Sec. 8.2.4.).

Concerning MSC, the enactivist notion of an autonomous system confronts the tension between subjective experience and a mere description of structure of a system (discussed in 4.2.3.). This tension can be seen, for instance, in terms of the notion of autonomous agent, which refers to a cognitive system that regulates its interaction with the environment (see Sec. 2.1.). The enactivist approach considers a self this kind of autonomous agent, but the same notion has been applied also to a living cell (Thompson et al., 2005) and within artificial agents (Barandiaran et al., 2009). However, it is clear that this kind of simple autonomous agent does not have experience (that we have), and the notion of an autonomous agent seems to give only a description of the functioning of a system. The enactivist solution is to highlight consciousness as an integral part of cognition from the beginning, claiming that the integration of the two is not the problem. However, the claim of their integration alone does not help; the enactivist theory should also be able to determine more precisely the relation between consciousness and cognition. For instance, enactivist theory should be able to answer the question of when an autonomous system is conscious and when it is not (an encouraging answer to this question invokes the complexity of sensorimotor functioning; see, e.g., Kiverstein, 2020). In any case, it remains important to bear in mind the difference between experience and neural activation, and that the two can be connected only through several explanatory steps. By paying attention to these steps, the neural basis of MSC can also be included in empirical paradigms. On the other hand, the enactivist viewpoint might assist in seeing

how the MSC can be tricky to reach by neural research. If MSC involves (or even is constituted of) interaction with environment, neural studies seem inevitably to give only a partial picture of it.

Altogether, the enactivist approach to mind advocates an idea of extended self that essentially involves processes that are not "inside the head" and thus, offers challenges to neuro-imaging studies. However, these challenges can be exploited in the development of paradigms in neurosciences. For instance, the notion of self-specific processing, which individuates self as an embodied agent distinguishing itself from the environment, can be seen as an elaboration of the enactivist notion of autonomous cognitive system (Sec. 2.2.). The neural studies are relevant for enactivist theories since they can indicate (at least some significant components of) the mechanisms of the self as cognitive agent.

4.4. Summary

The multidisciplinary framework, which was employed in this chapter, turned out to be advantageous in the examination of self. Learning from other disciplines assists in forming a complete picture of selfhood. It is, however, important to recognize that different disciplines contribute to different levels of explanation. On the one hand, careful conceptual analysis of dimensions of self-consciousness is useful for empirical investigations of self. The analysis by philosophers gives precise concepts and theoretical clarity that are needed in the interpretation of empirical results. Further, conceptual and theoretical work can directly contribute to empirical research by stating operationalizing claims, such as Legrand and Ruby's notion of self-specificity. Self-specific processes are essential for MSC, in which self is the subject of consciousness and these processes require neural activation that includes sensorimotor and affective features. The neural basis of self-specific processes notably involves at least parietal cortical components, as well as basic homeostatic regulation involving subcortical structures. In contrast, RSC operates with self as the content of consciousness and requires complex cognitive skills that are employed in evaluation, memory and attention. The neural

correlates of RSC comprise self-related processes that can be cortically located more frontally and temporally and essentially involve activation of the E-network.

On the other hand, knowledge of the underlying mechanisms of experience can be relevant for considerations of personal-level phenomena. The evident distinction between the neural mechanisms of MSC and RSC consolidates and strengthens the conceptual distinction between them. At the same time, the neural profile of self-consciousness indicates that the coupling of the two forms is evident. The neural processes underlying both forms occur together during RSC; self-referential neural processing also involves bodily and affective features, which supports the idea that MSC is a constitutive feature of RSC.

This way, theoretical and empirical viewpoints can advantage each other and co-evolve. However, wide neural activation is involved in tasks related to self, and understanding of these activation patterns requires a more detailed conceptual map of different sub-features of self. Within RSC, for instance, the major conceptual distinction between first- and third-personal stances toward oneself provides interesting empirical research questions. In addition, MSC has been targeted only in a handful of experiments and should be included in future empirical paradigms more often. Moreover, the connections between different aspects of self need to be examined more closely to complete the neural profile of self.

5. Connections between minimal self-consciousness and reflective self-consciousness

The major research question of this dissertation is how are minimal and reflective self-consciousness connected to each other, and this chapter tackles the question. Many times, the theories of self-consciousness have not touched upon relations between the features of self-consciousness more deeply than by mentioning that they should be elaborated and by recognizing the fundamental status of MSC (Bayne & Pacherie, 2007; Gallagher, 2013; Musholt, 2013; Zahavi, 2010a). Gallagher's (2013) pattern theory of self has also been criticized for not showing the relations between the aspects of self (e.g., Beni, 2016; Kyselo, 2014) but merely giving a list of aspects of self without telling how they are integrated. Thus, there is a need to clarify the connections within self-consciousness.

The first step in this chapter is to analyze the necessary relation that connects minimal and reflective self-consciousness; MSC is a constitutive feature of RSC. The analysis applies the popular conceptual distinction between phenomenal and reflective consciousness to self-consciousness, since the consideration of consciousness assists in seeing the differences between the concepts of MSC and RSC. In addition, the implications that MSC bears in consciousness studies are briefly discussed since they aid in indicating the significance of MSC. The second step in this chapter is to elaborate on the connections between minimal and reflective self-consciousness. In general, the operations within self-consciousness run bottom-up or top-down, that is, from minimal to reflective or from reflective to minimal self-consciousness. However, in practice, the dynamics of these operations are intricate, occurring in loops and comprising numerous factors. The key finding in this chapter is that minimal and reflective self-consciousness are intimately related to each other. This implies that studying the interconnections between these two forms and their unity is crucial in understanding selfhood.

5.1. Minimal self-consciousness makes reflective self-consciousness possible

The examination in Chapter 2 indicated that reflective self-consciousness depends on minimal self-consciousness. To recap, one cannot be reflectively self-conscious without minimal self-consciousness. MSC is a necessary condition of RSC, but RSC is not necessary for MSC: RSC implies MSC, but not the vice versa. However, MSC is not sufficient for RSC since there are (minimally self-)conscious states that are not instances of RSC but are directed, for instance, to perception of the outer world. The significance of MSC can also be seen in the study of consciousness.

5.1.1. Self-consciousness and consciousness

The characterization of consciousness can illustrate the distinction between minimal and reflective self-consciousness, since the two forms of self-consciousness have different relations to consciousness. Generally, consciousness studies draw a distinction between phenomenal and reflective (or access) consciousness (Block, 1995; Revonsuo, 2006).[66] *Phenomenal consciousness* refers to first-order or primary subjective experience; there is something it is like to undergo experiences (Nagel, 1974, see Sec. 1.3.2.). Instead, *reflective consciousness* is a higher-order form of consciousness that takes content of phenomenal consciousness as an object of attention for using it in reasoning and in directing the control of action.

As presented in Chapter 2 (following e.g., Kriegel, 2006; Levine, 2001), a phenomenally conscious mental state involves two characters: qualitative and subjective character. *Qualitative character* refers to the content of consciousness ("what" of experience), whereas *subjective character* refers to the what-it-is-likeness of experience ("how" of experience). The subjective character can be characterized as a kind of background feature of

---

[66] However, not all theories of consciousness draw the distinction between phenomenal and reflective or access consciousness (see, e.g., Block, 2009; Fazekas & Overgaard, 2018), but the distinction has been criticized, for instance, by the global workspace account of consciousness (see, e.g., Naccache, 2018).

experience: one does not pay attention to it although she experiences the contents of consciousness through it. Of these two characters, the subjective character is the more fundamental one since it makes a mental state conscious in the first place. It can be argued that there can be conscious mental states that lack qualitative character (such as deep meditation; see Chapter 8) but are conscious by virtue of subjective character only. However, the same does not go the other way around: without subjective character, there is no consciousness of qualitative character and the mental state remains nonconscious.

In terms of consciousness, MSC refers to the subjective character of phenomenal consciousness, whereas RSC refers to particular instances of generic reflective consciousness. In RSC, self is taken as the content of reflective consciousness. In contrast, MSC is the subjective character that is a necessary constituent of a conscious state. MSC is present in all experiences, including the reflectively self-conscious ones. The different positions within consciousness clearly show the difference between the two forms of self-consciousness. MSC is fundamental for both consciousness and self-consciousness, being always present in experiencing, whereas RSC is ascribed only to the cases of reflective consciousness that involve self. The relation between consciousness and self-consciousness is roughly illustrated in Figure 5.1.
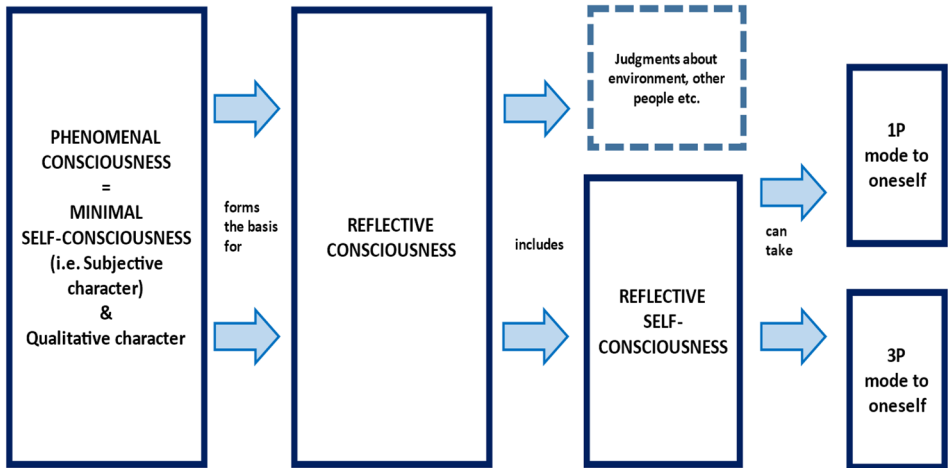
Figure 5.1. How the notions of self-consciousness and consciousness are joined together. Minimal self-consciousness (=MSC) as a subjective character is a structural feature of phenomenal consciousness itself, while reflective self-consciousness (=RSC) is a variety of the generic reflective consciousness. MSC forms the basis for the higher form of consciousness. RSC manifests itself in 1P- and 3P-modes.

5.2. Theories of consciousness

Since MSC is a structural feature of consciousness, an account of MSC bears implications (not only theories of self-consciousness but) also for consciousness theories. These implications are briefly considered below since they can further illustrate the conception of MSC and highlight the connectedness of consciousness and self-consciousness.

5.2.1. Anonymity theories of consciousness undermined

Proponents of the concept of minimal self-consciousness argue that the existence of minimal self-consciousness undermines so-called anonymity theories of phenomenal consciousness (Grünbaum & Zahavi, 2013; Zahavi, 2010c, 2014; see also Kriegel, 2006). An *anonymity theory of phenomenal consciousness* denies that MSC is part of what is consciously experienced (Grünbaum & Zahavi, 2013, 230). According to anonymity theories, experience is strictly impersonal and does not in itself involve any self-awareness. Zahavi

argues that an account that denies MSC in experience fails to recognize an essential aspect of experience and thus remains as an implausible account of consciousness (Grünbaum & Zahavi, 2013; Kriegel, 2006; Zahavi, 2010c, 2014)[67]. In Zahavi's (2010, 59) words: "Such a denial would be tantamount to a denial of the first-person perspective. It would entail the view that my own mind is either not given to me at all —I would be mind- or self-blind— or present to me in exactly the same way as the minds of others."

One example of an anonymity theory of phenomenal consciousness is strong intentionalism (e.g., Byrne, 2005; Dretske, 1999; Tye, 2003). According to the proponents of this view, "the phenomenal character of conscious states is identical to or supervenes on the state's intentional content" (Grünbaum & Zahavi, 2013, 231). In other words, they dismiss the subjective character and consider the qualitative character—that is, content— decisive for phenomenal consciousness. Thus, conscious experience does not in itself involve self-knowledge: it gives immediate knowledge about the world and only derivatively information about subject's conscious life. The intuitive appeal of this position is that it seems to apply to, for example, describing visual perception. In taking a look around, I engage a visual experience of the world; the involved phenomenal properties are experienced as properties of the objects I am seeing. When one attends to particular features of experience, one does not attend to oneself but rather to features of the objects that one is looking at; this experience is "transparent" (e.g., Dretske, 1995; Tye, 2003). However, this implies that living through the experience does not

---

[67] Zahavi (2014, Ch.3) distinguishes his account of pre-reflective self-consciousness from several kinds of theories, including phenomenal externalism (which claims that phenomenal consciousness does not entail self-consciousness), higher-order representationalism (which claims that self-consciousness is not intrinsic to phenomenal consciousness), and self-representationalism (which claims that phenomenal consciousness entails objectifying self-consciousness). Two first mentioned are considered in this chapter; the third is mentioned in Sec. 2.2.1. In this section, the main reference is Grünbaum & Zahavi (2013) because it gives a clear contrast between anonymity theories and a minimal self-consciousness theory (whereas Zahavi [2014] discusses "anonymity objections," among many other views) and uses argumentation that is especially relevant for this dissertation since it emphasizes the connections between minimal and reflective self-consciousness.

entail direct knowledge of the experience but only of the content. Instead, "knowledge of being engaged in the conscious act of seeing (rather than imagining) has to be inferred from the way in which the world is being represented" (Grünbaum & Zahavi, 2013, 231).

However, Grünbaum and Zahavi (2013) argue that an anonymity theory cannot account for a person's ability to refer to herself with a 1P concept and the unique features of self-conscious thoughts (discussed in Sec. 3.2.). Yet, accounting for these unique epistemic and motivational features—that a subject necessarily knows that she is referring to herself and this knowledge involves significance in her practical reasoning—is a major task for a philosophical account of (reflective) self-consciousness. In other words, anonymity theories are undermined since they cannot fully account for higher forms of self-consciousness, whereas the theories equipped with the notion of MSC manage it.

Grünbaum and Zahavi (2013) acknowledge that there are at least two ways in which anonymity theories can try to account for 1P self-awareness. First, by appealing to the perspectival character of perception: an impersonal perceiver is picked out by a definite description, "the perceiver of this perceptual presentation." And second, by appealing to general rules of reference, in which case the reference of the 'I' is determined by a simple rule: "'I' refers to the creature that produces it". Yet, neither of these appeals captures the characteristics of self-conscious thought.

According to the first route of anonymity theory, the perspectival character of perceptual content is used as self-specifying knowledge. The idea is that perceptual content has an implicit "back-reference" to the perceiver. This kind of back-reference explains how perception can directly control action, and according to anonymity theory, it can also ground one's ability to think 1P thoughts. After all, the perspectival features of perception provide the perceiver with knowledge of herself. However, Grünbaum and Zahavi (2013) argue that the appeal to perceptual content cannot determine the reference of 'I' by providing a necessary form of self-knowledge. In this appeal, the identity of self is specified by "a logical equivalent to a definite description: the

perceiver of this perceptual presentation" (Grünbaum & Zahavi, 2013, 233). However (as presented in Sec. 3.2.), any definite description cannot account for the unique epistemic and motivational features of first-person thoughts; we can always imagine a case in which the thinker refers to herself without knowing she is doing that. Likewise, there are cases in which the subject loses the perspectival form of self-knowledge but not her ability to first-person self-refer. This is present, for instance, in the pathological experience of thought-insertion such as "I am being visually presented with a tree but I am not really the one seeing it; this chap Chris is" (reported in Frith, 1992). The existence of this kind of experience indicates that the necessary features of self-knowledge involved in the use of 'I' are not determined by perceptual content.

The second route that an anonymity theory could take in order to account for our ability to have self-conscious thoughts is by appealing to general rules of reference. In this route, the reference of the 'I' is determined by a simple rule, not by a special form of knowledge. According to the simple rule, "a token of the 'I' refers to the creature that produces it" (Grünbaum & Zahavi, 2013, 233). In this case, the meaning of the 'I' is grasped by grasping this rule (Barwise & Perry, 1981; O'Brien, 2007, Ch. 4). However, Grünbaum and Zahavi (2013) point out that this attempt is problematic since it cannot account for why self-conscious thinking entails knowing and understanding oneself in the 1P way. The attempt produces a dilemma: it either over-intellectualizes reflective self-consciousness, or it describes only the causal role of self-presentations and is insufficient to explain self-*conscious* thinking.

On the first horn of the dilemma, one's understanding of 1P self-reference is explained by her conscious grasp of the rule. However, this proposal runs into problems because it presupposes the kind of understanding it is supposed to explain by containing indexical elements. Further, if a conscious grasp of the rule involves an ability to state the rule explicitly, the attempt is simply unconvincing; although people in general think and use self-conscious thoughts easily, only a few people can explicitly express the rule.

On the second horn, the rule is understood as "a law-like-generalization that describes the workings of the system" (Grünbaum & Zahavi, 2013, 234). In this case, the rule need not be explicitly expressed, but the application of the rule describes the inner structure of the system that puts the information involving direct relevance to the system into a specific information file (Perry, 1993). This gives a functional account of the motivational uniqueness of self-conscious thinking. However, the functionalist proposal seems insufficient to account for the special form of the understanding that is present in self-conscious thinking; the thinker necessarily understands that she is thinking about herself. The very understanding is not explained in the functionalist story. That is, this kind of anonymity theories fall into the problems of a functionalist view. The case is similar to Searle's Chinese Room argument; merely operating in a way that follows a rule does not entail understanding (Searle, 1980). For instance, a computer or robot might follow the token-reflexive rule, sorting information into two files, of which one is labeled by a greater relevance to the system, while the other file is simply for storage. However, merely operating this way does not entail self-conscious thinking or understanding that the system is referring to itself; we do not consider the computer or robot conscious for carrying out these kinds of operations.

Altogether, the anonymous attempts try to explain 1P thoughts by an appeal either to a definite description or to a functionalist reading. However, (as presented in Sec. 3.2.) 1P thoughts cannot be captured by any 3P description. Further, the functionalist outlook can be criticized in the same way that functionalist theories of consciousness are generally criticized: although a functionalist theory might manage to offer a functionalist explanation, it is not sufficient to explain some crucial features of consciousness. Merely operating according to rules does not entail that there is any understanding, since the same could be done also in the absence of consciousness. Thus, it seems that, because of excluding minimal self-consciousness, anonymity theories fail to account for the difference between 1P and 3P self-reference, which yet is a major feature of reflective self-consciousness.

145

5.2.2. Higher-order accounts of consciousness undermined

The second group of theories of consciousness that the concept of minimal self-consciousness arguably undermines is higher-order accounts of consciousness that ascribe consciousness to an extrinsic or relational property of those mental states that have it (Carruthers, 1996; Gennaro, 2020; Lycan, 1996; Rosenthal, 2005).[68] In terms of the distinction between subjective and qualitative characters of phenomenal consciousness, Kriegel (2006) notes that philosophers seem to be divided on which of the two characters is the more fundamental. Consistently with the discussion of anonymity theories above, Kriegel remarks that the philosophers who offer representational theories of consciousness (Dretske, 1995; Tye, 2003) seem to conceive of qualitative character as more basic and focus mainly on contents of consciousness. Instead, Kriegel (2006) notes that the philosophers who advocate higher-order monitoring theories of consciousness (see, e.g. Rosenthal, 2009) seem to focus on subjective character. In other words, both proponents of MSC theory and higher-order theorists of consciousness highlight the subjective character (recognized by Zahavi, e.g., in 2014; Gallagher & Zahavi 2007). However, the accounts differ from each other, since MSC theory sees subjective character as an intrinsic feature of consciousness whereas higher-order theories consider it extrinsic.

In higher-order theories of consciousness, the general idea is that a mental state is conscious when it is taken to the object of a proper kind of higher-order representation (HOR). These theories have the intuitive appeal that conscious mental states are the mental states that one is "aware of" in some sense. For instance, one's desire to drink coffee becomes conscious when she is "aware" of

---

[68] Generally, a theory that explain consciousness in terms of higher-order states is a higher-order representational (HOR) theory of consciousness (Gennaro, 2020, 1). The common distinction within HOR theories is drawn between theories that advocate higher-order thought (HOT) or higher-order perception (HOP). According to HOT theorists, HOR considers thoughts and involves a conceptual component (Rosenthal, 2005). HOP theorists, instead, consider HOR as a perceptual or experiential state that does not entail conceptual content (Lycan, 1996). However, different versions of HOR theories are beyond the scope of this dissertation. For the general debate of HOT theories, see, e.g., Block (2011); Rosenthal (2010).

the desire. This is described in the Transitivity Principle: "A conscious state is a state whose subject is, in some way, aware of being in it" (Gennaro, 2020, 3). By contrast, it would seem odd or even contradictory to say that one is having a conscious state if one is completely unaware of being in that state; a mental state of which one is totally unaware is clearly an unconscious state (argued e.g., by Gennaro, 2020; Rosenthal 2009, 2010). In higher-order theories, a mental state becomes conscious only when it is taken as an object of a relevant second-order mental state. Thus, the first-order mental state does not involve minimal self-consciousness in itself, but reflective self-consciousness is a necessary condition for being a subject of phenomenal feelings (as presented in Gallagher & Zahavi, 2007).

There is general criticism towards these higher-order accounts, classically that higher-order monitoring generates an infinite regress or leaves us with explanatory vacuity. However, an especially relevant objection in this context is that according to an MSC theory, a higher-order representation presupposes the existence of a prior minimal self-consciousness as its condition of possibility (Gallagher & Zahavi, 2007; Zahavi, 2014). MSC is "the categorical basis of our capacity for first-person thought, which explains why we can usually ... report on our experiences immediately and effortlessly" (Zahavi & Kriegel, 2016, 51; Sec. 2.2.1.). We are not "mindblind" before reflection, but reflection grasps experience that was already there prior to the grasping. Thus, MSC grounds higher-order representations and actually determines the sphere of what we may reflect about.

However, the case against higher-order theories is not this simple since HO theory can also be formulated as a claim that when a conscious mental state is a first-order world-directed state, the higher-order thought (HOT) about that state is unconscious (e.g., Brown et al., 2019; Gennaro, 2020; Rosenthal, 2010). This version of HO theory replies to the critique of infinite regress and seems to settle also the above presented shortcoming of requiring reflective self-consciousness for minimal self-consciousness. Instead, the HOT turns conscious only when it is taken as an object of introspection by a yet-higher-order (third-order) thought. In other

words, one is continuously "aware of" one's mental states but this awareness is generally not a conscious awareness.

From the viewpoint of minimal self-consciousness theory, this is still unconvincing. There seems to be something odd in saying that a mental state becomes conscious because one is aware of it although this awareness is not conscious. Generally, according to the phenomenological conception, MSC is not construed in terms of objects of experience ("experiences are not given as objects for me pre-reflectively," Zahavi, 2014, 16). Pre-reflective self-consciousness is non-observational; it should not be understood as monitoring since it does not involve a position or perspective of an observer or spectator. In addition, it has been argued that object-consciousness entails an epistemic divide between that which appears and to whom it appears (i.e., the object and subject of experience) and for that reason object-consciousness and the transitivity principle might alone be unsuited as a model for *self*-consciousness proper (Zahavi, 1999, 2014; Legrand, 2011; see also Sec. 2.1.2.). According to the MSC theory, consciousness is not objectifying, but subjectivity is an intrinsic feature of consciousness. The intuition that consciousness requires some kind of awareness can be covered in that minimal self is the subject of awareness. Since HO theories (e.g., the version proposed by Gennaro, 2020) recognize that something like self-as-subject is present in first-order experience, it seems unnecessary to postulate any HOR. However, it remains unclear how the notion of MSC and different versions of HO theories relate to each other. Would MSC be a property of the mental state that is an object of a higher order state? Or would MSC be formed in the relation between first- and higher-order mental states?

Furthermore, in more careful scrutiny, it seems to me that phenomenology and higher-order theories are not (always) even interested in the same question. HO theories highlight that a first-order mental state becomes conscious when it is monitored or meta-represented by a higher-order mental state. However, if these higher-order mechanisms are not conscious, it seems that a HO theory does not claim anything about the structure of the conscious mental state (e.g. it does not claim that consciousness would always

involve self-reflection or introspection as phenomenologists seem to interpret it). Instead, the higher-order theory concerns the emergence of a conscious state from non-conscious mechanisms. On the other hand, phenomenologists argue that consciousness should not be understood only in terms of its objects and thus, they consider the claims of experience as inner awareness unreasonable. However, a relevant critique of HO theories of this kind could simply point out that terms such as 'awareness' seem misleading if the higher-order processes in question are not conscious. In other words, phenomenology and HO theories differ in the "level" of mental phenomena that they target. While phenomenologists state that first-order experience is pre-reflective, HO theorists state that first-order mental states become experienced only when they are accompanied by higher-order mental states. That is, phenomenologists are interested in first-order 'conscious mental states', whereas the conception of the 'first-order' in HO-theories seems to refer to a 'mental state'. This conception is clear at least in the versions of HO theory that focus on the neural processes that are needed for consciousness. For instance, according to Brown and colleagues (2019, 756), higher-order representations "are usually associated with areas of PFC"; this claim does not involve any characterization of experience but describes enabling conditions for experience. If the claims of HO theories simply concern neural mechanisms, they do not seem to confront the phenomenological claims concerning the structure of consciousness and the nature of MSC. If MSC is considered in terms self-specific neural processes (as proposed in Sec. 4.1.2.), it would be interesting to compare these processes with the idea of activation of PFC that is presented in HO theories. However, this kind of comparison in terms of neural processes would differ from the original phenomenological critique of HO theories.

In summary, the theory of MSC undermines the versions of HO theories which claim that RSC is necessary for self-consciousness, whereas the versions of HO theories that focus on non-conscious mechanisms seem to be interested in different matter at the outset. Thus, it is unclear how the latter versions of HO theory and the MSC theory are related and whether the

phenomenological critique applies to these versions too. Actually, this kind of HO theory might criticize phenomenologists for not giving any kind of explanation of how a mental state obtains a subjective character. The relation between these views remains tense, but is not evident whether they are inconsistent. Because of this tension, the relation between the notion of MSC and different versions of higher-order theories of consciousness would be interesting to study in more detail and it could provide a useful elaboration of the theories; however, the study is beyond the scope of this dissertation.

Overall, it is worth noticing that even in the mentioned functionalist and higher-order theories of consciousness, some kind of minimal selfhood plays a remarkable role as a formal part of consciousness. In anonymity theories, the perceptual contents take place in a spatiotemporal perspective; a perspectival minimal self is needed as a structural element of consciousness even when it is not considered to entail subjectivity. In higher-order accounts, instead, the subjective character of consciousness is endorsed but it is explained as an extrinsic feature of conscious mental states. Thus, minimal self-consciousness indeed is important to recognize as a constitutive component of experience, although one would not commit to all claims of MSC theory.

5.2.3. The notion of minimal self-consciousness advances the notion of subjectivity

Although the notion of MSC is significant for phenomenal consciousness, many theories of consciousness do not involve a reference to self or self-consciousness.[69] If a theory underlines the notion of subjectivity, it is not anonymous in the sense that it would deny the idea of MSC altogether; however, it does not draw the link between subjectivity and MSC. Because of the shared emphasis of subjectivity, it might be argued that the theories involving subjectivity implicitly involve MSC. Below, I briefly

---

[69] Including the leading scientific theories of consciousness such as integrated information theory (Tononi et al., 2016), global neuronal workspace theory (Dehaene et al., 2011), and recurrent processing theory (Lamme, 2010).

consider MSC in terms of these kinds of theories of consciousness that are neither anonymity nor higher-order theories.

The idea in many theories of consciousness lacking the concept of MSC is that consciousness can be described simply as a presence of experience, without reference to self-consciousness. This can even be considered a conceptual advantage of a theory (e.g., Baker, 2012; Revonsuo, 2006): excluding self-consciousness from consciousness makes a theory conceptually simpler, avoids redundant or ambiguous notions, and can preserve the term 'self-consciousness' with reflective self-consciousness as it is used in everyday language. Experiences are simply co-conscious, linked to each other with some binding relations in phenomenal space, not involving a 'self' or 'me.'

This idea of consciousness as presence of experience can be accompanied with the concept of MSC. However, the concept of MSC aims to elaborate the conditions of experience. First, the concept of MSC points out that in order that there could be co-conscious experiences that are phenomenally present, there must be a self as the subject of experience (or as a cognitive system) *who* has the phenomenal space in which the experiences can emerge. Second, the concept of MSC entails that the presence of experience cannot be explained merely in terms of contents of mental states occurring in the subject, since mental states can be also unconscious. That is, experience not only takes place in the subject, but it is something like for the subject. This point highlights the subjective character of experience. These elaborations can be incorporated in a theory of consciousness: consciousness occurs only in a sophisticated cognitive system, and conscious mental states have subjective character.

Roughly, it seems that theories of consciousness that advocate subjectivity implicitly involve MSC. The notion of MSC aims to underline and elaborate the nature of subjective character. There still can be disputes about the details and emergence of this character, but I want to point out that it is also important to see the shared argument in theories of phenomenal consciousness and theories of MSC. Both argue for the significance of subjectivity and closer examination of it in order to understand conscious

phenomena. Thus, instead of being stuck in the debate on the notion of 'self', the theories of consciousness and MSC could accentuate the theoretical support they give to each other. Even if one would not commit to subtleties of the concept of MSC, one can recognize its basic ideas to be significant for the definition of phenomenal consciousness. That is to say, the proponents of MSC promote the same kind of arguments that are used in consciousness theories arguing for the subjectivity of consciousness.

5.3. Dynamics of self-consciousness

The rest of this chapter examines the dynamics of self-consciousness. By 'dynamics', I refer to the connections between aspects of self-consciousness; the way in which the aspects are related to each other is crucial in creating the shades in self-experience. Above, the focus was on the difference between MSC and RSC; MSC is necessary for RSC but not the other way around. This necessary relation has been highlighted in theories of self-consciousness, for instance, by Zahavi and Gallagher (2007). However, they do not elaborate the relations within self-consciousness from the opposite direction, which concerns the manner in which RSC can affect MSC. Nevertheless, the idea of the multidimensional self implies the idea of close connections within the features of self-consciousness even if it is not spelled out. Thus, since MSC and RSC are intertwined, it is important to study their interrelations more closely, not to say only that MSC is more fundamental.

I argue that the connection between MSC and RSC involves delicate reciprocal relations, and the acknowledging of these relations gives a fuller account of self-consciousness. These complex relations cannot be simplified to necessary and sufficient conditions but are more exhaustively described in terms of a pattern theory of self. However, Gallagher's (2013) original pattern theory of self has been criticized of not bringing out the relations between the aspects of self (e.g., Beni, 2016; Kyselo, 2014). According to this critique, the pattern theory gives a plain list of aspects of self but does not demonstrate how they are connected

together. However, the main idea of pattern theory is that the aspects of self are integrated and together form a pattern of self. Thus, the (pattern) theory of self needs to be elaborated in a way that also covers the connections linking the aspects together. Gallagher (Gallagher & Daly, 2018) has answered the critique by giving an analysis of the significance of the narrative aspect and how it is connected to other aspects of self. In the following, I will elaborate the picture of the connections within a self by examining the dynamics between minimal and reflective self-consciousness.

The emphasis on dynamics implies that a mere list of constituents is not enough; rather, an account of their connectedness is needed in order to understand the whole they compose. In a simple metaphor, if one is baking a cake, it does not suffice to know that she needs flour and sugar and eggs and butter and spices. Even knowing the right amount of the ingredients is not enough, but one must know how to put the ingredients together, that first one needs to whisk sugar and eggs into a froth, then add the dry ingredients without whisking, then bake the cake for the right time and so on. In the same way, knowing self-consciousness requires knowing the elements of MSC and RSC and how these elements are processed: which elements are linked together, is the self-consciousness flowing from minimal experiential features to reflective ones or the other way around, what kind of loops occur between them, how "thick" their relation is and so on. The spectrum of self-conscious mental states is considered below by some specifying remarks on the connections between MSC and RSC.

5.3.1. Simple picture of the relations: Bottom-up and top-down

The evident and straightforward interaction between minimal and reflective self-consciousness runs bottom-up and top-down, that is, from minimal to reflective self-consciousness and the other way around.

5.3.1.1. Bottom-up

In a bottom-up connection between the two forms of self-consciousness, a minimally self-conscious experience is taken into

closer scrutiny by means of reflective self-consciousness. Something in the first-order experience appears so important that one takes it as the object of her consciousness; that is, MSC has an influence on RSC. For instance, while one is focused on woodcutting (and not on herself), a sudden feeling of pain in her hand emerges, prompting her to pay attention to the hands and examine possible wounds. In this case, the self-consciousness leads directly to action: MSC has (reflex-like) an influence on RSC by presenting an urgent happening that immediately makes one stop her current activity and take herself into the locus of attention. Or in less urgent examples, minimally self-conscious feelings can trigger further self-reflection and analysis. For instance, when one is excitedly walking on cliffs and enjoying the landscape to the sea, she can start to think why exactly she is so excited. Is she excited for the pure joy of the view, or is she actually afraid of falling down from the cliff? Maybe the excitement is a combination of both? In that case, is she actually feeling joy because she is overcoming her fears? Should she walk even nearer the edge of the cliff in order to intensify the overcoming?

Another kind of bottom-up self-consciousness occurs during mind-wandering; when one is not focused on anything particularly but lets her mind wander without effort, she often meets thoughts that are related to herself (discussed in terms of default mode network, Sec. 4.1.1.). She might, for instance, run through a memory of spending a weekend in the city, and suddenly become captured by an emotional scene she had there, or by a thought about what kind of packing she should do next time. In this case too, MSC influences RSC by providing information about self to the central content of consciousness but without inducing immediate action. One can play with the self-related thoughts for a while, maybe evaluating her past actions or forming some undetermined future plans, or just daydreaming without active cognitive control. Since the thoughts are not urgent, she lets them flow and be followed by other thoughts.

The bottom-up direction in SC seems to be the default also in the sense that a first-order experience is needed to trigger a second order self-reflection. Everyday consciousness involves MSC as the

subjectivity of consciousness but is often directed to other matters instead of self. However, people are biased to react to self-related stimuli, and the first-order phenomenal consciousness feeds reflection by giving "material" concerning self. In other words, the bottom MSC is required to generate the cognitively higher RSC and top-down processing.

5.3.1.2. Top-down

In a top-down connection between the two forms of self-consciousness, reflective self-consciousness influences minimal self-consciousness. That is, I want to point out that RSC modifies MSC; thoughts about self can have effects on affective, embodied or behavioral aspects of self and thus alter the manner of experiencing. The capacity of sophisticated top-down self-consciousness has been emphasized as essential for being a human; one is not at the mercy of environment, desires and reflexes but is a reasoning agent who can orient her thoughts and action towards her values (see Chapter 3). Reflective self-consciousness can be used to direct action or to adjust the features and flow of MSC.

Classically, the thinking of "I want to be this-and-that kind of a person" leads one to actions and to modulate her MSC. For instance, the thought, "I want to be kind" can assist one in refraining from an impulsive offensive action that would be her first reaction to an insulting comment. When she holds back the impulse, her affects can also change; for instance, her feelings of being ready for defense or fight might be displaced by an interest to understand better what the comment was intended to mean after all. Or in another situation, the thought "I want to be kind" can motivate one toward a prosocial action in order to help a friend in need. Although she initially had been in a bad mood and a hurry, seeing the friend and engaging in the prosocial behavior initiated by RSC can also change the minimally self-conscious mood.

This kind of top-down self-consciousness can also be present as directed toward one's mind but without action in the environment. For instance, one can observe and regulate her emotions by top-down guiding in order to reach a calmer state of mind. In this case too, one is an emotional–cognitive–affective agent employing top-

down processing in order to change her current experience (discussed also in Sec. 4.1.2.).

The terms bottom-up and top-down explicitly point out a hierarchical structure within self-consciousness. In this hierarchy, MSC is present in the low foundational level, whereas RSC operates at the higher level of the hierarchy. From the bottom up, MSC provides constituents and contents to RSC, and from the top down, RSC controls and modifies MSC. These directions of effect are important to acknowledge, and they highlight the significance of RSC for meaningful self-contemplation and advanced cognitive control. However, often the relation between MSC and RSC is integrated and has a more dynamic nature: it is neither bottom-up nor top-down only, but these processes are coupled together.

5.3.2. More sophisticated picture: Mixture of the two forms

The bottom-up and top-down relations give an elementary picture of the connections between minimal and reflective self-consciousness, but in order to achieve a more comprehensive picture, more extensive dynamics of these connections need to be taken into account. I elaborate the close connections between MSC and RSC below, and remark how self-consciousness involves self as both a subject and an object of consciousness. The dynamics of self-consciousness can be seen already in the cases described above, which involved several aspects of self and the relations between the aspects affected in the self-experience. These effects concern both the content of consciousness (i.e., RSC) and the way in which the contents are present (i.e., MSC).

Many times, bottom up and top-down processes of self-consciousness occur in loops, and the looping highlights the importance of their connections. For instance, in the bodily case of yoga practice, one gives top-down (efferent) motor guidance in order to perform a pose, then checks her bottom-up (reafferent) feelings from the body and accordingly changes or deepens her pose by top-down (efferent) adjustments. Or during an emotive planning of her future, one can struggle with emotions; the first thought about a new option might fill her with a sudden bottom-up feeling of excitement that she then regulates top-down in order to

calm herself to concentrate on the facts. Then, bottom-up processing captures her with a distressing fear of failure that she top-down regulates. Then, the excitement follows again, then the distress, and she continues the planning and the balancing of emotions with information in loops.

Further, it is important to notice that the relations between MSC and RSC have effects both ways. On the one hand, MSC affects RSC by giving it material. On the other hand, RSC has effects on MSC by coordinating the manifestation of features of self-consciousness. However, RSC does not only "arrange" pieces of experiences but has a role in forming experiences. This is clear, for example, in taking the deliberative stance towards oneself by which one formulates how she feels or thinks (see Sec. 3.3.). In this case, the features of MSC depend on RSC, not just the other way around. For instance, when deliberative reflection leads one to see her anger as childish, the reflection helps to reduce the feelings of anger.

The neural profile of self-consciousness (Chapter 4) supports the view that self-consciousness is a complex whole in which the interaction of features of self is crucial. Selfhood is linked to wide neural activation in diverse neural networks, which manifests that various dimensions of self do not function separately but self-consciousness brings these dimensions together. For instance, the empirical studies pointed out self as a cognitive-emotive agent in which the processes of RSC and MSC occur together and affect each other. This is the point that I am advocating for: while one uses her RSC, she is both the subject and object of her reflection. This means that MSC and RSC are intertwined and have reciprocal connections.

The dynamic structure makes the system of self-consciousness very intricate, and the complex system is also vulnerable to defects. Even the "simple" connections do not always function. In the bottom-up direction, MSC can fail to trigger RSC even if it was needed. For instance, while being fully absorbed in action, one might miss paying attention and RSC to a bleeding cut in her hand of which she is minimally self-conscious. Or in the top-down direction, one might fail in her emotion-regulation; although she

uses RSC to modify the affectivity in her MSC, the emotions might be so overwhelming that she cannot change or ignore them. The intricacy of self-consciousness can also be illustrated by considering altered states of consciousness, in which the typical dynamics of self-consciousness change or fracture (more detail in Chapters 6-8).

Altogether, the two forms of self-consciousness are intertwined; variations in one or both or in their relations contribute to self-experience. Since the connections within self-consciousness play a crucial role, an account of self-consciousness should include them. This could be formulated as an argument:

1) Self-consciousness involves self as the subject of experience (=MSC) and self as the object of experience (=RSC).

2) Understanding a whole entails understanding its constituents.

3) Understanding a whole entails understanding the dynamics of its constituents.

4) Thus, understanding self-consciousness entails understanding MSC, RSC and their dynamics.

Premise 1 has been argued in Chapters 2 and 3. Premise 2 was presented in Chapter 1 as a beneficial way to approach the complexity of selfhood. The current chapter has argued for premise 3; a mere list of the aspects is unable to capture the shades of self-consciousness. Thus, self-consciousness is essentially a whole or unity; although it is conceptually (and empirically) important to acknowledge the difference between MSC and RSC, they are coupled together. A similar idea of the crucial interconnection has also recently been presented within the concepts of minimal self and narrative self (e.g., Belt, 2019; Bortolan, 2020).

For an illustration, the dynamics of self-consciousness are roughly outlined in Figure 5.2. The figure illustrates that a particular self-conscious mental state is formed by a network of connected features of self. Not only particular features but also the way in which the features are connected—their dynamics—is important for the self-experience. In addition, the networking of the features and relations that are not the center of attention produces effects on the whole. The circles represent features of self, and lines relations

between the features. Thickness represents the intensity of the feature/connection; the thicker circles represent the features that are especially strong, forming the focus of the current self-conscious state, and the thicker lines pick up the relations that are especially strong and integrate the experience. An 'aspect' of self (as presented in the pattern theory and listed in the figure) can involve several interlinked features and is linked to other aspects in the network. The connections are illustrated by lines instead of arrows for simplicity, but the lines can be considered as double arrows, involving both bottom-up and top-down relations between MSC and RSC. The lines draw only some possibilities of the connections, and the concepts in the circles roughly illustrate the manifold features of self-consciousness.
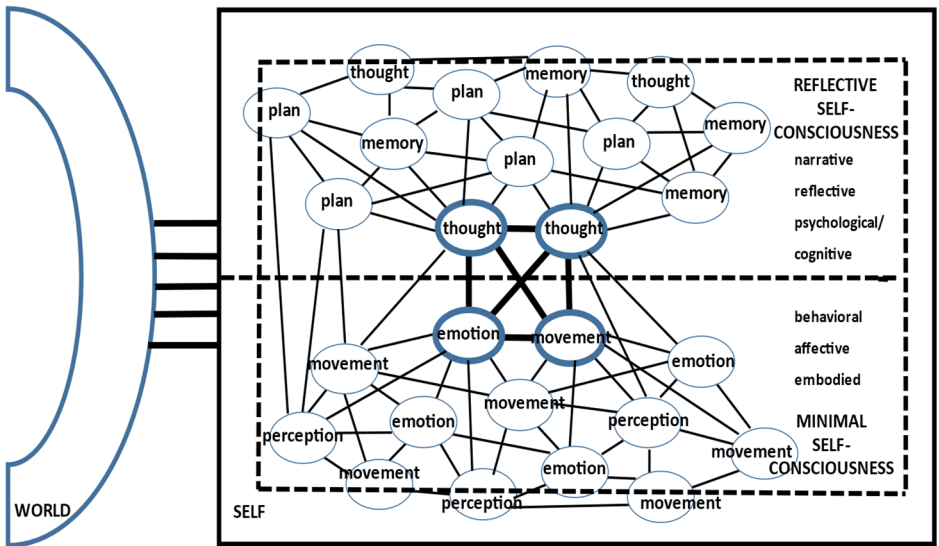
Figure 5.2. The dynamic structure of self-consciousness.

For example, Figure 5.2 could illustrate a situation in which one is eagerly practicing a new dance move and at the same time watching her mirror reflection in order to observe her progress.

159

The thick circles in minimal self-consciousness could be motor activity and joy of learning. These features involve embodied, affective and behavioral aspects of self that are included in MSC (as presented in Chapter 2). Especially the motor activity intuitively involves many components (movements of legs, arms and head; finding the right rhythm; balance; breathing; etc.) that are connected together, and their network is represented by lines in the figure. The thick circles in reflective self-consciousness could roughly represent the self in 3P mode to oneself and in 1P mode to oneself (which were discussed in Chapter 3) at the same time. In the 3P mode, self is observed as an object in the mirror and recognized as oneself, which can be associated with the psychological/cognitive aspect. In the 1P mode to oneself, one might have the joyous thought, "I know this move!" which can be associated with reflective aspects. These reflective features are linked to a network of other features involving narrative aspects of self—for instance, to a memory of one's teacher demonstrating the move the first time or a self-narrative of being a skillful dancer. Thus, a state of self-consciousness can be seen as an amalgamation of several aspects of self, and the examination of a state shows the tight and extensive links between the aspects.

Figure 5.2 might be further applied to functional and empirical research, illustrating the connections between functional or neural mechanisms that process self. Indeed, Ebisch and Aleman (2016) propose the same kind of network in their hypothetical and schematic graph of the functional connections between the default mode network (intrinsic self) and sensorimotor network (extrinsic self). In that case, circles represent rather local neural activation, and lines represent connections between them forming neural networks. In addition to the network connecting the features, Ebisch and Aleman (2016) highlight neural hubs (the thick circles in Figure 5.2) as having a critical role in exchanging information across the neural networks and balancing their function.

5.3.3. Balance within self-consciousness

The idea of balance between different features of self is also mentioned in philosophical accounts of self-consciousness, and I

160

want to develop it in this dissertation. The ideas of balance in self-reflection and psychological well-being have been presented, for instance, by Moran (2001, discussed in Chapter 3) who emphasizes the importance of both theoretical and deliberative stances. The first mentioned is linked to descriptive reflection of oneself, whereas the latter involves practical reasoning. Moran argues that both stances toward oneself, and alternation between them, are needed for mental well-being. Moran's idea is beneficial because it elucidates the practical importance of RSC in everyday life.

Developing Moran's idea further, I want to point out that the balance idea of self-consciousness should also involve MSC. Interestingly, it seems that overplay of either of the two forms of self-consciousness can lead to unbalance that is disadvantageous for psychological well-being. If RSC is exaggerated and one continually thinks of herself, she can run into negative thought circles and shut her eyes from other stimuli and even enter a depressive rumination. A focal function of RSC is to assist in understanding and balancing the conflicting thoughts. However, this example shows that too intense self-reflection can be harmful. On the other hand, absorption in mere MSC and minor exercise of RSC can result in impulsive, thoughtless and even irrational behavior. The excessive contribution of MSC can lead to dominance of affects, in which one merely reacts and does not utilize RSC to modify her feelings, and this also can be harmful in the long run. Thus, a domination of either of the two forms of self-consciousness can lead to distress that could be removed by finding a balance among the aspects of self.

In this sense, Figure 5.2 can be understood as an illustration of the extensive and balanced connections within different features of self. That is, balance refers to the connections or connected pattern of self, not (only) to the manifestation of individual features. The idea is not to strive to be in a state in which every aspect of self is represented exactly in the same amount; many times some aspect(s) can be predominant. Instead, I propose that the idea is to associate balance with a state in which the aspects are connected together in a way that enables flexible modification of the state. For instance, there could be a state of self-consciousness that is

primarily about some feature of self—for instance, a self-conscious thought—but it can be in balance in the sense that the thought has dynamic and flexible connections to other features of self. Instead, a state of self-consciousness is unbalanced if it is dominated by some feature and cannot be modified because it does not involve the required flexible connections between the dominating features and the others. The balance of self-consciousness is clearly disturbed in pathological experience, which will be considered in the next chapter.

5.4. Summary

Although the distinction between minimal and reflective self-consciousness is often made, their connections have been understudied. Typically, their relations have been considered merely by remarking that MSC grounds RSC: MSC is the most fundamental form of self-consciousness and a necessary condition of RSC. I pointed out that, as a constitutive feature of consciousness, the notion of MSC is significant for theories of consciousness. On the one hand, MSC undermines anonymity and higher-order theories of consciousness. On the other hand, the notion of MSC has resemblances to theories that highlight the subjectivity of consciousness.

In exercising RSC, both forms of self-consciousness occur together since one is both the subject and object of self-consciousness. I argued that in order to capture the subtleties of self-consciousness, it is important to study their coupling more carefully. Understanding self-consciousness involves knowing not only the elements from MSC and RSC but also how these elements are processed together. In a simple picture, the relations within self-consciousness run both bottom-up and top-down: MSC triggers RSC, and RSC is used in modifying MSC. In a more sophisticated picture, it is noteworthy that in practice the simple relations occur in loops and interact. These interactions are crucial constituents in a pattern of self-consciousness.

Further, I argued that targeting the dynamic of self-consciousness is important because the dynamic is connected to

mental well-being. Well-being is associated with a balance within self-consciousness that is founded on flexible and efficient connections between the two forms of self-consciousness.

PART II

Part I examined the two forms of self-consciousness and their amalgamation. It was argued that both minimal and reflective self-consciousness involve several features and that the connections between these forms should be investigated more closely. The significance of these connections indicates that a mere list of the features is not sufficient in capturing the subtleties of self-consciousness but interactions in the whole pattern need to be studied. Part II elaborates the dynamic in self-consciousness by analyzing it through concrete example cases. The conceptual framework, that was developed in Part I, is applied in accounting for altered states of consciousness in Cotard syndrome, depersonalization, and meditation. These states involve profound alterations in the structure of self-consciousness which can illuminate the function of typical self-consciousness too. These cases indicate that the connections within self-consciousness are not only theoretically interesting but their significance can also be seen in practice: only by acknowledging these connections can instances of self-consciousness be explained. While Part described the mutual influence between MSC and RSC, Part II demonstrates it: self is colored by the interconnections within self-consciousness and a change in these connections results in a change in self-experience.

6. The case study of Cotard syndrome exposing the structures of self-consciousness

In this chapter, I analyze concepts of self-consciousness in relation to a pathological case called Cotard syndrome (CS in brief). The distinction between two forms of self-consciousness is used as a general or guiding first step in analyzing CS, but the distinction will also be deepened. Both MSC and RSC involve several processes, and some of the processes can be severely distorted while others are preserved and may become emphasized. Further, the dynamics between MSC and RSC will be taken into account; instead of being isolated, these two forms are linked to each other, and this linking becomes obvious when it is disturbed in CS.

First, I will argue that pathological cases can be used as a test cases for theories of self-consciousness. On the one hand, pathologically altered consciousness offers an opportunity to examine structures of self by giving contrast conditions to normal states of consciousness. On the other hand, it should be acknowledged that pathological experiences can differ radically from normal ones and be especially difficult to grasp explicitly. A phenomenologically informed analysis is useful in providing a comprehensive picture of pathologies.

Second, I will defend the position that CS is an extreme abnormality of self-consciousness and elaborate what it can reveal about the structures of self. I will examine different explanations for CS and, by combining them, formulate a conception of self-consciousness that can encompass anomalous self-experience too. I will argue that a fine-grained conception of self-consciousness is needed in order to account for CS, and that CS involves multiple distorted aspects and relations within self. I will present that in CS, MSC is not lost, although it is distorted and thus far from normal. The analysis will indicate that CS originates in diminished MSC, especially in diminished affective aspects of self but also in experiential aspects. Further, I argue that the connections between MSC and RSC are distorted and that RSC also becomes dysfunctional in CS. RSC operates in the mode of hyper-

reflectivity: it is exaggerated and non-voluntary and manifests mainly a 3P approach to oneself. A loop between hyper-reflectivity and diminished MSC is overplayed in self-consciousness, while a typical network of tight connections between MSC and RSC breaks down. As a result of these dysfunctions, patients enter nihilistic delusions.

6.1. Altered states of consciousness as a methodological tool to study self-consciousness

The umbrella concept of self-consciousness comprises various phenomena, and the exact definitions for different forms or aspects of self-consciousness are currently under debate (as has become clear in the earlier chapters). Neurophilosophy proposes that a beneficial way to clarify the notion of self-consciousness is to look at real-world instances of it. As Metzinger (2004, 313) puts it:

> If we are seriously interested in a conceptually coherent and empirically plausible theory of the self-conscious mind, then it is important to test our conceptual tools at least against some examples of the enormous phenomenological richness of our target phenomenon.... Empirical constraints are relevant and indispensable for anybody who is seriously interested in the philosophy of self-consciousness.... What we need is a conceptually coherent theory of the self-conscious mind, which is phenomenologically and empirically plausible at the same time.

The use of empirical "test" cases or "constraint" enables an elaborated evaluation of theories. A theory of self-consciousness should be able to cover self-consciousness in all of its varieties: if a theory cannot do so, it should still be developed to be exhaustive and count the whole phenomenon. Thus, a theory of self-consciousness that fails to embrace all cases in which self-consciousness is present is undermined: the concept of self-consciousness cannot be accurate enough if it cannot be applied to cases that deviate from the exemplar. On the other hand, a concept

strengthens when it can also manage rare cases; its explanatory power widens and as do the reasons to endorse it. In addition, empirical case studies can indicate the richness of self-consciousness by bringing out its features that might not get attention in everyday life and a simple theory might not distinguish.

One promising methodological invention in the research of self and consciousness is studying them through different altered states of consciousness. The idea is to provide an analysis that uses an altered state of consciousness (ASC in brief) as a contrast condition that can elicit the features of normal consciousness. In other words, ASCs can be seen as a methodological tool that can assist in sorting out the aspects and functions of self-consciousness. An ASC can be defined as "a temporary change in the overall pattern of subjective experience" (Farthing, 1992, 205) or "a state in which the neurocognitive background mechanisms of consciousness have an increased tendency to produce misrepresentations such as hallucinations, delusions, and memory distortions" (Revonsuo et al., 2009, 187). The wide theme of ASCs cannot be discussed in detail here (see, e.g., Ludwig, 1966; Revonsuo et al., 2009; Tart, 1990; Vaitl et al., 2005; Wittmann, 2018), but we can simply note that ASCs have been shown to be useful in investigating the foundations of self-consciousness. Altered self-consciousness has been studied in a wide range of ASCs, which Ciaunica and colleagues (2021, 2) classify under four headings:

i) meditative practices (Dahl et al., 2015; Garfield, 2015)

ii) experiences under psychedelic drugs (Carhart-Harris et al., 2012; Letheby & Gerrans, 2017; Millière et al., 2018)

iii) induced illusions (Blanke & Metzinger, 2009; Lenggenhager et al., 2007)

iv) pathological conditions including schizophrenia (Fuchs, 2015; Parnas & Handest, 2003; Sass & Parnas, 2003), psychosis (Nelson et al., 2009), and depersonalization disorder (Billon, 2017; Ciaunica et al., 2021; Gerrans, 2019; Seth et al., 2012; Sierra & David, 2011)

The strategy of examining ASCs seems especially relevant for detecting minimal self-consciousness and the dynamics in self-consciousness. MSC is a necessary subjective component of experience that is intertwined with the contents of consciousness. This means that we cannot capture MSC in itself in typical consciousness but only in conjunction with a content of experience. Thus, MSC flees our introspection and is also extremely difficult to operationalize. In everyday experience, MSC is a tacit feature that we take for granted. In addition, the dynamics that ground self-consciousness as a whole are something that we take for granted in typical experience; we do not pay attention to the complex character of self-consciousness but flexibly modulate the modes of self-reflection. However, in ASCs the pattern of features of self changes, and the typical operations of self-consciousness might fail.

That is, the contrasts between altered and normal experience bring out the tacit features of self that we ordinarily do not notice: only when those features change or go missing does one understand what they were in the first place. Thus, profiles of altered states of consciousness may disclose the intricacy of self-consciousness better than normal experience. Altered states of consciousness assist in finding a phenomenological definition of selfhood (which was presented as the first step that is needed to take in empirical studies of self in Sec. 4.2.1.) by providing examples of variations in the pattern of self, and these variations can elicit the components composing self-experience.

6.1.1. Pathological cases as a challenge and an opportunity for theories of self-consciousness

Altered states of consciousness include pathological states in which normal experience is somehow distorted and as such becomes more striking. As pathological self-experience is an important contrast to normal, it affords an opportunity to see into the structures of self-consciousness (Daly & Gallagher, 2019; Metzinger, 2004; Zahavi, 2005, 132-146). In Metzinger's words (2004, 312), case-studies from cognitive neuropsychiatry allow for "reverse engineering": an empirically plausible investigation of identity disorders would yield a better understanding of identity and

selfhood under standard conditions. In addition, the case studies of identity disorders can give an effective cure for "Philosopher's Syndrome," that is, "mistaking a failure of imagination for an insight into necessity" (Dennett, 1991, 401).

In terms of the pattern theory of self, Daly and Gallagher (2019, 5) argue that "all psychiatric disorders are self-disorders, understood in a wider sense to mean varied disorders in self-patterns." That is, a pattern theory provides a unified theoretical framework to examine self-disorders. Pathological cases manifest themselves as disrupted aspects of self and changes in the dynamical pattern itself. On the other hand, the changes in the pattern can elucidate its original texture. The idea of psychiatric disorders as disrupted patterns of self can also be considered in terms of neural processing. The connectivity between neural systems is crucial for normal self-consciousness and this integrated connectivity, or a complex neural "pattern," breaks down (or at least alters) in psychiatric disorders (e.g., Ebisch & Aleman, 2016; Frewen et al., 2020; Zhao et al., 2013).

Particular pathological cases can function as challenges to theories about self, providing tricky cases that a theory needs to be able to explain. A popular example has been thought insertion in schizophrenia, which has been used in discussions in the philosophy of mind (e.g., Gallagher, 2000; Gunn, 2016; Lane, 2012; López-Silva, 2019). A patient suffering from thought insertion claims that somebody else's thoughts occur in her mind, and philosophers have been eager to explain what exactly happens in this peculiar condition and to use it in arguing for or against some theories of self (for discussion, see, e.g., Henriksen et al., 2019). Some have highlighted that the patient is so alienated from her cognition that her experience lacks some kind of sense of mineness. Others have highlighted that even this kind of extreme experience does preserve basic forms of selfhood, including the subjectivity of MSC.

However, pathological cases are also challenging for theories of self in the sense that we need to have a picture of the general nature of pathological experience before it can be used in an inference concerning normal experience. In other words, the

170

challenge is to which extent we can treat pathological experience in the same way as normal experience. On the one hand, it seems that pathological experience is so different from normal that it might not be used at all in order to learn about the normal structure of self. On the other hand, if we do use pathological experience in theories of self, it still needs to be noted that single descriptions of (for instance) delusional experience might not be employed in an analysis in fully the same way as ordinary experiences, but also the wider pathological context should be acknowledged.

6.1.1.1. From discontinuous interpretation to phenomenologically informed psychiatry

The first side of the challenge asks whether we can draw conclusions about aspects of normal self-experience from pathologies that seem to be very different in kind. Aren't pathological cases so abnormal that they should be excluded from the study of notions of normal self-consciousness? One approach to this question can be made in terms of continuous and discontinuous interpretations that invoke a patient's rationality. According to Billon (2016, referring to Bortolotti, 2009, 2018), a *continuous interpretation* makes sense of patients' claims and, by contrast, a *discontinuous interpretation* argues that we cannot make sense of the patients' reports. In other words, the difference between continuous and discontinuous interpretations concerns the patients' rationality: continuous interpretations assert that patients are minimally rational, whereas discontinuous interpretations dispute patients' rationality. Thus, discontinuous interpretations can, for instance, ascribe delusions to a form of irrational thinking, while continuous interpretations endeavor to draw up a more extensive image of how abnormal experience can provide reasons to form delusions.

Continuous interpretations seem more promising for tracing the subtleties of self and providing psychiatric understanding. Although pathological experience can differ significantly from normal, it also has features that are present in normal experience. Instead of simply denouncing pathologies as irrational and pathological experience as totally unreachable, it seems more

fruitful to study more closely the nature of pathological experiences. The shift from normal to abnormal is not abrupt but gradual—at least in the case of the origins of progressive mental illness, in contrast to sudden brain damage from an accident. This gradual shift gives reasons to study the fine-grained alterations of normal self-consciousness in the first place. This kind of approach has been argued for, at least in phenomenologically informed psychiatry (e.g., Fuchs, 2015; Henriksen et al., 2019; Parnas et al., 2013; Parnas & Zahavi, 2002; Sass et al., 2018; Sass & Parnas, 2007), which advocates for the continuous interpretation. As Sass and colleagues (2013, 438) describe it, the phenomenological point of view "ascribes a significant (though certainly not exclusive) pathogenetic role to the patient's general experiential/affective orientation, rather than only to perceptual, cognitive, or intellectual deficit." The emphasis on experiential and attitudinal factors cultivates a multi-leveled, dynamic, and holistic conception of pathogenesis in addition to neurobiological research. Daly and Gallagher (2019) argue that this kind of extensive framework is also compatible with a pattern theory of self. The phenomenologically informed viewpoint is used here since it can give an elaborated pattern theoretical picture of self in pathological experience. That is, fine-grained conceptual analysis goes hand in hand with fine-grained phenomenological analysis.

6.1.1.2. Nature of delusions

The second side of the challenge of using pathological examples in building a theory of self-consciousness requires a more precise continuous account of pathological experience. This involves large themes in the philosophy of psychiatry and cannot be considered here in detail.[70] In order to give a general picture of the models of pathological experience, below is a brief discussion of delusions, since a particular delusion is characteristic for Cotard syndrome, which will be analyzed in the rest of this chapter.

---

[70] For philosophy of psychiatry; see, e.g., Fulford et al., 2013; Murphy, 2020.

According to common definition, delusions are "false beliefs that are held with certainty and cannot be corrected" (Feyaerts et al., 2021, 1; see, e.g., Berrios, 1991; Bortolotti, 2009). This definition follows Karl Jaspers' (1997) "three criteria: a belief is delusional when a) its content is false or impossible, b) it is held with unshakable conviction, and c) it is incorrigible by rational argument or counterevidence" (Feyaerts et al., 2021, 2).[71] Feyaerts and colleagues (2021) propose that this definition has led most psychological and neurocognitive models of delusion formation to a shared general doxastic approach. Within this approach, delusions are formed as a response to anomalous experience by using the same normative-psychological framework that we use for non-delusional beliefs, and theories differ in treating these responses as rational or irrational. One-factor theories claim that a delusion is a reasonable hypothesis, which explains an anomalous experience. Instead, two-factor theories claim that the adoption or maintenance of a delusion also involves reasoning deficits or biases. Recent predictive coding theories propose that a single deficit in a hierarchical Bayesian information-processing system explains both anomalous experience and delusional belief.[72] However, all these theories have been criticized for failing to give an exhaustive picture of delusions. For instance, Feyaerts and colleagues (2021, 7) argue that these current theories could be updated and enhanced by "adding more phenomenological specificity to the two factors."

From a clinical-phenomenological perspective, it would be beneficial to acknowledge the wider psychopathological context within which delusions originate (Feyaerts et al., 2021). This also questions the conception of simply taking delusions as false beliefs. As Zahavi (2005, 145) notes: "Delusional statements ... are attempts to express highly unusual and frequently dreadful experiential

---

[71] However, as Feyaerts et al. (2021, 2) point out, Jaspers himself considered these criteria only external indicators that do not define the essence of delusion: "to say simply that a delusion is a mistaken idea which is firmly held and which cannot be corrected gives only a superficial and incorrect answer to the problem." Nevertheless, this definition is largely restated in major psychiatric manuals.

[72] For a predictive coding approach to Cotard syndrome, see Gerrans's (2015) model in Sec. 6.5.2.

situations that inevitably stretch ordinary language to its limit." Feyaerts and colleagues (2021) propose that instead of being mere mistaken beliefs, delusions are elaborated expressions of a global reorganization and shift in basic structures of experience. These basic structures include a minimal sense of self, bodily awareness and lived space and time. That is to say, "rather than being a matter of specific misperceptions within everyday reality, delusional atmosphere involves a qualitative shift in the overall way reality is experienced" (Feyaerts et al. 2021, 7).

Feyaerts and colleagues (2021) argue that the simple definition of false belief has led to the predominant focus of research on the (irrational) content of delusions, while alterations in the structures of experience (i.e., in self-consciousness and experience of reality) have generally been neglected. However, these global alterations seem to be in a key role in the origins of delusion, and thus acknowledging these experiential transformations provides advantages over current content-focused conceptions. In terms of self-consciousness, the phenomenological approach recognizes the role that MSC has in self-experience, and thus it seems most promising in capturing the complexity and dynamics of self-consciousness or a self-pattern. The following analysis of CS here supports a phenomenologically informed approach but is compatible with one- and two-factor models too.[73] However, the analysis here does not further elaborate any general model of pathology and employs the common way to use notions such as 'delusional beliefs' in referring to the cognitively high level of self-reflection and notions such as 'affectivity' and 'eroded first-person perspective' in referring to the cognitively lower level of MSC.

Still, a major challenge in using pathological examples in theorizing self is to respect patients' experience. Although pathological cases can function as tools in theories of self, it is important not to treat single curious descriptions of an anomalous experience as purely instrumentally and without the context. In the end, the interest of psychiatric research is to understand the

---

[73] In the same way as, e.g., Billon's (2016) interpretation of CS can be combined to either one- or two-factor theories, but is committed to neither, see 6.5.3.

anomalous and distressing experience in order to help the patients. However, by seeing the wider picture of pathological experience and being careful with the presented challenges, the pathological experience offers an exemplar way to apply the neurophilosophical idea of multidisciplinary cooperation that can benefit all disciplines. That is, pathological cases are interesting for the philosophy of mind, and on the other hand, conceptual analysis and theories can assist in empirical and psychiatric endeavors.

Fine-grained conceptual analysis of self can have very practical implications. Precise knowledge of different aspects of self can directly contribute to empirical and clinical developments. As Zhao and colleagues (2013, 1) point out: "A better future understanding of how changes in the neural systems sub-serving self-processing contribute to different aspects of symptom abnormality in psychiatric disorders will require that more studies carry out detailed individual assessments of altered self-processing in conjunction with measurements of neural functioning." Conceptual analysis aims exactly at a detailed assessment of different aspects of self, and thus, it can contribute to empirical studies (for instance, by giving conceptual tools to the interpretation of neuroimaging results, Sec. 4.2). In addition, the conceptual recognition of alterations in structures of self-experience can help in determining and diagnosing mental illnesses and to guiding patients to the right kind of treatment (see, Sec. 7.5.).

6.2. Cotard syndrome as a challenge to self-theories

In this dissertation, it has been proposed that MSC is necessary for being a self and that RSC has special epistemic and motivational features. Especially, the theories of MSC emphasize that selfhood is a necessary feature of our conscious lives. Whenever we are conscious, we are self-conscious at the same time: MSC is a feature a conscious creature cannot lack. Also generally, some form of self-consciousness has been highlighted to be the central feature of mind throughout the history of philosophy (Sec. 1.1.). For instance, Descartes' insight "Cogito ergo sum/I think, therefore I am," is one of the most famous ideas in western philosophy. The idea in

Descartes' reasoning is that the existence of self is the most certain thing one knows: while we can doubt all other things, we cannot doubt the existence of our consciousness.

Yet, there is a well-documented, although a rare, mental illness called *Cotard syndrome* (CS in brief) that seems to be a counter-example to the above philosophical claims. Patients suffering from CS have a delusion that they are dead: they claim that they do not have bodily organs, thoughts, or a self. In short, they are dead and do not exist at all anymore! Compared to our normal experience, this kind of condition is really difficult to understand. At first sight, it seems either just irrational or to undermine the conceptions of MSC as a necessary feature of consciousness. If one can have experiences of being dead or non-existing or having ceased to be a self, these experiences seem to be straightforward counterevidence to the claim that selfhood is necessarily present in experience.

However, it would be too hasty to conclude that CS simply undermines the notion of MSC. Rather, this peculiar syndrome raises the question of *what kind of self* or *which dimension of self* do the patients lack? The answer to this question also sheds light on subtleties of self. In other words, a closer scrutiny of CS is interesting not only for making sense of the patients' anomalous experience but also for clarifying the notions of self-consciousness. In order to use CS as an opportunity to investigate the subtleties of self, the characteristics of the syndrome need to be carefully observed.

The following analysis is not meant to be a complete or final description of CS; there is considerable variation between individual cases of CS, and not all clinical aspects can be considered here. However, the aim is to respect the experience of the patients suffering from CS and not to oversimplify their experience by picking only sporadic peculiar descriptions of self for scrutiny in order to prove some theoretical claims. In order to give a comprehensive picture of CS, diverse symptoms and many aspects of self are taken into consideration, especially emphasizing the experiential dimension of self. The analysis here aims to clarify the way different features of self and the pattern unifying them are

distorted in CS. The considered pathological examples are significant for theories of self even though the examples do not give an exhaustive account of the syndrome, and the analysis here can be elaborated in further analyses. The general idea is that a multidimensional approach to self can aid in understanding the experience patients are going through and, thus, also give tools to help them.

6.2.1. Symptoms and clinical picture

Cotard syndrome is a complex psychiatric condition of nihilistic delusions.[74] Patients suffering from CS can deny being alive, having guts, thinking, or even existing (Billon, 2016, 356). Most strikingly, CS involves the belief that the sufferer has died, and this individual belief is called *Cotard delusion*. However, this particular delusion is not a necessary feature of CS, but the syndrome can also manifest as a combination of other nihilistic delusions about the self, the body, and the outer world (see, e.g., Berrios & Luque, 1995; Billon, 2016; Enoch & Ball, 2001; Radovic, 2017; Young & Leafhead, 1996). Below are few case reports illustrating the symptoms of CS:

> A 48-year-old man with no medical history, apart from a previous short depressive illness, was seen by a psychiatrist after a self-electrocution attempt. Eight months later, he first told his general practitioner that his brain had died. He further explained that "I am coming to prove that I am dead," that he no longer needed to eat or sleep and was condemned to a kind of half-life, with a dead brain in a living body. He acknowledged that his abilities to see, hear, think, remember and communicate proved that his mind must be alive: he could not explain how his mind could be alive if his brain was dead, but he

---

[74] Jules Cotard (1882) introduced the term 'délire des négations' in referring to cases "in which patients show a marked tendency to denying everything," and this term has been translated as 'nihilistic delusions.' However, e.g., Berrios & Luque (1995a) point out a major difficulty in this translation, since French 'délire' has more complex meaning than English 'delusion': "Cotard never meant it to be a thought, but instead a symptom cluster." The label 'Cotard syndrome' was introduced by Régis (1893) and later popularized by Sèglas (1897). For the term, see, e.g., Berrios & Luque (1995a, 1999); Enoch & Ball (2001).

was certain that this was the case. Psychotropic treatment had little therapeutic effect and his delusion receded only to return. (Charland-Verville et al., 2013, 1997)

Stating that she was no longer anything, the patient begged for her veins to be opened, her arms and legs to be cut off, and her body to be opened up, so that it could be seen that she had no more blood and that her organs no longer existed. (Cotard, 1882, 367)

She had the constant experience of having no identity or "self" and being only a body without content. In addition, she was convinced that her brain had vanished, her intestines had disappeared, and her whole body was translucent. She refused to take a bath or shower because she was afraid of being soluble and disappearing through the water drain. (Debruyne et al., 2009, 197)

He often said, "I am a ghost" or "I am dead" and sometimes denied the existence of certain parts of his body saying, for example, "I have no blood." (Enoch & Ball, 2001, 165)

At one point he had stated that he was a ghost and that no one could see him. He explained that his physical body had been transformed into the immaterial body of a ghost. Nevertheless, he indicated that he visualized his body without any difficulty. About 2 months prior to admission he had concluded that eating was unnecessary since he was already dead. (Silva et al., 2000, 188)

The reports above illustrate clearly how the experience of CS patients is highly anomalous. In one sense, they claim that some of their parts are dead, but, in another sense, they still recognize the evident presence of their bodies and consciousness as in some way transformed. Generally, CS involves *self-oriented delusions*: most nihilistic delusions concern oneself rather than the world (Billon, 2016). Billon (2016, 357-358) further defines the nihilistic delusions to the following categories:

Desomatization delusions concern the body and are most frequent. Patients deny having real and lively bodily organs ("I used to have a heart. I have something which beats in its place . . . I have no stomach, I never feel hungry" or "My insides are rotting").

Dementalization delusions concern the subject's mind. Many patients deny having memories or feelings. Some deny having any thought at all ("If I had a brain I would think, but I do not think") or even a mind ("I don't have a mind").

Death and nonexistence delusions. In severe forms, patients deny being alive, and they may even negate their very existence.

World-oriented delusions. Patients can also deny some features to things other than themselves. Some otherwise believers deny that God exists; others deny the existence of their family, their doctors, their city, their marriage or even the whole world.

Time-oriented delusions concern the existence of time or some of its essential properties (for example, the property that it "passes") and are not uncommon.

Other symptoms. In addition to these nihilistic delusions, patients have typically suffered at least from anxiety, depersonalization, and depression or schizophrenia.

Altogether, patients suffering from CS feel that they are not really involved in their action but function like mindless and heartless robots or zombies and just watch themselves from outside. They even have stopped using the first-person pronoun 'I' and refer to themselves in the third-person: for instance, one patient called herself "Madam Zero" in order to emphasize her nonexistence (Metzinger, 2004).

Cotard syndrome was first described more than a century ago by French neurologist Jules Cotard (1840-1889) and later named after him. However, it has been debated whether Cotard's

description of combinations of nihilistic delusions comprises a discrete pathology or if the syndrome rather is symptomatic of other pathological conditions (Enoch & Ball, 2001, 169). In current classification systems, the syndrome as such is not cataloged, but it has to be diagnosed as part of an underlying disorder. Yet it has been argued that even if CS was more like a symptom of its underlying condition, its recognition as a phenomenological entity is important, since the syndrome has more characteristics than simply the delusion of being dead (which is also what Cotard originally meant; see, e.g., Berrios & Luque, 1995a; Huarcaya-Victoria et al., 2020). Further, for instance, Debruyne et al. (2011) pay attention to the very specific presentation of the syndrome and argue that a better understanding of the psychopathology would be beneficial in the treatment of the patients.

Since Cotard syndrome has not been classified as such in psychiatric systems, detailed and systematic information of its incidence, etiology and pathogenesis is still lacking. As CS lacks classification in psychiatric manuals, not all cases may have been recognized, or they might have been diagnosed as some other disorder. The literature on the syndrome is still "scarce and scattered" (Debruyne et al., 2009) and for the most part restricted to case reports. The studies that give information about the prevalence are conducted in psychiatric institutions and present the incidence of CS at less than one percent in psychiatric patients (Debruyne et al., 2011; Ramirez-Bermudez et al., 2010). When this is generalized to the whole population, it seems evident that CS is very rare.

A further reason for the poor understanding of CS is that particular cases can vary greatly from each other. Patients have different delusions: one can be convinced that she is dead and wants to get buried, while other experiences having no identity and being a translucent body without content (Debruyne et al., 2009). Some researchers have explicitly classified different types of CS and argued that it is beneficial to be able to distinguish the different versions. For instance, Berrios and Luque (1995b) specified three factors or versions of CS: a) psychotic depression, where patients suffer from anxiety, depression, auditory hallucinations, and only

few nihilistic delusions; b) Cotard type I, where patients suffer from a variety of nihilistic delusions; and c) Cotard type II, where patients suffer from anxiety, depression, delusions, and auditory hallucinations and thus constitute a mixed group. More recently, for instance, Huarcaya-Victoria et al. (2020) conducted a review giving partly overlapping results of these three types of CS cases; a) psychotic depression, b) delusions/hallucinations, and c) mixed. In addition, the successful treatment of particular cases or versions of CS varies: even if a treatment relieves the symptoms of one patient, it might not work for another. In recent years, interest towards CS has increased: new reviews have been written, more cases analyzed and new neuro-imaging methods have been used in the examination of CS (Charland-Verville et al. 2013; Debruyne et al. 2009; Huarcaya-Victoria et al. 2020). Thus, there seems to be a growing consensus that it would be beneficial to study the syndrome more closely.

6.3. Depersonalization

Interestingly, many of the symptoms of CS patients are also present in another mental disorder, though in a considerably milder form. This condition, *depersonalization*, can be defined as "alienating feelings of detachment or estrangement from one's self, one's body and/ or one's surroundings" (Ciaunica et al., 2021, 2; Dugas, 1898; Medford et al., 2016; Sass et al., 2013; Sierra, 2009; Sierra & David, 2011). Depersonalization involves "a disturbing change in the quality of subjective conscious experiences" (Ciaunica et al. 2021, 2); characteristic symptoms "include emotional numbing, loss of sense of agency, inability to focus, feelings of the unreality of the external world, altered experience of the body, time, and space, and heightened self-observation" (Sass et al., 2013, 433; Simeon & Abugel, 2006). The feelings of the unreality of the external world are referred as "derealization."

The resemblance between depersonalization and CS is so clear that it has been noted in several accounts of CS (e.g., Billon, 2016, 2017; Enoch & Trethowan, 1991; Gerrans, 2015; Radovic, 2017;

Radovic & Radovic, 2002; Young & Leafhead, 1996).[75] The difference between CS and depersonalization is that patients suffering from depersonalization are not delusional: they do not claim that their selves actually are nonexistent but that they feel *as if* their selves are unreal. Instead, patients suffering from CS have entered delusions and present stronger straightforward claims about their nonexistence. I agree with the point presented in some CS studies that it is beneficial to note the similarities between depersonalization and CS, since a picture of the more common depersonalization can give tools to understanding CS. Further, depersonalized experiences point out that the feelings in nihilistic delusions are not limited to the pure cases of Cotard delusion. Since depersonalization might be considered a simpler version or a preliminary stage of many CS cases, it is useful to view the features of depersonalization first.

Transient episodes of depersonalization are rather common, with an estimated lifetime prevalence of 26–74% in the general population (Hunter et al., 2004; Michal, 2009). However, the experience of depersonalization can turn into a chronic condition that is diagnosed as Depersonalization/Derealization Disorder (DPD in brief), with an estimated prevalence of 1–2% (Tibubos et al., 2018).[76] It has been proposed that transient experiences of depersonalization can be a normal and adaptive response in stressful and traumatic situations, while chronic depersonalization may result from a catastrophic appraisal of these normally transient

---

[75] On the relation between CS and depersonalization, Radovic (2017, 681-2) remarks, "Roughly, there seem to be two positions in the debate on how depersonalisation relates to Cotard's delusion. First, there is the view that there is no qualitative difference, only a difference of intensity, between a non-delusional state (depersonalisation) and a delusional state (Cotard's delusion). Second, the alternative view stresses that Cotard's syndrome includes a rather specific kind of phenomenology that distinguishes it from the symptom cluster typical of depersonalisation. For instance, Enoch and Trethowan (1991) seem to lean towards the latter interpretation ... whereas Young and Leafhead (1996) and Billon (2016b) argue that depersonalisation and Cotard's syndrome share the same phenomenology, in the sense that there is no distinct qualitative difference."

[76] Further, the experiences of depersonalization can be seen as a syndrome that involves four dimensions (within the 29 items of the Cambridge Depersonalisation Scale): 1) anomalous body experiences; 2) emotional numbing; 3) anomalous subjective recall; and 4) alienation from surroundings. See Sierra et al. (2005); Sierra & Berrios (2000).

experiences and leads to increased anxiety and stress (e.g., Hunter et al., 2003). In addition, depersonalization can be experienced in other mental disorders such as schizophrenia, depression and post-traumatic stress disorder.

The character of depersonalization, alienation from one's self, shows it to be a profound alteration of self-awareness. Since experiential alterations in depersonalization involve an attenuation of self-presence, it has been considered a disturbance of minimal self-consciousness especially (e.g., Sass et al., 2013). However, as Ciaunica and colleagues (2021, 2) indicate, depersonalization concerns multiple aspects of self, which they describe in following terms:

> a) Low-level sensory and bodily aspects of the self (detachment from one's body or body parts);
>
> b) Experiential aspects (detachment from one's subjective feelings and emotions); and
>
> c) High-level cognitive and narrative aspects (disconnection from one's personal stories, memories, thoughts and future plans). These high-level alterations have been described as a loss of the narrative flow or "plot" in patient's life.

In examining the experience of self-detachment in DPD, Ciaunica and colleagues (2021) use the distinction between reflective and pre-reflective self-consciousness which coincides with the distinction between MSC and RSC. Ciaunica and colleagues (2021, referring to Zahavi, 2005; Fuchs, 2015) describe pre-reflective self-awareness as a "general medium through which our bodily sensations, movements and thoughts are integrated and articulated as a unitary whole." In daily life, this pre-reflective sense of self is taken for granted and it does not require explicit attention. In other words, it is "transparent" and the transparency enables the feeling of being fully immersed in the world. Ciaunica and colleagues (2021, 15) present two key ideas concerning depersonalization: 1) DPD alters the transparency of pre-reflective sense of self and 2) DPD alters the capacity to flexibly connect and switch between a)

the I- and body-as-subject and b) the I- and body-as-object of an experience. In other words, depersonalization involves a "divorce" or a "fracture" between an observed and an observing self.

Furthermore, Ciaunica and colleagues (2021) argue that changes in the pre-reflective sense of self can highlight its fundamental nature in everyday experiences. Ciaunica and colleagues (2021) describe the tacit transparency of pre-reflective self-consciousness by means of the metaphor of a crack in a window glass: Since the glass is transparent, one might not notice that there is a window in the first place. However, when the glass cracks, one becomes aware of the window; the window does not disappear but the crack disrupts the transparency. Similarly, depersonalization "cracks" the transparency of one's embodied sense of self; the sense of self reappears in a modified form and is felt with a continuous disruption.

These two pathological cases, depersonalization and CS, are particularly interesting for the philosophy of self-consciousness since altered sense of self is their key symptom. This kind of abnormal self-experience offers an opportunity to scrutinize components and dynamics of self-consciousness that remain tacit in everyday experience. Below, CS is analyzed with the same kind of approach that Ciaunica and colleagues (2021) showed to be helpful in the study of depersonalization. The approach explores the subtleties of self in pathological cases, which can especially reveal the normally tacit MSC and the dynamics between different forms of self-consciousness. However, the analysis of CS here is more detailed, involving multiple aspects of self within MSC and RSC and elaborating their relations.

6.4. Previous models highlighting RSC in concern of Cotard syndrome

In the following, I will use Gallagher's pattern theory (2013; Gallagher & Daly, 2017; see Sec. 1.1.3.) as a heuristic framework by which I approach different interpretations for CS. I consider which aspects of self have been disordered in CS and whether some aspect(s) or relation(s) between the aspects is particularly

crucial in the propagation of CS. In other words, I aim at defining what kind of distorted pattern of self characterizes the syndrome. The nihilistic delusions in CS are interesting for philosophy of mind since they seem to contradict with profound structures of typical self-experience. Although the condition of CS is rare, it has received attention from neurophilosophers, starting from Metzinger (2003, 2004) and including, for instance, Billon (2016, 2017), Gennaro (2020), Gerrans (2015), and Radovic (2017). However, different writers have focused on different features of CS or on some specific theory in philosophy of mind. The pattern theory of self has not yet been applied to CS, but it has the advantage of being able to provide an integrated picture of CS.[77] In addition, the pattern theory is useful since it enables comparison and conjoining of different explanations of CS.

In this subchapter, I focus on insights and shortcomings of the interpretations of CS that emphasize RSC (including psychological/cognitive, reflective, and narrative aspects of self). In the next subchapter I consider interpretations that consider more MSC (involving embodied, experiential, and affective aspects of self) and importantly supplement the appeal to RSC. In Section 6.6., I argue that in addition to the explanations of CS that highlight disturbance in some particular aspect of self, it is important to recognize that the connections between MSC and RSC are dysfunctional in CS.

The dramatic character of CS clearly involves changes in the content and functioning of RSC. According to Newen's (2018) brief consideration, CS involves phenomenological disembodiment and radically diminished affective flow "even though the cognitive ability to form self-conscious thoughts is still intact." It is true that a CS patient can exercise her RSC and take herself as the object of thought. However, one should not settle with this remark since self-conscious thoughts also go through significant alterations in CS, and it is useful to analyze these alterations in order to understand the syndrome. In contrast to the normal familiar and active agent,

---

[77] Newen (2018) briefly mentions CS in terms of his pattern theory of self; however, Newen focuses on development of the pattern theory, not an analysis of the syndrome.

the self (or various features of self) appears dead or nonexistent in CS. Further, RSC is unable to change this content although the function of RSC is to operate on self-conscious thoughts and modify them. Below, I consider several malfunctions of RSC in CS. I argue that general deficits in psychological-cognitive aspects do not explain CS. Instead, CS involves specific distortions in self-reflection, including alterations in or loss of the use of first-person expressions. Further, I consider delusions of being dead in terms of narrative aspects and as an attempt to capture the feelings of not being alive but only the inauthentic remains of a self.

6.4.1. Problems in psychological-cognitive aspects

Since RSC is connected to general cognitive capacities of reasoning, one explanation for CS could simply be impaired reasoning or rationality. Indeed, the claim that one is dead seems blatantly untrue since one can still examine her movements, see the outer world and her portrait in a mirror, and undergo mental states—all these observations offer clear proofs that she does exist and is alive. However, these proofs do not convince the patient; she keeps on claiming her nonexistence. As Metzinger (2003, 455) points out, CS violates the global logical coherence of the patients "web of belief" and the "nature of a specific belief content...initially raises serious doubts about the patient's status as a rational subject." Accounts of CS can take this line and give a discontinuous interpretation that considers the syndrome simply as irrational thinking (see, e.g., Berrios, 1991).

However, it is more fruitful to consider a continuous interpretation of CS that does make sense of patients' claims (see Sec. 6.1.1.). That is because a continuous interpretation shows that patients' abnormal experiences provide them reasons for their delusional claims; a patient develops a conclusion from a dramatic shift in her self-experience. Metzinger (2004, 317) also agrees with this and remarks that a closer look at the data reveals that a CS patient counts as a rational subject. Thus, "something that seems an a priori impossibility on logical grounds—a conscious subject truthfully denying its own existence—turns out to be a phenomenological reality. And phenomenology has to be taken

serious." Further support for favoring a continuous interpretation of CS is that the patients' delusions are usually highly specific and that the feelings of unreality of self can also be found in other conditions. As Billon (2016), for instance, highlighted, patients with CS usually have normal procedural rationality: they are capable of reasoning according to the laws of logic and probability theory. Thus, independent of their strange claims about nonexistence, no reason exists to assume that they would be so irrational that we could not make sense of what they are saying. In addition, some perfectly rational people say that they feel as though they do not think or exist. These people are patients suffering from depersonalization, which was discussed in Section 6.3.

Thus, it seems that continuous interpretations do not appeal only to RSC but ascribe CS to a distortion of MSC in the first place. The idea is that alterations in MSC result in delusional beliefs in RSC. It is clear that RSC is dysfunctional in CS, but according to a continuous interpretation, the syndrome cannot be accounted for only by observing the contents of the delusional beliefs but also by tracing the roots of the delusions. In other words, the experience of CS patients is not caused in a top-down manner by strange beliefs (as, e.g., Campbell, 2001 argues; for criticism of Campbell, see Billon, 2016). However, the dysfunctions of RSC are worth further analysis in order to understand better both the experiences of patients suffering from CS and the functioning of RSC.

One suggested psychological factor that is explanatory in the outbreak of CS has been attribution style. For instance, Critchley (1964; used in explanation of CS by, e.g., Young & Leafhead, 1996) analyzed bizarre psychiatric syndromes associated with parietal brain dysfunction and suggested that the role of premorbid personality characteristics is crucial in these syndromes. The progression of CS is more likely in patients who have an internal attribution style, which is often present in depression. By contrast, an external attribution style is present more often in paranoia and has been connected to the progression of Capgras syndrome, which involves delusions regarding identical impostors who have replaced familiar people. The idea is that both of these bizarre syndromes originate from cognitive dysfunction in the process that

associates face and body recognition with an affect of familiarity. Without the normal feelings of familiarity, patients experience unusual feelings of derealization and depersonalization. Debruyne and colleagues (2009) note that whereas this hypothesis has been confirmed for Capgras syndrome, empirical evidence on this dysfunction has not been established for CS. Thus, the internal attribution style (together with the lack of the feeling of familiarity) might be one factor underlying CS, but it is not decisive. Even if the role of self-misattribution was a component of CS, it cannot explain all features of the anomalous experience, including dementalization and nonexistence delusions (see also Ramirez-Bermudez et al., 2010) and thus, more explanatory features are needed.

6.4.2. Changes in reflective aspects

CS involves specific dysfunctions in self-reflection. Typical RSC is approached by first-person thoughts and the use of first-person expressions, but CS is characterized by the loss of first-person self-expression. Instead, the beliefs of nonexistence in CS are accompanied by third-personal thinking and observing of oneself, and a tendency to avoid the use of the first-person pronoun. As Billon (2017) describes, patients suffering from CS (and depersonalization) do not feel like 'I's and instead consider their condition as something that cannot be referred to in the first-person in a legitimate way. According to Billon (2017), it should be elaborated that many CS patients frequently, or even systematically, avoid the use of first-person. [78]

---

[78] According to Billon (2017, 738-41), depersonalized patients also share the feeling that they should not refer to themselves using the first person and "[do] not clearly feel like an 'I.'" Billon (2017) lists four ways that depersonalized patients use in expressing this "not being an 'I'": 1) Being multiple: it feels to a patient that she is not unified enough to be an 'I' but is instead multiple or plural (this is called nosism). For instance, "I used to say 'we' rather than 'I.'" 2) Third-personal access to oneself: many patients feel more entitled to refer to themselves in third person (this is called illeism). 3) Illeist and nosist inclinations. Sometimes patients do use the first-person correctly but they say that they have an inclination to do otherwise (to use nosist or illeist reference instead). 4) Robot-like uses. Sometimes patients explicitly highlight that their use of 'I' is not standard, but something is amiss in the referent. Patients compare their use of the word to machines: also a machine could use the word, although it is clear that the machine cannot seriously be considered as an 'I.'

Billon (2017, 742) remarks that some CS patients refer to themselves by using the third-person pronouns 'she' or 'he' or proper names. For instance, one patient substituted her name 'Madame Leblanc' for the first-person (Levassor & Dromard, 1908), and another famous patient called herself 'Madam Zero' in expressing her nonexistence (Léger et al., 1969). Some patients even use the impersonal 'it' or 'this' instead of personal pronouns; one patient described herself 'this is empty', (Séglas & Meige, 1895, 475). And another patient, 'père Lambert', replaced 'I' with 'this': "You ask how the père Lambert is going? But the père Lambert is not there anymore, this is a machine made to his resemblance" (Falret 1864, 287–9).

Further, Billon (2017, 743) argues that even when CS patients use the first-person 'I', their uses are "not standard and comprehending." When a patient is confronted with questions of how she can walk and talk if she is not alive, as she claims, a patient can answer something like "It is true that I talk, walk and work, but it is like an automaton" (Séglas and Meige, 1895, 667). Billon also remarks that many CS patients use the first person only in negative phrases. That is, in the claim "I don't exist," the use of 'I' can be interpreted as metalinguistic negation rather than descriptive; it expresses something like "'I exist' is not true" or "'I exist' is not appropriately assertable." In other words, 'I' is being mentioned rather than used in its ordinary sense. These abnormalities in CS patients' language use can be considered dysfunctions of reflective aspects of self. They are worth noticing and indicate distortions in RSC but need to be connected to other dysfunctional aspects of self in order to give a full account for CS.

6.4.3. Changes in narrative aspects

Radovic (2017) presents a promising interpretation of spelling out the experiences that patients undergo in delusions of CS. Radovic argues that the phenomenology of deadness can be traced to a sense of inauthenticity (or artificiality) as opposed to authenticity (or genuineness). The concept of authenticity helps to understand better how "it involves no real paradox to affirm one's own death, from the patient's point of view. Life in the ordinary natural sense,

as the patient knows it, does not exist anymore" (Radovic 2017, 679). Radovic's account is also advantageous for encompassing several earlier models of CS, including Ratcliffe (2008) and Young (2012).

Although Radovic (2017) does not refer to a narrative, his account is discussed here under the heading of narrative aspects, since it gives a wider picture of the functioning of RSC in CS patients. Also generally, narrative aspects have been considered significant in psychiatry. For instance, Daly and Gallagher (2019) argue that a narrative is a means of disclosing and connecting all the other aspects. As such, the narrative is one key in tracking the evolution of psychiatric disorders, and it is beneficial to acknowledge narrative aspects. Applied to the case of CS, it needs to be noted that instead of presenting sporadic claims of their death, patients try to express and make sense of their anomalous experience by forming beliefs and weaving the best narrative whole from those. Further, the observation of the narrative aspects shows that a patient's narrative has not ended altogether (as would be the case in really being dead), but it has altered as a story of an inauthentic being. This supports a continuous interpretation and aids in seeing that the patients have not lost their cognitive capacities; on the contrary, they are trying to give a rational account for their anomalous experience.

In Radovic's (2017) interpretation, a CS patient feels a distressing sense of not being alive and, accordingly, conceives herself as "an inauthentic proxy of an earlier authentic individual who has ended natural life or no longer possesses an authentic self." In claiming that she is dead or nonexistent, a CS patient means rather that "being alive is an inadequate description of my present state, hence I am dead." Allusions to death do not mean to exclude all the everyday signs of life but signify the lack of authenticity in the feeling of being alive. Inauthenticity implies that the proper "nature" is lacking from the self: "it looks right but feels wrong" (Ramachandran & Blakeslee 1998). Radovic's (2017) interpretation also helps to understand CS patients' references to fictional characters such as ghosts and living corpses. These characters are considered to meet the feeling of being an inauthentic personality

that is not a complete living human being but alive merely in an imitative sense.

Radovic (2017) notes that the concept of inauthenticity relates to terms like "unfamiliarity" and "misidentification" but gives a more exhaustive account of the experience of CS patients. "Inauthenticity" significantly includes the descriptive element of the tension between a superficial appearance and a deeper nature. According to Radovic, the notion of a feeling of inauthenticity elaborates the descriptions of CS such as, "We must think of the experience as signifying an alteration in one's existential state, rather than its total negation" (Young 2012, 133) and "The sentence 'I have ceased to exist' ... What is lost is the sense of existence that ordinarily operates as a background to all experience" (Ratcliffe 2008, 169).

To sum up, Radovic (2017, 697) argues that a CS patient suffers from the feeling of inauthenticity and accordingly describes herself "as some deficient remains resulting from a metamorphosis." In this metamorphosis, the patient's genuine or authentic self has been replaced by an inferior ersatz self. She denies that manifest signs of being alive, such as walking and talking, represent a proper natural life. Instead, the patient feels she is in a condition of quasi-existence that mimics a proper existence: She apprehends the corrupt remains as the inauthentic kind that cannot be considered a self. This interpretation has the advantage of being applicable to cover various versions of CS; although the clinical picture between patients differs and they undergo different delusional experiences (from desomatization delusion to death delusion and time-oriented delusions), the feeling of inauthentic being characterizes all these delusions.

Altogether, CS involves significant malfunctions of RSC concerning psychological-cognitive, reflective and narrative aspects of self. The notion of inauthenticity might be considered as a collecting concept that covers many accounts of CS and also manages to explain the peculiarities in self-reflection. Yet, the concept is rather general, and in order to understand CS in more detail, we next look at various continuous interpretations that focus

on minimal self-consciousness (the analysis of RSC will be returned in Sec. 7.2.).

6.5. Previous models highlighting MSC in concern of Cotard syndrome

In this subchapter, I continue the examination of insights and shortcomings of models of CS and extend consideration to the aspects that are involved in MSC.

6.5.1. Problems in minimal embodied aspects of self

The most frequent feature of CS is desomatization delusions, which concern the body, and thus, CS noticeably involves altered bodily self-consciousness. Some explanations of CS highlight this. For instance, Radovic (2017) discusses the disturbed sense of body-awareness that can be present in CS in two different ways: in some cases, the patient feels reified or de-animated like a mere thing, whereas in other cases, the patient is not able to sense her body.

However, as presented in Chapter 2, the minimal embodied aspects of the pattern theory (Gallagher, 2013; Gallagher & Daly, 2018) are rather non-conscious and as such cannot be considered a form of self-consciousness. Thus, the feelings involving altered bodily self-consciousness are accounted here as features of self that are experienced and involve experiential or affective aspects (Secs. 6.5.3. and 6.5.2.), and the minimal embodied aspects are considered in terms of neural functions. The quick look into the neural dysfunctions illustrates that the extensive changes in self-consciousness correlate with extensive changes in neural processing in CS patients. The evident neural changes support the continuous interpretations that the patients are not just playing with their beliefs, but their condition involves wide malfunctioning in the structures of self, including MSC.

Cotard syndrome has been associated with a range of neurological conditions (see, e.g., Ramirez-Bermudez et al., 2010; Swamy et al., 2007). Generally, the pathophysiology of CS is characterized by an effect on the frontotemporoparietal circuitry (Debruyne et al., 2011; Swamy et al., 2007). Swamy and colleagues

(2007, 99) summarize that the most common neurological abnormalities in CS involve: 1) structural brain in relation to bilateral cerebral atrophy; 2) functional brain changes in relation to hypoperfusion in the frontal and parietal cortices; and 3) neuropsychological disorders that include impaired face recognition. Especially, Swamy and colleagues (2007) note that a disturbing focus in the right frontal and temporal lobes superimposes on widespread cerebral atrophy. The focus on the right hemisphere is consistent with the research that links the origins of content-specific delusions to the right hemisphere, especially the frontal lobe (Swamy et al., 2007; see, e.g., Malloy & Richardson, 1994).

Further, the first PET study of a CS patient showed a complex pattern of hypo- and hypermetabolism (Charland-Verville et al., 2013). The former localized especially in the dorsolateral prefrontal cortex and cingulate regions, the latter in regions of the medial and inferior frontal cortex, thalamus, basal ganglia and cerebellum (as compared to age-matched normal controls). Charland-Verville and colleagues (2013) highlight this cortical hypometabolism in midline and dorsolateral regions to be much more severe and widespread than in major depressive disorder. They also note that the hypometabolic regions are critical for conscious awareness and key parts of the default-mode network that have been considered central in self-related processing (see Sec. 4.1.1.). Accordingly, they conclude that Cotard delusion reflects a profound disturbance in brain regions that are responsible for our abiding sense of self.

That is, CS is characterized by widespread cerebral atrophy and extensive cortical hypometabolism. Thus, interestingly, when patients claim that their brains are dead, in a sense they are literally right: their brain activity is extremely low! Yet I want to emphasize that the loss of the sense of being a self is not explained simply due to reduced neural activity, since some neural areas are characterized by hyperactivity, and the wider patterns of neural activity are also disordered in CS (these are considered again in Sec. 6.6.2.). Further, a neural explanation alone cannot account for the whole phenomenon, but a complete account of CS essentially

requires explanatory factors that involve a first-person experiential point of view.

6.5.2. Lack of affective aspect?

The classical neurocognitive accounts (e.g., Gerrans, 2002; Ramachandran & Blakeslee, 1998; Young & Leafhead, 1996) have argued that the CS involves a diminution of affective phenomena. This flattened affectivity has been associated with face recognition (Young & Leafhead, 1996), wider perceptual processing (Ramachandran & Blakeslee, 1998), and cognitive processing (Gerrans, 2003). The idea in these classical accounts is that a substantial attenuation of affects deprives perceptions and thoughts of "emotional coloring." In Billon's words, the classical interpretations present a continuous interpretation of the delusions of CS: A disruption of some affective processes causes abnormal phenomenal experience "which causes, through endorsement or rationalization and maybe because of some cognitive biases as well, the delusion that P" (Billon, 2016, 363). Thus, these classical accounts seem to consider CS as a condition in which affective aspects of self are missing, and the lack of affectivity accounts for the feelings of nonexistence.

However, although CS involves disruptions in affective processes, an appeal only to affectivity seems insufficient to explain CS patients' experiences. The mere absence of affective features does not seem to provide a global experience that is described by claims such as one does not think or even such as "I am dead" or "I do not exist." Billon (2016) argues that the classical neurocognitive accounts cannot justify beliefs with such contents and thus cannot provide a convincing continuous interpretation. Even if the appeal to lack of affectivity could explain desomatization delusions, it has difficulties explaining dementalization, death and nonexistence delusions.

In addition, I want to point out that it seems evident that CS patients do not lack affectivity altogether: they are still reacting to stimuli and demonstrably they *do* feel gnawing anxiety about their abnormal experience. Thus, it seems that although the affective

aspect of self is distorted in CS, the lack of it cannot alone explain patients' delusions.

6.5.2.1. Gerrans's elaborated account

The appeal to lessened affectivity in CS has been elaborated in more recent interpretations. For instance, Gerrans (2015) sees the lack of affectivity as a central component in wider theoretical frameworks that can explain CS. Gerrans's (2015) explanation of CS draws from neurophysiology, the appraisal theory of emotion, and predictive coding theory.[79] According to Gerrans (2015, 3), in self-awareness, the organism undergoes episodes of memory and foresight for self-relevance, and this narrative capacity bears affective significance, grounding the feeling of being a temporally integrated "extended" self with a past and future. In other words, affective self-relevance is present in all levels of self from the simple interaction with the world to the higher levels of narrative cognition. With this conception, Gerrans (2015) emphasizes the role of affectivity in both minimal and reflective self-consciousness.[80] Gerrans's (2015) account gives an extensive description of the dysfunctions in self in CS and depersonalization, and it is discussed below.

 According to the appraisal theory of emotion, appraisal systems encode the significance of information for the organism. These systems convey affective significance of encounters with the world to the subject of experience. According to Gerrans (2015), the neural mechanism of the appraisal system is the Anterior Insular Cortex (AIC in brief) that encodes emotional significance of body states (see Sec. 4.1.2.). Referring to Craig (2009, 67), Gerrans (2015, 3) notes that AIC represents a sentient self that provides the basis

---

[79] Later, Gerrans (2019) elaborates the predictive processing approach to depersonalization. Gerrans (2013) also discusses delusions and aberrant predictive processing in terms of wider abnormalities in the default-mode network. In addition, in the same way, Seth et al. (2012) consider DPD in terms of predictive processing and suggest that DPD arises from imprecise interoceptive predictions.

[80] However, in the definition above, Gerrans's concept of 'self-consciousness' seems to be rather strong, referring to a reflective sense involving narrative thinking about oneself. Thus, it is slightly unclear whether his conception of self-awareness is thin enough to meet the notion of MSC as subjective character of consciousness.

for the continuity of subjective emotional awareness in the present. Gerrans (2015) highlights the extension of the temporal range of affectivity by integrating it with representation of the past and future episodes of experience and semantic knowledge.

Gerrans (2015) continues his explanation with the predictive processing theory[81] and formulates a predictive processing account that highlights the role of affectivity in experience. According to the predictive coding theory, the "mind is organized as a hierarchical system that uses representations of the world and its own states to control behavior. ... All levels of the cognitive hierarchy exploit the same principle: error correction" (Gerrans, 2015, 8, referring to Friston, 2003; Hohwy, 2013; Hohwy et al., 2008; Jones & Love, 2011). A cognitive system models its domain in order to predict its current and future states; when predictions are satisfied, the model is reinforced, and when they are not, the model is revised and new predictions are generated. If and when there is a discrepancy between actual and predicted state (called surprisal), an error signal occurs. The signal is conveyed to a higher-level supervisory system which, by using its larger database of potential solutions, generates an instruction to cancel the error and minimize surprisal (Friston, 2003; Hohwy et al., 2008). This process iterates until error signals are canceled by suitable action.

Gerrans (2015) highlights that the affective processes represent the significance of the information for the subject, and affective responses provide the necessary personal perspective on information. According to Gerrans's (2015) account, a patient suffering from depersonalization "has normal perceptual and sensory responses to the world but those responses are not

---

[81] Although launched relatively recently (originating, e.g., in Friston, 2003, 2010), the predictive processing framework has been highly influential in the philosophy of mind (Clark, 2013, 2016; Hohwy, 2013), and it has been applied in the study of self (e.g., Kiverstein, 2020; Limanowski & Friston, 2020; Newen, 2018; Seth, 2013; Seth & Tsakiris, 2018). The predictive processing approach is useful and important for theories of self since, as Ciaunica et al. (2021, 8) note, the "idea of a hierarchy of 'self' priors within the brain's overarching model of the self supports the philosophical idea of a (lower level) non-conscious and bodily base for higher forms of self-consciousness (Gallagher 2000)." However, the predictive processing theories are presented only roughly here, mainly as a potential add to a pattern theory of self, but cannot be elaborated more.

integrated into a bodily representation which informs her of their significance. The world feels derealised or ... de-affectualised." However, in accordance with the predictive coding architecture of mind, the patient has a model that predicts activity in the AIC when she is encountering the world. Thus, absence of AIC-produced affectivity is a prediction error that triggers metacognitive responses. The metacognitive responses refer to the processes that take experience to the focus of cognition and in this case involve increased attention to the experience that lacks affective components.

Thus, Gerrans (2015) highlights that the lack of affect alone does not produce the error signal engaging higher-level cognition. The high-level response induces only when lack of affect is unpredicted. The failure to resolve the prediction error results in anxiety in depersonalization and CS (Gerrans 2015, 9–10). Attention amplifies the error signal in order to make it more precise and maintains it for higher-level systems to interpret the experience and generate appropriate responses. Without a proper predictive model or response, the processing involves anxiety. Since anxiety is driven by the need to resolve uncertainty, the highly anxious patients cannot divert attention away from the experience. However, the experience is inexplicable and irresolvable, and the patients do not succeed in resolving the uncertainty but end up with the personal-level interpretation of the experience of "it feels like it is not happening to me." According to Gerrans (2015), this interpretation "is not a direct report of the experience, which is more like a total deaffectualisation, but amplifies it." The amplification is experienced as depersonalization, which lacks the affective significance that creates the normal sense of being a self. Gerrans (2015, 11) argues that the lost affective response cannot be restored from the top down and that delusions are high-level responses to this kind of obstinate signal of prediction error.

Gerrans's (2015) model also explains the differences in casual etiology between depersonalization arising in the CS and in DPD. CS results from malfunctioning in the mechanisms that appraise perceptual and interoceptive information for self-relevance. The felt significance disappears, since the AIC is not getting any

information from affective systems to integrate and convey to higher-order cognition. When the patient focuses on this experience, she feels alienated from the world and depersonalized. By contrast, in DPD, the hypoactivation of the AIC is caused by inhibitory activity in the Ventrolateral Prefrontal Cortex (VLPFC). The VLPFC is a structure which plays a crucial role in the regulation of affective feeling; it enables the subject to redirect attention to alternative interpretations of self-relevance by inhibiting an experience that would otherwise monopolize cognition. Gerrans (2015) argues that hyperactivity in VLPFC leads to hypoactivity in the AIC, and this reduced activity produces the loss of a sense of presence. As Gerrans puts it, "In depersonalization it seems that almost all expected affective feelings are absent because of hyperactivity in the VLPFC" (Gerrans 2015, 12–14). Seth and colleagues (2012) also suggest that the transition from DPD to full-blown delusional CS may reflect aberrant high-level inferences resulting from attempts to explain away persistent interoceptive prediction errors.

Both cases, CS and depersonalization, are characterized by an involuntary deactivation of systems that produce the experience of emotion, and a patient attends to this experience, trying to interpret it in order to respond. However, the increased attention does not lead to an increase in precision (of the signal in predictive processing); the attention only makes the absence of the predicted response more salient. These predictions are representations of the expected self-relevance that normally characterizes experience of self-awareness, and the lack of this characteristic makes the patient conclude that her self does not exist. However, some of the information necessary to generate self-awareness is still in place; "the body, the world and first order representations of their interaction are all represented in experience. What is lost is a sense of the significance of those interactions for the body that mediates them" (Gerrans, 2015, 14-15).

Altogether, Gerrans's (2015) account elaborates the flattening of affectivity in CS by evoking emotion appraisal theory and predictive coding theory. In general, this kind of wide theoretical framework assists in seeing the major role that emotions have in

normal experience and in understanding why the lack of emotions causes anxiety. Since the emotions are predicted to occur, and the absence of them accounts as prediction error, that catches one's attention. Gerrans (2015) argues that the anomalous experiences in CS cannot be explained by lack of affectivity alone, but by excepted significance of the affectivity.

However, Gerrans's (2015) view seems to be too rough for explaining CS, although it gives novel ideas on the aetiology of depersonalization experiences. Gerrans (2015, 14) argues that "in the Cotard syndrome something is amiss with the mechanisms that appraise perceptual and interoceptive information for self-relevance. The AIC is not getting any information from affective systems to integrate and relay to higher-order cognition ..." The general idea is plausible; the lack of affectivity in low-level features of self captures attention of higher-level features of self and results in anxiety. However, Gerrans does not specify dysfunctions in affective systems or higher-order systems of self or their relations more carefully. In other words, Gerrans does not analyze the experience of CS in detail but gives only a rough picture of origin of the nihilistic delusions in CS and neuropsychological references to couple of brain areas. It is rather evident that CS involves diminished affectivity, which is linked to hypoactivity of AIC; however, CS also involves other complex symptoms and other complex neural dysfunctions. Thus, it seems oversimplified to account CS only in terms of lack of affectivity or only as a dysfunction of particular neural areas.

In addition, Gerrans's model entails the acceptance of predictive processing theory, which is maybe not necessary for accounting for CS and might also bring apart from an extensive enough framework to explain CS. It has been proposed that the predictive processing framework is compatible with the pattern theory of self and could be used in explaining psychiatric conditions (Gallagher & Daly, 2018; Newen, 2018). However, these propositions are very general so far, and it seems that a more detailed account of connecting the pattern theory of self, dysfunctions of self, and predictive processing theory is still needed. This kind of elaborated view could provide a fuller assessment of

the possible usefulness or incompatibilities of connecting these theoretical frameworks, but it cannot be conducted here.

6.5.3. Problems in experiential aspects

It was presented above that an appeal only to RSC cannot fully account for CS, but MSC also needs attention. And since the disruptions in the affective aspect could not account for the most extreme delusions in CS, it seems intuitive to appeal to experiential aspects of self in interpreting the syndrome.

Experiential aspects of self are the core of MSC, that is, the subjective character of consciousness (by contrast to the qualitative character that refers to contents of consciousness, as presented by Kriegel, 2006, see Sec. 2.2.1.). According to one kind of interpretation, it is the very subjective character that *disappears* in CS patients and this disappearance accounts for their delusions of nonexistence. This seems to be the idea, for instance, in Billon and Kriegel's (2015) considerations in which CS appears as an example of access consciousness without typical phenomenal consciousness: patients would have only qualitative contents of consciousness without subjectivity. This kind of interpretation would also undermine the theories of MSC which claim subjectivity to be a necessary feature of experience.

However, under closer scrutiny, this kind of explanation simply does not work: the idea of the subjective character is that it is present always when one is undergoing experiences. And CS patients clearly are having experiences: exactly the anomalous experience makes them feel so anguished. Even when an experience is not recognized as "mine" or it does not involve its typical familiarity, it is not a non-experience but an anomalous experience. Even if a patient feels that she is not connected to her body (or her body is rotting or an automat), feels that she is not an authentic agent of her own mind ("I do not have a mind"), or feels separated from her earlier autobiographical self ("I do not exist"), she is still the one who is having these uninviting feelings. Even if her judgments are delusional, she exists as the subject of these thoughts and experiences. Surely, her self-experience is severely altered and diminished from the normal. However, her MSC still

200

remains; she does have a 1PP to the world. She is able to perceive the outer world and she is able to view her stream of consciousness in time. And crucially, the perspective still is subjective: there is something it is like for her to undergo experience. That is, the case is not that there is not what-it-is-likeness at all in CS but rather that patients' what-it-is-likeness is radically different from normal.

Another way to explain the anomalous experience in CS is to invoke to *diminished* subjective character of experience. Billon (2016) represents this kind of interpretation and thus, has himself resigned the above-mentioned interpretation of lack of subjective character (Billon & Kriegel, 2015) and developed his account of CS further. Billon (2016) argues that the subjective character of CS patients' mental states is widely attenuated and that this can explain the core of both depersonalization and CS, namely the feeling that the self or some of its parts are missing. According to Billon (2016), the feeling of nonexistence reflects "impairment of self-awareness, caused by a deficit in subjectivity." When a patient undergoes a widespread and intense attenuation of the subjective character of experience, her self-awareness impairs severely and she ceases to properly feel herself. The attenuated subjective character results in feelings that she is not an 'I' and can thus also explain the nonexistence delusions.

Billon (2016) elaborates a promising continuous interpretation of CS that involves the attenuated subjective character and a strong appeal to the similarities between CS and depersonalization. In this interpretation:

> 1) The delusions of CS patients result from experiences that are similar to those of depersonalized patients, which are essentially characterized by an attenuation of their subjective character (desomatization, dementalization and nonexistence experiences), but also, sometimes, by an attenuation of their present and their actual character (derealization and detemporalization experiences).

> 2) The subject would take these experiences at face value (she would endorse them), believing that she lacks bodily parts, thoughts, that she is not an 'I,' etc.,

3) Given the similarity between the experiences of depersonalized patients and those of CS patients, what differentiates them and explains why the latter (but not the former) believe that she is dead, is plausibly either some kind of rationality deficit (more precisely, a cognitive bias) or the intensity of the condition. In the latter case, delusional patients have a more intense and more widespread attenuation of the subjective character.

4) Different forms of CS could be explained by differences in the extension and the intensity of the attenuation of the subjective, present or actual character of experience.

Billon's (2016) interpretation is more convincing than the previously mentioned models since it gives more extensive explanation for the delusions in CS. Billon's and Gerrans's (2015) explanations have in common the appeal to diminished MSC, but they differ in the emphasis of affectivity versus broader frames of experientiality. In addition, the obvious discrepancy between them is that Gerrans's view is committed to predictive processing theory, but it is unclear whether the predictive processing theory is necessary or if it ignores some features of CS. Generally, Gerrans's model of the lack of affectivity and familiarity earns credit for its detailed theoretical background that can point out the significance of the role of flattening affectivity better than previous models. Further, Gerrans's (2015) view seems to aim at accounting for at least some of the dementalization delusions, since he pays attention to how patients get estranged from their episodic memory and longer-lasting self. Although these details deal more extensively with CS than classical models do, it still seems that Billon's (2016) interpretation takes more explanatory factors into account and has a more elaborated conception of self-consciousness than Gerrans's (2015) account, and thus it provides a more thorough understanding of CS.[82]

---

[82] Further comparison between Gerrans's and Billon's accounts for CS would be interesting, but it is out of the scope of this dissertation. It might be argued that the appeal to reduced

However, I want to point out that Billon's (2016) interpretation might simplify the progression of CS a little by claiming that a patient just takes her anomalous experiences "at face value", since the transition to a delusional state arguably is a more complex process (see, e.g., Feyarets et al., 2021; Fuchs, 2015; Sass & Parnas, 2003). In addition, Billon's claim that depersonalization does not involve derealization and detemporalization experiences might be questioned, since these kinds of experiences seem to present in at least some depersonalization patients (see, e.g., Ciaunica et al., 2021). Yet Billon's (2016) general idea of developing a very detailed description of the experiences of CS patients is in line with more extensive theories of delusion formation and seems to give a comprehensive picture of CS.

Further, in terms of the notions of self-consciousness used in this dissertation, an unnecessarily complicatedly presented point in Billon's view is the postulation of what he calls "present and actual characters of experience" in addition to subjective character. Billon (2016) describes these characters: phenomenal states are normally experienced as "being present (as occurring now) and actual (as occurring in the actual world rather than a merely possible or imaginary world)." Billon proposes that these three characters are structural, self-locating features of experience. The subjective character locates experiences to a specific subject, the present character locates experiences to a specific time, and the actual character locates experiences to a specific possible world. According to Billon (2016), impaired present and actual characters account especially for derealization and detemporalization. However, postulating these two characters seems to be unnecessary since MSC is a subjective perspective, or 1PP, and this definition already involves the notions of 'present' and 'actual characters,' which Billon (2016) postulates. Subjective perspective *is* the temporal and spatial self-locating feature of experience in the first

---

affectivity, which Gerrans advocates, manages to account dementalization and all the delusions of CS in the end, since affectivity has a crucial role in self-consciousness. However, in the above considered versions of their models, Billon is more precise in describing the factors of CS and thus, the appeal to subjective character is vindicated as the most accurate account for CS. The usefulness of a model of this kind is indicated also in Section 7.1.

place. Thus, the key explanation of CS could be stated only by appealing to MSC that also includes the features of Billon's present and actual characters. In other words, the appeal to MSC can give a conceptually simpler account than Billon (2016).

In any case, the account for CS benefits from acknowledging the disturbances in the experiential aspect of self. This kind of account is consistent with classical accounts that emphasize the role of affectivity, but with the appeal to the experiential aspect, it is easier to make sense of some of the most extreme delusions in CS, such as the claim that one does not exist. Altogether, the examination above indicates that the best explanation for anomalous experiences and delusions of CS patients involves diminished experiential and affective aspects and distortions in embodied aspects underlying them. These distortions manifest the experiences of diminished immersion and inauthenticity that are also present at the level of RSC. In other words, CS involves dysfunctions in all the considered aspects of self, and the intertwining of these dysfunctions results in nihilistic delusions. However, since several aspects are involved, it seems that their intertwining also requires more attention.

6.6. Problems in the connections between minimal and reflective self-consciousness

Two previous subchapters evaluated the insights and shortcomings of different interpretations of CS. Each interpretation presented a relevant point but seemed to highlight some particular aspect of self. I ended up proposing, in agreement with Billon (2016), that the best interpretation explicitly recognizes the significant role of the experiential aspect along other aspects. However, to my understanding, none of the mentioned interpretations is sufficient alone, since they failed to embrace self-consciousness as a whole.

In this subchapter, I introduce a novel viewpoint on the study of CS, which investigates the dysfunctional connections between aspects of self in CS. I argue that in order to account for CS, it is not sufficient to discuss only a single aspect of self, but it is crucial to look at the interplay between the aspects. This kind of

investigation can explain CS better than previous theories since it acknowledges the whole pattern of self instead of its particular features. Within the framework of whole self-consciousness, the options for explanation of CS seem to be that it is 1) a disturbance of MSC, 2) a disturbance of RSC, 3) a disturbance only between MSC and RSC, or 4) a combination of the mentioned options: 1 and/or 2 and the resulting 3. The previous subchapters concerned options 1 and 2, and I argued that both are important but not alone sufficient to account for CS. Thus, my interpretation of CS represents option 4. That is, although the examination in this subchapter is focused on the connections between MSC and RSC, the idea is not that only a disturbance in these connections could explain CS (option 3). Instead, the idea is that in order to account for CS, both MSC and RSC and their connections need to be acknowledged.

In addition to the above-considered aspects of self, the connections between them distort in CS. The disturbances in the relations between MSC and RSC have not been explicitly highlighted in the explanations of CS, although they are present in many accounts. For instance, both Billon (2016) and Gerrans (2015) argue that CS involves distorted MSC, which causes alterations in self-reflection. This idea involves, rather straightforwardly, a reference to the crucial bottom-up connection between MSC and RSC (see Sec. 5.3.1.). In addition, the connections between minimal and reflective self-consciousness are at least indirectly referred to in many neurological studies on CS that show dysfunctions in the connections between different neural processes (Sec. 6.5.1.; e.g., Debruyne et al., 2011; Swamy et al., 2007). Further, alterations in the relations of MSC and RSC were proposed to be one key factor in depersonalization (Sec. 6.3.; Ciaunica et al., 2021). Thus, the most convincing interpretations of CS presented in this chapter so far seem to emphasize a bottom-up direction within self-consciousness. According to these interpretations, CS originates in alterations in low-level features of MSC that bring on alterations of higher-level features of RSC. This highlights the experiential and affective aspects of self but also the importance of the connections between MSC and RSC.

However, I propose that the relations between MSC and RSC in CS cannot be expounded only by the bottom-up relation, because minimal and reflective self-consciousness are reciprocally connected to each other, and the relations between them are distorted in several ways. In other words, the dynamics between MSC and RSC are altered in CS, and these alterations play a remarkable role in the syndrome. Thus, a systematic analysis of the structures that hold self-consciousness together is useful in order to form a complete picture of CS. Below, I argue that in order to account for CS, two general kinds of distortions in the links between MSC and RSC should be acknowledged: increased connectivity in a vicious circle between diminished self-affection and hyper-reflectivity (Sec. 6.6.1.) and decreased connectivity of typical, more diverse relations between MSC and RSC (Sec. 6.6.2.).

6.6.1. The (vicious) circle between diminished self-affection and hyper-reflectivity

Generally, CS involves not only distorted minimal self-experience but also continual non-voluntary and third-personal reflection of this experience. The same kinds of alterations have been described in other psychopathological conditions. Especially interesting is a conception of schizophrenia developed by Sass and Parnas (e.g., 2003, 2007). Sass and Parnas consider schizophrenia a disturbance in minimal self and propose that this "disturbance has two main aspects, which may seem mutually contradictory but are in fact complementary: hyperreflexivity and diminished self-affection" (Sass et al., 2013, 431). Since these two facets of disturbance in self seem to be present also in CS, the model by Sass and Parnas can be beneficial to apply in the analysis of CS. The theoretical background of the studies by Sass and Parnas lies in phenomenological philosophy that highlights a careful analysis of experience and is compatible with a pattern theory of self, both highlighting the multidimensional nature of self-consciousness (see Sec. 6.1.1.). For instance, Daly and Gallagher (2019) argue that a pattern theory should extend this kind of phenomenologically informed framework and expand it to other psychiatric cases. Below, I apply the ideas of self-disturbance by Sass and Parnas to

CS. [83] Their ideas have inspired the analysis of CS that I am conducting here; however, I do not directly use their model but apply it to the conceptual framework of this dissertation. The model is helpful since it recognizes alterations in the links between MSC and RSC.

Roughly, *diminished self-affection* is an attenuation of self-presence—that is, depersonalization (Sass et al., 2013, 432). According to Sass et al., (2013, 431), diminished self-affection refers to "a reduction in the very sense of existing as an aware subject or agent of action, i.e., to a diminished sense of existing as a first-person perspective on the world." Typically, self-affection is an essential but ineffable feature of MSC. One schizophrenic patient described the diminution of this feature: "I was simply there, only in that place, but without being present" (Sass et al., 2013, 431). Sass and colleagues (2013) point out that this kind of description of diminished self-affection is present not only in schizophrenia but also in depersonalization.

Hyperreflexivity, in turn, refers to an exaggerated and non-volitional self-consciousness (Sass et al., 2013; Sass & Parnas, 2003, 2007). This "focal, objectifying or alienating attention" towards oneself is directed to processes and features that are normally experienced as part of oneself and remain in the background of awareness (Sass et al., 2013, 431). Hyperreflexivity manifests itself in different ways. In operative or basal hyperreflexivity, the normal experience is interrupted by an automatic popping-up or popping-out of phenomenal processes that in everyday experience remain in the tacit background of awareness. Normally these tacit processes are involved in implicit (minimally self-conscious) self-affection, but in hyperreflexivity, the processes are experienced with an object-like quality. In consequential hyperreflexivity, the primary disturbances result in further attention and a process of self-scrutiny and self-objectification. Finally, a patient might enter a compensatory hyperreflexivity by voluntarily engaging in reflective

---

[83] The model by Sass and Parnas was developed with respect to schizophrenia in the first place. Although an elaborated comparison between CS and schizophrenia would be interesting, it cannot be conducted here.

self-monitoring in order to compensate for the deficiencies in her self-presence. However, instead of restoring the lost tacit self-awareness, the excessive self-monitoring can only exacerbate the problem by producing further objectifying experiences and deepening the alienation (Parnas & Handest, 2003; Sass et al., 2013; Sass, 1994; Sass & Parnas, 2003, 2007). Altogether, a hyperreflexive patient reflects her experience, but the reflection lacks the normal first-personal familiarity, and instead one reflects herself as an unfamiliar object.

According to Sass and Parnas (e.g., 2003, 2007; Sass et al., 2013), hyperreflexivity and diminished self-affection are "mutually implicative aspects or facets of the intentional activity of awareness" (Sass & Parnas 2003, 430). Hyperreflexivity brings something normally tacit into the focus of attention, whereas diminished self-affection is the other side of the same process—the fact that what once was tacit "is no longer being inhabited as a medium of taken-for-granted selfhood" (Sass & Parnas, 2003, 430). This double-faceted disturbance distorts the pre-reflective sense of presence, and thus, Sass and colleagues describe this disturbance as a disorder of minimal self. The description is well grounded since the disturbance is characterized by altered subjectivity, and the model of Sass and colleagues is advanced in recognizing the major role that MSC has in schizophrenia and depersonalization.

However, it seems that, by definition, hyperreflexivity can also be regarded as a manifestation of RSC. RSC is the capacity to take oneself as the object of one's cognition and as to think of oneself as oneself. Clearly, in hyperreflexivity one does take oneself as the object of her attention: "Hyperreflexivity is exaggerated self-consciousness, a fundamentally non-volitional tendency for focal, objectifying or alienating attention that is directed toward processes and phenomena that are normally experienced as part of oneself." Thus, the characterization of hyperreflexivity is consistent with the dysfunctions of RSC in CS and depersonalization discussed in this chapter. In the original model of basic self-disturbance in schizophrenia, Sass and Parnas emphasize that although hyperreflexivity includes hyperreflectivity (that is, "an exaggerated intellectual or reflective process"), it is not an intellectual, volitional

or "reflective" kind of self-consciousness at its core (see, e.g., Sass & Parnas, 2003, 2007; Nelson et al. 2009). Hyperreflexivity primarily refers to automatic operative hyperreflexivity, which can lead to further consequential scrutiny and self-exacerbating alienation, which is a more reflective form of hyperreflexivity (e.g., Sass & Parnas, 2007).

A more recent formulation of the self-disturbance model (Sass et al., 2018) emphasizes the point that both hyperreflexivity and diminished self-affection can occur in primary and secondary ways. Primary hyperreflexivity is more passive and grounded in the more foundational affliction of self-experience. Secondary hyperreflexivity is a reaction to more primary and operative experiential disruptions or traumatic environmental circumstances. Similarly, depersonalization can occur in the primary sense as a severe erosion of first-person perspective or in the secondary sense of self-protective or defensive response to trauma. Sass and colleagues (2018, also e.g., Sass et al., 2013) propose that the secondary fashion of these disturbances is also present in some non-schizophrenic conditions involving dissociation, such as depersonalization and derealization.

I do not elaborate on the self-disturbance model of schizophrenia by Sass and colleagues here and do not directly employ it in the study of CS. However, I apply it, and instead of hyperreflexivity I use the concept of *hyper-reflectivity* that explicitly grasps RSC.[84] That is, I elaborate a model of CS in which diminished self-affection is coupled with hyper-reflectivity. In other words, here the focus is mainly on hyper-reflectivity of hyperreflexivity. The concept of RSC may leave some aspects of the concept of hyperreflexivity aside, but the connecting of the concepts is well-founded in the similarities in their definitions.

---

[84] To recap, the complete self-disturbance theory is beyond the scope of this dissertation. Thus, although I use the ideas presented in the theory, I do not use the theory as such since it was developed to explain schizophrenia in the first place and may involve components that apply to schizophrenia particularly but not to CS. For this reason, I prefer the concept of 'hyper-reflectivity' instead of 'hyperreflexivity' in the original theory of Sass and Parnas. I am interested especially in the abnormalities of RSC here and thus, consider the use 'hyper-reflectivity' more appropriate.

Further, the concept of RSC is used in many theories of self, and the use of it can spread the insights of the self-disturbance model to theories that are unfamiliar with concepts such as hyperreflexivity.

When hyper-reflectivity is coupled with diminished self-affection, these two processes form a vicious circle that dominates patients' experience. According to Sass and colleagues (2013, 2018), the loop between hyper-reflectivity and diminished self-affection is characteristic in depersonalization. Although "depersonalization is, by definition, a way of feeling the self to be unreal (diminished self-affection), it apparently involves or engenders experiences involving intense self-consciousness (hyper-reflectivity)" (Sass et al., 2013, 437-8). It has been pointed out in this chapter that CS bears remarkable similarities to depersonalization and might be considered its more severe form in which patients enter a further delusional state. Accordingly, a strongly depersonalized sense of self is an essential feature of CS and thus, it seems obvious that the vicious circle, which is present in depersonalization, is present also in CS. Indeed, the feelings of estrangement from one's self (diminished self-affection) dominate the experience of CS patients and catch the patient's attention (hyper-reflectivity).

Further, the vicious circle might also play a role in the formation of the delusions of CS. When a patient faces profound alterations in her subjectivity, and the vicious circle continues for long or becomes very severe (or both), a patient can end up forming beliefs about her nonexistence. Since the patient's self-affection has been diminished and continual hyper-reflectivity keeps on reminding her how inauthentic her body, thoughts, and the whole 1PP feel, she concludes that she is dead. Only that belief seems to account for the totality of her strange feelings. When the belief has been formed, it becomes difficult to change since it seems to offer the best explanation for her experience, and living through the experience repeatedly gives support to the belief.

The idea of the vicious circle between diminished self-affection and hyper-reflectivity is also quite compatible with Gerrans's (2015) account, although Gerrans (2015) and Sass and colleagues (2013) represent different theoretical frameworks. In Gerrans's (2015; Sec.

6.5.2.) model, the predictive processing framework gives theoretical support to the anxiety that is present in the vicious circle between diminished self-affection and hyper-reflectivity. Altered MSC involves a prediction error that cannot be corrected by RSC, and thus, altered MSC remains as the un-voluntary and constant object of RSC. Further, the abnormal brain functions described by Gerrans (2015) might ground the feelings in the vicious circle. According to Gerrans, the absence of AIC-produced experience is a prediction error that drives metacognitive responses, including increased attention to the experience as a patient tries to understand it. However, the patient is unable to get further information from the experience but, due to her anxiety, cannot divert her attention from the experience. In addition, the idea of a vicious circle seems to be present in descriptions of DPD that highlight how depersonalization lacks a sense of self that normally is tacit in experience and how this lack leads to anxiety and intense attention to the diminished self-experience (e.g., Ciaunica et al., 2021). Thus, all these accounts bring forward the important interconnections between MSC and RSC; these connections can become dysfunctional in the vicious circle that occasions and maintains anxiety. In summary, these considerations of the vicious circle managed to show that CS involves not only a bottom-up direction in self-consciousness but also heavy top-down processing.

6.6.2. Decreased connectivity between minimal and reflective self-consciousness

In addition to the vicious circle, I propose that it is good to remark that another dysfunction in the connections between MSC and RSC underlies patients' abnormal experiences in CS: decreased connectivity between minimal and reflective self-consciousness. At first sight, this point might seem to be in contradiction with the vicious circle that involves increased and continual interaction between MSC and RSC. However, a general decrease in the connectivity of the two forms of self-consciousness can be seen as a complementary process of the vicious circle: while the normal activity is diminished or weakened, the abnormal vicious circle appears or becomes even more emphasized and prevalent. That is,

the vicious circle brings out an altered exaggerated connection between MSC and RSC, but CS also involves connections that are not that active.

The decreased connectivity is rather clear in terms of extensive neural hypoactivity in CS. At the neural level, (as described in Sec. 4.1.) self-consciousness involves widespread activation within several neural networks. MSC as an embodied subjective perspective has been connected to sensorimotor processing recruiting the parietal lobes (e.g., Legrand & Ruby, 2009). Meanwhile, RSC can be connected to the general E-network that includes processes of activation in the frontal and temporal lobes. It has been proposed that any kind of dysfunction in the connections within the networks of self also produces a mental dysfunction or disorder (e.g., Daly & Gallagher, 2019; Zhao et al., 2013). This highlights that the connections within the neural networks of MSC and RSC are crucial for the functioning of the whole self-consciousness—and these connections break down in CS.

Generally, CS is characterized by dysfunctions in frontotemporoparietal circuitry and extensive cortical hypometabolism (Charland-Verville et al., 2013; Debruyne et al., 2011; Swamy et al., 2007). The neural dysfunctions involve extensive cortical hypometabolism in midline and dorsolateral regions (Charland-Verville et al., 2013), including frontal and parietal cortices (Swamy et al., 2007) and AIC (Gerrans, 2015). These hypoactivations can be seen underlying diminished self-experience; distortions in MSC are related to dysfunctions in parietal regions and AIC, whereas distortions of RSC are related to decreased activation in frontal and temporal lobes. Further, the connections between different features of self are decreased. For instance, Charland-Verville and colleagues (2013) highlight extremely decreased metabolic activity in precuneus, a central node in the default mode network; normally, the precuneus shows high metabolic activity when an individual is awake and is the most connected area in the brain. When these kinds of central nodes in the neural networks underlying self-consciousness are not as active as usual, they do not convey information to other networks in the

typical way. Thus, CS involves a deficient pattern of connections binding together the features of self.

On the other hand, the decrease in the typical connections between MSC and RSC can be sustained by the process in which the vicious circle and reflection on anomalous experiences replace the typical connections. The idea that CS involves exaggerated observational hyper-reflectivity is supported by empirical data that show hyperactivation, for instance, in regions of the medial and inferior frontal cortex that can be connected to RSC (see Sec. 4.1.1.). Thus, CS is characterized by an unbalanced self-pattern in which the typical connections between MSC and RSC have been reduced and substituted with a biased vicious circle.

Altogether, empirical studies support the idea that CS involves a diminution of numerous aspects of self, and this diminution also concerns a decrease in connections within a self-pattern of a CS patient. The typically tight and diverse connections between MSC and RSC break down in CS, and this kind of deactivation indicates the significance of these connections in normal self-experience.

6.7. Summary

The phenomena involving anomalous experience, such as Cotard syndrome, produce a challenge for theories of self-consciousness that is worthwhile to take. A theory of self-consciousness should be fine-grained enough to cover the whole variety of cases of self-consciousness, and thus, examples of pathological self-experience can be seen as important test cases for theories of self-consciousness. That is, pathological cases can assist in formulating the comprehensive concepts and theories of self-consciousness.

Patients suffering from CS claim that they are dead: they do not have a living body, thoughts or a self anymore. These claims are in sharp contrast not only with our normal feelings in everyday life but also with the philosophical conception of selfhood. In philosophy, the existence of self has been the most certain thing one knows, and minimal self-consciousness has been considered to be a necessary feature of experience. Thus, at first sight, the claims of CS patients seem to be just irrational or illogical. However,

despite having peculiar delusions about their nonexistence, CS patients are still rational. This invites a continuous interpretation, which implies that the most comprehensive way to understand patients' peculiar claims happens through investigating their experience: highly abnormal first-order experiences give patients reasons to believe in their own nonexistence. In other words, alterations in MSC result in delusions in RSC.

After examining several accounts of CS, it transpired that the explanations of CS that are restricted to only to some specific aspect(s) of self are insufficient. Although these explanations give important information about some features of CS, I argued that a full account of CS needs to recognize the connections between features and inspect self-consciousness as a whole. In terms of RSC, CS involves alterations in psychological-cognitive, reflective and narrative aspects of self. In terms of MSC, CS involves alterations in embodied, affective and experiential aspects. In addition to these alterations, I proposed that CS is characterized by two kinds of dysfunctions in the links between MSC and RSC: on the one hand, the vicious circle between diminished self-affection and hyper-reflectivity, and on the other hand, decreased connectivity in the typical wider network of connections.

7. An analysis of Cotard syndrome as an extreme diminution of first-person self-consciousness

In this chapter, I elaborate an account of CS that is not limited to a particular aspect(s) of self but acknowledges the distortion in the whole pattern of self. Partly drawing from the models mentioned in the previous chapter, I formulate a multifactorial model of CS that considers patients' altered experience to result from a combination of three factors: 1) diminished minimal self-consciousness leading to 2) intensified third-personal and non-voluntary reflective self-consciousness and 3) abnormal connectivity between minimal and reflective self-consciousness. I analyze each of these three factors by using the concepts from Part I of this dissertation. I argue that together the factors form a defective pattern of self, lacking typical diversity in self-experience and flexibility to shift between different modes of self-consciousness. I also consider how this analysis of CS can help in clarifying the concepts and general theory of self-consciousness. The conceptual and theoretical clarifications of self can further assist in related research within empirical and clinical fields.

7.1. The structure of minimal self-consciousness

The experiences of CS patients can reveal structural features of self-consciousness, interestingly even about MSC, which is a tacit feature of experience and, as such, difficult to capture in everyday experience. That is, distortions in MSC can expose its presence in a typical experience. Further, the examination of CS can assist in unpacking the notion of MSC and reveal what kind of features it involves.

MSC was defined (in Secs. 2.1.-2.) as the subjective character of consciousness and subjective perspective. As a subjective perspective (or 1PP), MSC is a subjective spatiotemporal self-orientation within immediate experience. At the bottom, MSC refers to the what-it-is-likeness of experience that is a constitutive feature of experience and minimum self-consciousness. This definition gives the necessary

conditions for self-consciousness and consciousness. As Zahavi (e.g., 2010c, 2017) describes it, MSC is a very "thin" notion of self: it does not (necessarily) involve any explicit self-consciousness. However, MSC is not "thin" in a sense that would imply being totally invariant (as Zahavi remarks, e.g., in 1999, 2014, 2017; Sec. 2.4.). Instead, MSC is "rich" in involving affective and embodied components that are taken for granted in everyday experience. CS assists in bringing out this "richness" of typical MSC. On the other hand, CS highlights MSC as the necessary feature of conscious mental states: MSC is not lost even in the experience of CS patients, but it is present in altered and diminished form.

The classical accounts of CS appeal to diminished affectivity, loss of emotional coloring, as the cause of the experienced symptoms (see Sec. 6.5.2.). This flattening of affectivity can be considered diminished minimal self-consciousness, since MSC includes the affective aspects of self. Affectivity falls easily into MSC, since it does not refer to the content of consciousness in the first place but to the manner of experiencing. Typically, MSC involves affects pre-reflectively: even without paying attention to it, our everyday life is colored by affective components and typical emotional patterns. The lack of the usual affective features can explain the feelings of unfamiliarity that patients have toward themselves. When the self appears without the affects that normally are incorporated into it, the self feels unreal, more like an inanimate object than a properly existing subject.

Further, the appeal to the minimal self-consciousness takes into account features that might be excluded in the classical accounts of Cotard syndrome. As Billon (2016; Sec. 6.5.3.) points out, in addition to the delusions of desomatization and dementalization, CS patients suffer from world- and time-oriented delusions; patients feel the time or its durations to be unreal and the world non-actual. Also these symptoms are surely connected to the reduced affectivity: without affects, the time does not feel meaningful oneself and the environment is also encountered without the typical feeling of involvement. Time and space are still present in CS experience but in a distorted form, since they are felt as not significant for action anymore, and this is grasped in the delusional belief that the world

and time are non-actual altogether. However, as Billon (2016) argues, the delusions of time and space seem difficult to account by invoking a lack of affectivity alone. Instead, a more comprehensive picture can be drawn by invoking the subjective character of consciousness as a whole.[85] In other words, CS involves a general kind of erosion of 1PP; not only the normal affects have faded away, but the whole frame of subjective *perspective* has been altered in CS. The appeal to the perspectival and embodied whole of MSC in addition to affectivity can help to understand the severity of the condition of CS patients. Only very severely distorted first-order experience can give rise to claims that one is dead.

Thus, analysis of CS can contribute to a theory about the structure of MSC. Although MSC is the most elementary form of self-consciousness, phenomena such as CS show that MSC involves variance. MSC is a subjective perspective; both of these defining features—subjectivity and perspective—are needed in describing experience, and both of them can alter. 'Perspective' refers to temporal and spatial features of experience that can be altered to the extent that the world and time feel unreal or nonexistent. Further, the 'subjective' character of MSC refers to the how of experience—that is, to features of experience that are not about contents of consciousness but about the manner in which contents are present to the subject. This manner typically involves affective components; however, CS and depersonalization indicate that affective features are not necessary components of experience: patients lacking them still do have experiences. However, these pathological conditions imply that affectivity is a central feature of normal or default functional MSC, since the conditions lacking affectivity are characterized by the feeling that something essential is missing and cause suffering. Thus, the pathological cases are helpful in exposing that although MSC is a thin form of self-consciousness, the structure of MSC actually is "rich" in its manifestations since it involves embodied perspective and affectivity, which can occur in many forms.

---

[85]Or it might also be argued that affectivity is somehow crucially related to time and space, but this line of argumentation would also be about the character of MSC.

On the other hand, the pathological case of CS assists in revealing what is strictly necessary in MSC. Even with diminished affective and perspectival aspects, MSC prevails as the subjective character of experience. In other words, subjectivity cannot be reduced to affectivity or a perspective alone, but some what-it-is-likeness also prevails with flattened affects and eroded perspective. Although the experience in CS and depersonalization feels anomalous, it does not lack subjectivity altogether since it *does feel like* alien and distressing for the patient. A mental state without subjectivity would be a non-conscious state and that is not the case in CS: even while claiming that she is dead, a patient does undergo experiences and even describes the strange character of her experience. Normally, affective factors are included in the subjective perspective and make the perspective immersive with vivid self and world, but subjectivity can also manifest itself with diminished affective features. Thus, studying CS assists in specifying the character of MSC: although it typically involves affective aspects, experiential aspects are its necessary feature and, as such, the most fundamental form of self-consciousness.

To recap, it can be stated that minimal self-consciousness is not lost, but it is significantly altered in Cotard syndrome. Although patients claim that their selves do not exist anymore and do have some reasons for these claims due to their altered first-order experiences, they have not lost their minimal selves. However, their 'selves' in more everyday language indeed do seem to disappear. Self as the familiar entity that one reaches through self-conscious thinking does not feel as if it is there anymore. Instead, the self is reached by 3P descriptions and appearances, and continuance of these anomalous appearances highlights the alien feeling of estrangement.

7.2. The structure of reflective self-consciousness

The striking feature of CS is the anomalous beliefs in one's nonexistence, and these beliefs are kept as the content of consciousness and operated by RSC. Undergoing thoughts about self show that RSC of CS patients is still functioning in that a patient

can take herself as the object of her reasoning—and does it hyper-reflectively. However, when the patient thinks of herself as herself, she finds that something is disturbingly missing; her very being as a person in first person. Instead, her self-reflection occurs with diminished MSC and is characterized by the beliefs that she has ceased to be her proper self and that life has run out of her, leaving only inanimate remains. Thus, it is evident that her RSC is dramatically distorted. It was presented in Section 6.4. that CS involves alterations in three aspects of RSC: psychological-cognitive, reflective and narrative aspects of self. Below, it is considered what is missing in Cotard patients' RSC in terms of the shades of self-reflection presented in Chapter 3. CS is approached first with the distinction between 1P and 3P thinking of oneself, then with the distinction between deliberative and theoretical stance, and then with the distinction between voluntary and involuntary self-reflection. Lastly, the considerations are put together by considering that normal mental well-being is characterized by flexibility in RSC, and this flexibility is disturbed in CS.

Within the most general distinction of RSC, the one between first-person and third-person stances toward oneself, CS clearly points out a case in which one is trapped in a 3P approach, unable to gain the 1P mode to herself. Patients refrain from typical first-person expression and instead engage in third-personal objectifying self-reflection that is evident in terms of hyper-reflectivity. A hyper-reflective CS patient observes herself as an object (in the same way as external objects), not as a subject who is an active agent and authentic 'I.' The same kind of point was raised by Ciaunica and colleagues (2021) describing a fracture between an observing and observed self in depersonalization; the self is present as the observed thing, but not as a typical feeling agent that enacts the observing. This fracture is also present in CS but in a more severe form in which the observed self too is considered nonexistent; the self is not felt as if it is unreal but as really lacking mental or bodily features or the status of being alive.

While CS patients face themselves only with 3P mode, they do not reach the uniqueness of self-conscious thoughts and cannot exercise their reflective self-consciousness at its full power.

However, a self-conscious thought, that is, a thought referring to oneself by the use of the 'I', has unique epistemic and motivational necessary features (Sec. 3.2.). Firstly, entertaining self-conscious thought (e.g., by saying "I feel happy"), the subject cannot misidentify the referent of her claim. However, CS patients do not make this kind of 1P reference, instead favoring 3P references to themselves ("Madam Zero"). Secondly, self-conscious thoughts are motivationally unique since they matter to subject's practical reasoning in a special way. Lacking self-conscious thoughts, a CS patient has lost this intimate relation to her action too; she does not enter into normal practical reasoning but rather feels that that kind of reasoning does not concern or relate to her anymore. In other words, the loss of I-thoughts is connected to ceasing to be an active agent; instead one remains an inanimate passive spectator lacking first-personal (mental and practical) activity, only observing herself from 3PP.

These dysfunctions in RSC are also evident in terms of deliberative and theoretical stance toward oneself (Moran, 2001; Sec. 3.3.). Within this distinction, CS involves over-emphasized theoretical stance and diminished deliberative stance. That is, even when one is using a first-person reference to herself, the self is observed in a rather theoretical manner ("I am this and this kind of person" or in the case of a CS patient "I am not a proper person at all") and not as an agent initiating actions ("I will change"). This excessive theoretical stance might also be called an inauthentic stance toward oneself (in reference to Radovic, 2017; Sec. 6.4.3.). Instead of being a "real" authentic deliberative self, a CS patient performs only a theoretical approach towards herself.

The lack of 1P point of view is connected to the inability to flexibly modulate and switch between different modes of self-reflection. In their study of depersonalization, Ciaunica and colleagues (2021) described this inability in terms of self-as-subject and self-as-object (i.e., minimal and reflective self-consciousness). I want to emphasize the role of modulation of self-consciousness and extend the point presented by Ciaunica and colleagues (2021): the inability of flexible modulation does not concern only the relation between MSC and RSC but also the relations in the structure of

RSC in CS patients. Normally, one can flexibly modify and revise the mode of her self-reference in accordance with demands of the situation. However, in CS this normal flexibility to exercise different modes of self-consciousness is lost or diminished. One cannot change between 1P and 3P approach, or theoretical and deliberative stances; the third-personal or theoretical stance has superseded the deliberative. This inability to modulate the modes of self-reflection can be considered also in terms of voluntariness of self-reflection. Normally, RSC is characterized by the capacity to voluntarily alternate stances toward oneself, but the 3P stance is involuntarily sustained in CS.

Altogether, CS involves a failure of first-person RSC: this failure manifests itself as the involuntary lack of first-personal cognition and the involuntary lack of alteration in the modes of self-reflection. This strengthens the distinction between the 1P and 3P modes of self-reference and the special status that 1P has. The 3P mode can remain without 1P, but this condition causes suffering to the CS patients. When a patient resigns first-person expression, she no longer can alternate between 1P and 3P stances but remains captured by the 3P approach to herself and loses the feeling of being an agent capable of carrying through the alternation. The case of CS also supports the idea that reflective 1P is based on experience; the lack of 1P thinking can be considered a rather natural outcome of the feelings of diminished MSC.

Further, the case of CS strengthens the significance of voluntariness as essential to normal RSC. This capacity to control and modify contents can be considered a central function of RSC, and CS patients present a case in which prevention of this capacity involves distress. The inability to voluntarily modify self-reflection is also connected to the inability of taking a 1P stance; these inabilities are overlapping and feeding each other. In other words, the very 1P approach to oneself is linked to the volitional capacities of a person. When one fails to exercise deliberative stance and the modification of her mental states, she at the same time fails to exercise her volitional skills. Or the other way around, when one is unable to voluntarily modify the stance she takes on her self-reflection, she cannot enter the deliberative stance. In a sense, the

inabilities to voluntarily modify self-reflection and to take 1P stance might be seen as an extension of the vicious circle between diminished self-affection and hyper-reflectivity to RSC. The patient cannot "correct" her third-personal objectifying thoughts to first-person although she tries to reach deliberative stance and the feelings of being a competent authentic agent, and this failure erodes her 1P even more.

The distressing unbalance between 1P and 3P in CS supports the general idea that psychological well-being involves the 1P stance and the alternation between different modes of reflection (presented in Chapter 3). Moran (2001; Sec. 3.3.) argues that the deliberative stance is crucial for mental well-being, together with a flexible and functional alteration between the deliberative and theoretical stances. CS presents a case in which a patient is in a constant inauthentic theoretical stance, unable to switch to a deliberative stance and employ its capacities in order to modify her mental states. Since CS causes distress, it highlights the deliberative stance and the flexible modulation between the modes of reflection as significant for functional RSC.

Furthermore, it can be remarked that conditions such as CS actually elicit the richness of our self-reflection. The human mind is unique because of the sophisticated RSC we have, but due to the complex structure of RSC, it is also vulnerable to various malfunctions. Since one can think about herself in various manners, she can also become alienated from her experiences in many ways. By comparison, creatures that lack sophisticated RSC cannot suffer from a condition like CS since they are incapable of complex self-identification, -reference and -deliberation in the first place. An infant does not reflect on her self but is absorbed in feeling and experiencing. Or a fox is a fox carrying through behavior that is characteristic of its species: the fox does not consider whether it is a fox or not, it just undergoes fox experiences and actions. Humans, however, can get alienated from themselves. The fox is a self-conscious creature in the minimal sense: it is immediately aware of its ongoing experience. However, without RSC, the fox does not consider itself as "active subject or thinker" and because of this original lack, it neither can lose or miss this capacity.

7.3. The structure of self-consciousness as an interconnected whole

In addition to the elaboration of the structure of MSC and RSC separately, the study of CS points out the dynamics of self-consciousness as a whole. Generally, the examination of CS fortifies the conclusions made in the earlier chapters about the importance of the interconnections between minimal and reflective self-consciousness. CS shows that the relations between MSC and RSC are crucial in two ways (as presented in Chapter 5): there is a necessary connection between the two and a dynamical interaction. First, the constitutive role that MSC has in self-consciousness is evident in CS: MSC is present while a patient exercises her RSC, and first-order experience grounds the contents and delusion in RSC. Second, the dynamical interaction can be seen in the connections between MSC and RSC that involve wide disturbances in CS. On the one hand, the disturbances occur as increased connectivity in the vicious circle between hyper-reflectivity and diminished self-affection. On the other hand, the disturbances are evident in decreased connections between MSC and RSC. Altogether, the malfunctions in all features of the dynamics of self-consciousness display a distorted pattern of self(-consciousness).

7.3.1. RSC is disturbed because of the disturbances on MSC

MSC has a fundamental role in the hierarchy of self-consciousness since it constitutes all forms of (self-)consciousness. By definition, MSC does not refer to any content of consciousness but to the manner in which a content is present. This implies that the study of RSC is not totally independent of the study of MSC, since the two—contents and subjectivity—are intertwined in experience. Although the referent of the content of consciousness is the same (one's self), experience of it is not the same if the related MSC has turned anomalous. In depersonalization, the content of consciousness is oneself, but there is something strange in that content because it is not present in the familiar way it used to be. With diminished embodied, experiential and affective features of MSC, the self feels unreal or distant. And in CS, reduced MSC and the feelings of not

being alive have turned into experiences and claims of being altogether non-existing.

Further, CS patients' altered MSC is taken as the content of RSC and drives self-reflection to deal with the experience. In CS, the delusional belief of being dead is based on an extremely depersonalized experience that is grounded in distorted MSC. Since CS originates in alterations in MSC, deficiencies of first-person expression can be considered a rather rational reaction to the feeling of diminished MSC (e.g., Billon, 2016, 2017) and the delusions resulting from efforts to conceptualize or reflectively grasp anomalous experience (Feyaerts et al., 2021). On the other hand, the distortions of RSC can sustain and reinforce the feelings of failing to be a proper self. When the belief in one's nonexistence is formed, it seems to alter one's total self-model or -narrative.

In brief, it is important to note that disturbances in MSC inflict disturbances on RSC. This is schematically illustrated in Figure 7.1, in which diminution of MSC triggers excessive RSC, which processes the anomalous experience in order to handle it, and together these disturbances form a vicious circle.
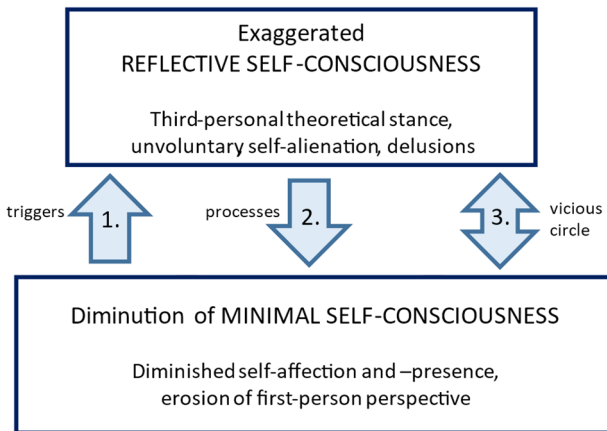

Figure 7.1. A simple model of the interplay between two forms of self-consciousness in Cotard syndrome.

Figure 7.1 depicts a general idea of the role of the two forms of self-consciousness and their interplay that is useful to make more fine-grained distinctions in explaining pathologies. For instance, Feyaerts and colleagues (2021) present a "phenomenological update" picture of the formation of delusions in schizophrenia. In this picture, an anomalous experience, including self-disorders, and delusional mood lead to a (primary) delusion by mediation of a process of conceptualization that takes place in an altered sense of reality. Comparing this picture with Figure 7.1, MSC includes delusional mood and altered sense of reality, which can be considered altered affective and experiential aspects of self. The simple picture of Figure 7.1 is elaborated below to include several aspects of self and their connections.

7.3.2. Cotard syndrome as a distorted pattern of self

In terms of pattern theory, CS can be considered a distorted pattern of self that is characterized by an extreme diminution of first-person. This diminution is manifested in both MSC and RSC and also in distorted connections between these two forms of self-consciousness. The pattern theory is useful in the discussion of CS since it gives a framework, in which several distorted aspects of self and their relations can be portrayed together.

CS involves distortions in the links between MSC and RSC in terms of hyper-reflectivity and increased wider connectivity, and lessons from these unbalances are significant for theories of self-consciousness. On the one hand, the vicious circle between diminished self-affection and hyper-reflectivity implies that features of minimal and reflective self-consciousness are deeply connected and form self-experience as complementary processes. MSC is present in RSC as its background and providing its contents, and RSC is used in modifying MSC and in initiating actions. In CS, the normal flexible modification in their relations breaks, and the vicious circle becomes dominant.

On the other hand, CS is characterized by an only partial activation of both MSC and RSC, and the normal connections between them are dramatically declined, which is also evident in decreased neural activation. This decline brings out the complexity

of connections that normally tie MSC and RSC together. Without the rich connections between them, something essential is missing in experience. Thus, in the same way as CS points out the significance of MSC in normal experience, it points out the connections between MSC and RSC to be crucial for typical balanced self-consciousness. These connections are tacit, and we do not pay attention to them in everyday experience, but they become obvious when one can no longer carry out the modification of them and is captured in a narrow and biased stance towards herself.

The distortions in the pattern of self-consciousness in CS are roughly illustrated in Figure 7.2. This figure can be compared with Figure 5.2, which illustrates a typical network of connections within self-consciousness. Instead of the typical extensive and dense network, in which one feature is connected to several other features, most features have only single connections and thus remain in isolation from each other in CS. This can be seen in MSC, in RSC and in the links between them. Further, instead of the typical wide and balanced interaction between MSC and RSC, the interaction is dominated by hyper-reflectivity in which exaggerated third-personal RSC scrutinizes diminished MSC. The third-personal hyper-reflectivity is illustrated in the figure with the thick circle. The striking features of the picture are that there are no other thick circles, but the third-personal self-reflection dominates the picture. Connections within reflective self-consciousness are significantly decreased from normal, and typical first-personal self-reflection, which could be considered the circle next to the thick one, is not strong. The radical diminution of MSC can been seen in that it does not involve thick circles, and the connections between the circles are scarce. In the connections between MSC and RSC, the dominance is on the connection between the circle of third-personal self-reflection and some particular diminished features of MSC, but typical extensive connections between them are missing. In addition, the radical diminution of self-consciousness is related to disturbed encounters with the world. Altogether, the Figure 7.2 illustrates imbalance and disintegration within self-consciousness; it shows diminution of first-person in several features of self and in the dynamics between the features.
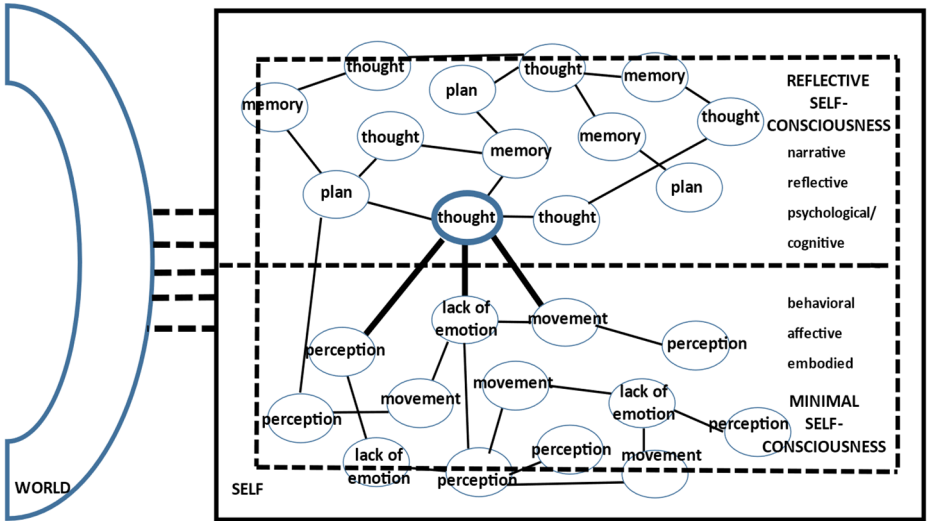
Figure 7.2. A distorted pattern of self-consciousness in Cotard syndrome.

The pattern theory of self is useful in the examination of CS, since it also assists in understanding the differences between CS cases. Cases with different groups of symptoms can be considered slightly differently altered patterns of self. For instance, the delusion of being dead results from a certain kind of alteration of a pattern of self but is not present in all cases of CS. That is, in addition to providing an explanatory framework in which alterations in different features of self-consciousness can be brought together, the pattern theory can also provide the means to see differences and similarities between particular cases of CS.

Altogether, the analysis of CS strengthens the conclusions drawn in previous chapters. A convincing interpretation of CS establishes the conception that MSC is the fundamental form of self-consciousness that constitutes the other forms; MSC is always present in RSC, and further, alterations in MSC inflict alterations on RSC. In self-reflection, minimal and reflective self-consciousness are amalgamated, and effects between them run in both directions. In addition, CS involves distortions in the connections between MSC and RSC, and these distortions show how the connections

play a remarkable role in the structures of the self-experience. These conclusions imply that the study of RSC should take notice of MSC and that the study of the unity of self-consciousness is needed in order to provide a complete picture of self-consciousness.

7.4. Theories of self-consciousness require the notion of minimal self-consciousness

It has been argued above that an explanation that invokes MSC manages to deal with CS best, and this insight from CS can help in sorting out theories of self-consciousness. To recap (Sec. 6.1.), a theory needs to be able to cover all its instances; if a theory of self-consciousness fails to embrace all cases in which self-consciousness is present, the theory is undermined. On the other hand, a theory is strengthened when it shows that it can encompass rare cases too.

Generally, the case of CS indicates that theories that appeal only to RSC or deny the substantial role of MSC are insufficient since the notion of MSC is needed in expounding on the syndrome. As described above, an explanation that invokes only RSC is not sufficient for understanding what happens to CS patients. Although CS clearly involves problems in RSC, it cannot be reduced to the failure of RSC only since reflectively self-conscious beliefs originate in altered first-order experiences. Thus, in order to understand patients' peculiar beliefs, their altered MSC requires attention.

Further, the failures of RSC in CS are related to the links between MSC and RSC and thus highlight the significance of these links. Instead of normal balanced interaction, in which MSC provides RSC with information on contents of consciousness that RSC regulates, CS involves the vicious circle between minimally self-conscious diminished self-affection and third-personal hyper-reflectivity in self-reflection. At the same time, CS involves a decline in typical more diverse interaction between MSC and RSC that supports the regulative functions of RSC. This distorted pattern points out the tight but flexible connection that typically ties MSC and RSC together and enables balanced self-consciousness. The

significance of this pattern implies that a theory of self-consciousness needs to acknowledge the reciprocal connections between MSC and RSC.

7.4.1. Anonymity theories undermined

Since the analysis of CS accentuates MSC, it also strengthens the arguments against anonymity theories. To recap, any theory that denies that MSC is part of what is consciously experienced is called an anonymity theory of phenomenal consciousness (considered in Sec. 5.2.1., Grünbaum & Zahavi, 2013; Zahavi, 2014). At first sight, CS and depersonalization might be described as anonymous experiences: not involving self or personal experience but instead a viewpoint of "Madam Zero." This kind of interpretation has been made, for instance, by Metzinger (2004) and Kriegel and Billon (2015). However, under closer scrutiny, these pathological experiences only elucidate the existence of MSC. The experience of Madam Zero might be called 'anonymous' in the sense that she lacks the normal feelings of being 'I,' but it is not anonymous in the sense of anonymity theories that deny MSC altogether. In other words, the experience of a CS patient is not fully anonymous or non-personal but rather the experience of someone who feels she lacks proper personhood. In the end, CS highlights the significance of subjectivity and first-personal character of typical experience.

The experiences of CS patients can be considered a further critique of anonymity theories along the argumentation line presented by Grünbaum and Zahavi (2013). This argument points out that anonymity theories fail to account for a person's ability to refer to herself with a 1P concept and the unique features of self-conscious thoughts, and this undermines anonymity theories as theories of (reflective) self-consciousness. The unique features of self-conscious thoughts are both epistemic and motivational; a subject of an 'I'-thought necessarily knows that she is referring to herself, and this knowledge involves significance in her practical reasoning. According to Grünbaum and Zahavi (2013), there are at least two versions of anonymity theories that try to account for these features. One version appeals to the perspectival character of perception and the other to general rules of reference. In addition

to the arguments Grünbaum and Zahavi (2013) present, the existence of experiences of CS patients seems to indicate that neither version of anonymity theory suffices to account for the way in which RSC manifests itself in practice.

According to the first version of anonymity theory, the perspectival character of perceptual content fixes the self-specifying knowledge, and a self meets a definite description: "the perceiver of this perceptual presentation." However, CS shows that perceptual content alone cannot provide self-knowledge of being an 'I.' A CS patient does have a perspective of the world, but she disclaims being a proper self or 'I.' This implies that a mere perspective does not guarantee or fix the use of a first-person pronoun or deliberative stance. Instead, typically the perception is accompanied by MSC, and this subjective perspective constitutes the first-personal thoughts. CS patients suffer from diminished MSC and thus do not enter the deliberative stance, although they remain perceivers.

The other version of anonymity theories appeals to general rules of reference, claiming that the reference of 'I' is determined by a simple rule: a token of the 'I' refers to the creature that produces it. However, Grünbaum and Zahavi (2013) argue that this version presents a dilemma. On the first horn of the dilemma, the rule is considered to be grasped consciously, and 1P self-reference is explained by this explicit grasp. However, CS presents a case in which even this grasping does not entail 1P self-reference. A CS patient can acknowledge this kind of rule but still disclaim referring to herself as 'I', feeling more entitled to refer to herself with third-person concepts. This implies that mere explicit knowledge of the rule does not provide first-person thoughts, but the use of first-person expression is grounded in authentic MSC.

On the second horn, the rule is understood as an implicit law-like generalization that describes the functional inner structure of the system. Thus, the motivational difference between the forms of self-reference is explained in functional terms only; the system follows the rule when it categorizes information with direct relevance to the system into a specific information file. Grünbaum and Zahavi (2013) point out that merely operating this way does

not entail understanding that the system is referring to itself. Further, CS indicates that this kind of implicit rule is insufficient to account for the first-person thoughts. One can quit using first-person expressions although she had earlier followed the rule. The best explanation seems to be that if there was this kind of rule, it alone does not fix taking a first-person stance; typical first-person thoughts require immersive MSC.

These considerations also recap the point that the third-personal definite description is insufficient to account for first-person features. Altogether, CS reinforces the conception of MSC as the fundamental form of self-consciousness; the subjective character is needed in order to ground and account for reflective forms of self-consciousness. This subjective character of experience cannot be covered only by perspectival features of experience or rules in reflective language.

7.5. Considerations for empirical studies

By this point, the focus of this chapter has been mostly on the question of what philosophers can learn from pathological cases that have been studied in empirical sciences. However, neurophilosophy is a two-way endeavor; the conceptual and theoretical work by philosophers can reciprocally inform empirical research and psychiatry about CS. Fine-grained concepts and analysis are needed in order to understand the alterations in patients' experiential, psychological and neurophysiological processes.

Thus, in addition to implications to theories of self-consciousness, the analysis of CS can contribute to the empirical research on self by indicating the complex structure of self. In terms of the neural profile of self, the analysis of CS supports the general picture that self-consciousness involves activation within several neural networks and that the connections between the networks play a significant role. Generally, MSC involves sensorimotor neural processing, and RSC involves self-related processing that activates a wide E-network (Sec. 4.1.1; Christoff et al., 2011; Legrand & Ruby, 2009). CS is characterized by various neural

dysfunctions in all these processes. The dysfunctions are briefly discussed below, relative to future neural research of CS and self in general.

Considering the widespread neural abnormalities in CS, it would be beneficial to elaborate the neural dysfunctions in the connections between MSC and RSC. Normally, self-consciousness is characterized by interplay between minimal and reflective self-consciousness. This connectivity breaks down in CS, which involves extensive neural hypoactivity (see Secs. 6.5.1. and 6.6.2.). In more detail, it seems that the typical balanced and wide connectivity between neural networks underlying MSC and RSC is replaced by the quite frontal vicious circle, involving hyperactivity in, for instance, the medial and inferior frontal cortex. These unbalances have a central role in CS, and thus the connectivity provides a target for further studies.

In order to track the altered pattern of self in CS, it is important to look at extensive changes in neural functions. For instance, Gerrans's (2015) model of CS and depersonalization emphasizes the neural dysfunctions of AIC and VLPFC especially; encoding of affective significance of body states is decreased in hypoactive AIC since VLPFC hyperactively regulates this processing of affective feelings. However, in addition to this abnormal activation of AIC and VLPFC, depersonalization and CS involve more complex problems in neural processing. Empirical studies of depersonalization have found alterations in, for instance, ACC, orbital gyrus (Lemche et al., 2013; Medford et al., 2016), and the vestibular system (Jáuregui Renaud, 2015). And CS also involves wide hypoactivity in the frontal and parietal cortices (Charland-Verville et al., 2013; Swamy et al., 2007). Thus, although AIC plays an important role in processing bodily and affective features of self, the decrease in activation of AIC only is insufficient to account for the emergence of alterations in the experience of the CS patients. In addition to the information about this kind of local dysfunction, it would be beneficial to track the changes in relations between neural processes. Generally, the decreased connections between MSC and RSC could be seen as decreased activity between the frontal-temporal and parietal areas, but it would be interesting to

also develop more specific hypotheses about the alterations in the connections.

One main point in this chapter is that CS origins in the diminution of MSC and the central role that MSC has in experience make it a relevant target for empirical studies too. It has been proposed that the sensorimotor neural processing underlying MSC involves parietal neural activation especially, which implies that it would be worth a closer analysis to look at the functioning of parietal lobes in depersonalization and CS patients. This could lead to empirical hypotheses about the dysfunctions of sensorimotor processing that are characteristic for depersonalization and CS. A simple hypothesis about the dysfunctions would focus upon a general decrease in parietal processing as characteristics of CS. This kind of hypothesis seems already to be affirmed by empirical data that show decreased parietal activity in patients (see, e.g., Swamy et al., 2007). More detailed hypotheses could concern various sensorimotor processes more specifically. For instance, it might be suggested that patients have problems in their primary somatosensory and/or motor cortex or that dysfunctions occur especially in reafferent processing. I do not elaborate more detailed hypotheses here but want to highlight the advantage of noticing the sensorimotor neural processing that underlies MSC in conducting empirical research on self. At the same time, knowledge of the dysfunctions of these processes in CS and depersonalization could aid in understanding the sensorimotor basis of MSC in non-pathological conditions.

Concerning RSC, the analysis of CS supported the distinction between 1P and 3P modes of self-reference. CS presents a case in which a patient can lack 1P mode to herself while 3P mode to herself still functions, and this implies that the processes underlying these modes are at least partly distinct. From an empirical point of view, this kind of dissociation in experience suggests that 1P and 3P stances on self involve distinct neural processes. Thus, in future research, it would be useful to elaborate the neural basis of RSC and trace the specific processes involved in 1P and 3P stances. For instance, Frewen and colleagues (2020; Secs. 4.1.1. and 4.2.2.) propose a quadripartite model of neuroanatomy of MPFC related

to different functions of self-related processing. Especially interesting here seems to be the difference between objective and executive selves that could be linked to a more third-personal theoretical stance toward oneself and a more strongly first-personal deliberative stance toward oneself. According to Frewen and colleagues (2020), both modes of self are related to top-down processing in DMPFC, but the processing of the executive self involves more superior aspects of DMPFC that are part of the central executive network too. In terms of these features of self-reflection, the analysis of CS seems to produce an empirical hypothesis that the objective self would be hyper-reflectively activated, whereas the executive self would have decreased activation. The existing studies seem to support at least the former point by showing hyperactivation in the medial prefrontal areas of CS patients, but more specific areas within the MPFC are not reported. In any case, in order to understand better the neural basis of the subtleties of RSC and their breakdown, it would be useful to specify the functions of different parts in the frontal lobe. And in addition to these rather local differences, connections in RSC could be studied by targeting altered connectivity within frontal and temporal lobes.

Further, the neural profiles of particular cases of CS vary, and it would be beneficial to clarify the neural dysfunctions in different versions of CS. The data concerning more common mental disorders show that schizophrenia involves a decrease in the connections between frontal and parietal lobes whereas depression involves an increase in these connections (Zhao et al., 2013). Related to these, CS would at first sight more closely resemble schizophrenia, and probably the "mixed" version of CS could have some similarities with the neural profile of schizophrenia. According to Huarcaya-Victoria and colleagues (2020) the mixed version involves depression, anxiety and auditory hallucinations, and resembles the Cotard type 2 described by Berrios and Luque (1995b). On the other hand, CS is most often linked to depression, and it would be interesting to understand better the progression of CS: how the increased neural activity of depressions drops to decreased activation of CS. Here, for instance, the suggestion of the

longitudinal progression of CS by Yamada and colleagues (2007) could function as a working hypothesis. They proposed that CS proceeds from the germination stage of depression and hypochondriacal symptoms (psychotic depression) to a blooming stage of delusions of negation together with other symptoms (Cotard type 2) and further to the chronic stage of delusions and depression (Cotard type 1).

Altogether, in order to understand CS, various explanatory factors need to be taken into account, and all these factors can also be seen in the neural dysfunctions of CS patients. Thus, a clear conceptual picture of CS can assist in clarifying different neural failures. In future studies, it would be interesting to compare the neural profile of CS with other pathologies (including depersonalization, depression and schizophrenia) and altered states of consciousness. These kinds of comparisons could aid in understanding the cognitive and neural mechanisms involved in the particular altered states on the one hand and in typical self-consciousness on the other hand.

7.5.1. Psychiatry of Cotard syndrome

Generally, the multidimensional approach to self invites an extensive psychiatric perspective. According to the pattern theory of self (Daly & Gallagher, 2019; Gallagher & Daly, 2018), all psychiatric disorders are disorders in self-patterns (that is, self-disorders in a wide sense). The alterations in patterns and basic sense of self in mental disorders imply that psychiatric inspection would benefit from sensitivity to the experience. This has been argued especially in phenomenologically informed psychiatry (see Sec. 6.1.1.). To my knowledge, this kind of phenomenologically informed approach has not been used in proposing any direct interventions for CS so far. However, as we have seen in this and the previous chapter, the phenomenologically informed approach is beneficial in providing a fine-grained analysis of CS. Since the approach validly describes the alterations in the structure of self, it is motivated to apply the approach also in wider psychiatric considerations of CS. Below, I conduct this application in order to outline some relevant points for the psychiatry of CS.

7.5.1.1. Diagnostics

Since CS is not classified in current diagnostic manuals, its diagnostics and a picture of its progression are not fully developed. Generally, CS manifests itself as a specific kind of altered self-pattern, and knowledge about these alterations can have also diagnostic value. Especially since CS originates in distortions of MSC, some of these distortions could be recognized before the outbreak of delusions. At the same time, it is good to keep in mind that CS is not simply the delusion of being dead but has a more complex character. CS involves different nihilistic delusions (from desomatization delusions to time-oriented delusions) and other symptoms (anxiety, depression, depersonalization), and there are many versions of the syndrome (see, e.g., Berrios & Luque, 1995b; Huarcaya-Victoria et al., 2020).

One focal feature of CS that the examination above highlighted is the close relation between CS and depersonalization. Because of the central role of this relation, it seems reasonable to suggest that depersonalization could be used as a criterion in developing the diagnostics for CS.[86] Indeed, CS has been considered a more severe version of depersonalization (see Sec. 6.5.3.). Thus, paying attention to depersonalization symptoms can help to recognize potential CS patients before the outbreak of delusions, and at least some CS cases might be prevented.

However, since CS is very rare and depersonalization is present in many altered states of consciousness, a more elaborated picture of its use as a possible diagnostic criterion for CS would be needed. An experience of depersonalization alone does not imply CS, but the depersonalization in question is extremely intense. The considerations in this chapter do not point out any single feature of depersonalized experience to be decisive for the transition to CS, but suggests generally that, for instance, diminished self-affection, feelings of unreality of self and world, and changes in the way of referring to oneself can precede delusions. These alterations could be investigated through precise examinations and interviews, such

---

[86] Sass et al. (2013) have proposed that depersonalization could be used in diagnosing schizophrenia too.

as the Cambridge Depersonalization Scale (CDS, Sierra et al., 2005) and Examination of Anomalous Self Experience (EASE, Parnas et al., 2005). In future studies, it might turn out that particular depersonalization symptoms are connected to the onset of CS, and/or the prolonged duration of depersonalization symptoms might be crucial in the development of more severe mental disorders.

In general, since CS involves a specific altered pattern of self-consciousness, many kinds of more accurate information about the pattern can be diagnostically significant. The link with depersonalization seems especially valuable since it concerns several aspects of self at the same time, including diminished affectivity in affective aspects, eroded first-person perspective in experiential aspects, and changes in self-conscious thoughts in reflective aspects. In addition, some kind of combination of personality traits or a form of depression or schizophrenia might be significant in the progression of CS. Further, better knowledge of the changes in neural mechanisms in CS could be used in diagnostics alongside the phenomenal character of the patients.

7.5.1.2. Treatments

Better understanding of the intricate dynamics of self can help also in the development of treatments for CS and depersonalization. At least in some cases, medication has not had the hoped-for effect on CS, and medication that works for one patient might not help another. As well, DPD still lacks treatments that would work for all patients (see, e.g., Ciaunica et al. 2021). Thus, more effective treatments and knowledge about the selection of treatments are greatly needed.[87] In finding the right kind of treatment, it is important to acknowledge that there are different types of CS, and the different types probably benefit from different treatments. For instance, Huarcaya-Victoria and colleagues (2020) differentiate

---

[87] However, all the possible treatments of CS cannot be considered here, but the focus is more on what new information the analysis in this chapter could offer for the determination of treatments. For treatments of particular cases see, e.g.,. Debruyne et al. 2009; Huarcaya-Victoria et al., 2020; Ramirez-Bermudez et al., 2010.

between three types of CS cases (psychotic depression, delusions/hallucinations and mixed), and argue that taking the differences between the types into consideration optimizes treatment and treatment response.[88]

The multidimensional approach to self generally proposes a wide selection of non-medical treatments for mental disorders. For instance, Daly and Gallagher (2019, 13) argue that the pattern theory of self "motivates a diversifying of treatments on offer to also include more personal and lifestyle investments such as CBT, narrative therapy, meditation, Qi Gong, art therapy, various movement therapies, etc." Further, Feyaerts and colleagues (2021) propose clinical-phenomenological improvements to current (cognitive-behavioral) therapies. Feyaerts and colleagues (2021, 9) argue that the focus of treating delusions should be shifted "away from narrowly targeting delusions themselves (via attempts to for example refute or challenge them), towards altering the experiential conditions that inspire and sustain them. Effective treatments will help to reduce the feelings of self-alienation and uncertain embeddedness in everyday reality that are conducive to delusional experience." These kinds of ideas seem to be relevant to apply to CS, since delusional CS seems to fit the description of Feyaerts and colleagues (2021). In addition, the alternative therapies mentioned by Daly and Gallagher (2019) aim to induce extensive changes in self-pattern, and it is worthwhile to investigate whether some of them can dissolve the pathological alterations of CS. Below, I briefly consider therapies that focus on embodied, social, and meditation techniques.

Since CS originates in distorted MSC, it seems that patients could benefit from interventions that acknowledge and target MSC more specifically. This idea is compatible with Gerrans's (2015, 11) claims that "loss of affective response is not something that can be restored from the top down." If the pathologically altered MSC in

---

[88] For instance, patients with the delusion/hallucination type of CS are not showing a suitable response to antidepressant treatment, and antipsychotic treatment is preferred in this case (Huarcaya-Victoria et al., 2020; see also, e.g., Berrios & Luque, 1995b; Huarcaya-Victoria et al., 2016).

CS and depersonalization cannot be fixed by RSC, it seems intuitive to aim to affect MSC first in the treatment. Especially, beneficial treatment could induce a stronger sense of embodiment and reduce anxious hyper-reflective self-awareness. This kind of recommendation has been suggested for treatment of depersonalization (Ciaunica et al., 2021), schizophrenic delusions (Feyaerts et al., 2021) and both (Sass et al., 2013). In order to counterbalance hyper-reflectivity and self-objectifying stance, different body-based therapies could be useful. As possible relevant body-based therapies, Feyaerts and colleagues (2021) mention dance and music therapies and physical activity that can be immersive enough to generate flow states.

In addition to the focus on embodiment, the role of social relations could help overcome the excessive self-focus. Feyaerts and colleagues (2021) point out the role of detachment in the development and maintenance of delusions and that treatment should also target the feelings of loneliness, disconnection and isolation that are often experienced in psychotic conditions. Generally, the importance of social components in treatment supports the significance of therapy and highlights the role of a competent therapist (Daly & Gallagher, 2019; Feyaerts et al., 2021). Further, for depersonalization, Ciaunica and colleagues (2021) propose that "moving with others" could dissolve the feelings of being detached from the world (see also Ciaunica & Fotopoulou, 2017). Ciaunica and colleagues (2021, 14) note that their recommendations of embodied and social activity as treatments for DPD are supported by self-reports of DPD patients "who often notice an improvement of their dissociation states after a period of intense physical training and/or social interactions."

Further, a potential technique for treating CS could shift the emphasis away from hyper-reflective contents of consciousness towards patients' overall experiential orientation. Here Feyaerts and colleagues (2021) mention the third-wave cognitive-behavioral therapies (Hayes & Hofmann, 2017), which include acceptance and commitment therapy and mindfulness-based therapies. Mindfulness meditation could be useful since it offers a way to approach the diminished self-affection with acceptance or an

approving attitude. With this kind of attitude, one can view her own mental states but let go of the related anxiety. Although self or the world would feel unreal or nonexistent, the feeling does not determine one's self; it can be accepted among the flow of other feelings. In other words, even if one cannot change her minimally self-conscious feelings by reflection, she might be able to modify her attitude towards these feelings and thus reduce the anxiety. Ciaunica and colleagues (2021) also raise meditation-based interventions as useful treatments for DPD, especially for their potential to regulate the anxiety in depersonalization. However, Ciaunica and colleagues (2021) consider the lack of embodied engagement component to be a shortcoming of meditation-based techniques as treatments of DPD. (More about the relation of meditation and depersonalization in the next chapter.)

These potential options for treatment are interesting, but it is also important to acknowledge the special characteristic of CS and carry out more research in order to find apt therapies. Even if the mentioned treatments can be beneficial for depersonalization, the condition of CS is more severe and it is not obvious how patients would react to the treatments. Since the problem of CS is, in the first place, the lack of emotional involvement and nihilistic delusions cornering one's body and mind, attending to these features might also increase anxiety. In addition, while claiming to already be dead and lacking proper selfhood, patients can lack the motivation to engage in new activities. In any case, the mentioned options of treatment —body-based therapies, social interaction and meditative practice— targeting the overall experiential orientation are worth more study. By acknowledging the complex distortions of self in CS, these treatments could be advantageous in helping patients reconnect to themselves and the world.

7.6. Summary

The analysis of CS in this chapter pointed out three factors. First, CS manifests a diminution of MSC: the lack of immersion in experience can be tracked to disrupted affectivity and erosion in the structure of 1PP. Erosion in the structure of 1PP also includes features that

involve body, space and time. MSC does not disappear but is severely diminished, which makes the self feel inanimate and the world without practical relevance. Second, CS involves distortions in RSC. Instead of first-person, a patient approaches herself with a third-personal or theoretical stance, even using third-personal ways to refer to herself. The patient excessively reflects herself in an objectifying and involuntarily manner. The patient cannot switch between different modes of thinking of herself but is caught in the third-person mode, which only feeds the feelings of self-alienation and inauthenticity. Third, the connections between MSC and RSC distort in CS. The vicious circle between diminished self-affection and hyper-reflectivity dominates self-consciousness, while typical, more diverse interaction between MSC and RSC is absent. Together these three factors display a distorted pattern of self that is characterized by only partly activated features of self and their connections.

Altogether, the investigation of the alterations in the structure of self-consciousness in CS gives us the possibility to understand better what the patients are undergoing and to clarify the concepts of self. Lessons from CS consolidate both the conception of MSC as the most fundamental form of self-consciousness and the reading that emphasizes its subjective character. Further, the analysis of CS highlights the importance of the modes of RSC and the connections between MSC and RSC. In addition, it strengthens the idea that mental well-being involves a balance between different aspects and in their relations, whereas psychopathologies involve an unbalanced pattern of self. This kind of detailed conceptual picture of CS is also useful for empirical and clinical studies, and invites an extensive approach to mental disorders, that acknowledge MSC and the dynamics within self-consciousness. In the end, theoretical insights of self can assist in finding ways to help the patients in their struggle with anomalous experiences.

8. Comparison between depersonalization and meditation indicating shades of self-consciousness

This chapter deepens the account of the significance of dynamics of self-consciousness. As we saw in the previous chapter, CS and depersonalization are characterized by feelings of estrangement from one's self, and patients feel strong distress about these feelings. However, in a traumatic and stressful event, a milder form of depersonalization can be seen as a rational reaction, which helps one to overcome and to get through the situation (e.g., Hunter et al., 2003; Stein & Simeon, 2009). Further, feelings of some kind of depersonalization are involved in experiences that are considered positive ones. Meditation is this kind of altered state of consciousness that aims to detach from normal operations of mind, including highly affective contents of self-consciousness, in order to reach a state of mind in which a meditator does not identify with the contents of RSC. Meditators describe this self-detachment as a pleasant experience, and often meditation is practiced because it is considered to promote many positive effects, such as the reduction of stress, improved cognitive performance and mental well-being (e.g., Dambrun, 2016; Fox et al., 2016; Gu et al., 2015; Tang et al., 2015).[89] Because of the similarities in experiences of non-identification, it has been suggested that meditative experience resembles depersonalization, however, the suggestion has not been studied more closely (Castillo, 1990; Ciaunica et al., 2021; Metzinger, 2004).

In this chapter, I conduct a comparison between meditation and depersonalization since this comparison can elicit subtleties of self-consciousness. First, I give a general description of meditation and point out how it is closely related to the notions of MSC and

---

[89] However, as also Ciaunica et al. (2021, 14) point out, it would be misleading to conclude that meditation necessarily leads to mental well-being (see, e.g., Lindahl & Britton, 2019). Instead, the self-exploratory practice of meditation can raise some hidden distressing feelings or traumatic memories, and "trigger debilitating impairments with potentially irreversible consequences." However, for simplicity, this chapter focuses on "positive" experiences of meditation that are connected to well-being.

RSC. Second, I look at the similarities between meditation and depersonalization, finding detachment or non-identification as their shared factor. Third, I examine the differences between the two states. Notably, depersonalization involves dysfunctions in MSC and RSC and their connections, whereas meditation is described involving enhanced MSC and flexible RSC, resulting in extensive connections between the two. Finally, I briefly outline the different patterns of self that are involved in these altered states of consciousness and consider how they point out the connective links in self-consciousness.

8.1. Meditation as an altered state of consciousness

Generally, meditation can be defined "as a form of mental training that aims to improve an individual's core psychological capacities, such as attentional and emotional self-regulation" (Tang et al., 2015, 213). Different forms of meditation are found in almost all cultures. However, nowadays, 'meditation' is quite commonly used to refer to practices that have been developed and exercised in the East for centuries or millennia and that typically derive their origin from elaborated religious or philosophical ideas.[90] Just recently, scientific research has become interested in these practices and their positive effects, including improving attention and health (see, e.g., Gu et al., 2015; Tang et al., 2015). Because of the great number of different trends and schools, details of different meditation techniques cannot be wholly covered in this dissertation, but, for the sake of clarity and brevity, only some main ideas of meditation and their relation to self-consciousness are discussed. In addition, the analysis here aims at a comparison of shades of self-consciousness between meditation and depersonalization, focusing on main ideas of meditation that are relevant especially for this comparison. The examples in this chapter are highly useful and capable of showing

---

[90] Within western philosophy, meditation might be connected to methodological skepticism or epoche. See also fn. 93.

dynamics of self-consciousness even without an exhaustive description of the meditative states.[91]

At first, it is important to be clear about positioning meditation as an altered state of consciousness. Meditation is typically defined as a technique used to train one's mind or attention. With meditative training, one can bring about or arrive at altered states of consciousness. In other words, a distinction can be drawn between meditation as 1) a technique of mental training and 2) a meditative altered state of consciousness that is the result of such mental training (e.g., Revonsuo, 2010; Tang et al., 2015). The latter is many times also the aim of meditative practice and is referred to as a "higher" state of consciousness. The detailed character of these altered states of consciousness depends at least on the used technique and on the previous experience of a meditator. Typically, it is rather difficult to get into an altered, let alone "higher," state of consciousness as a beginner to meditative practice. Only after training can one manage to enter an altered meditative state and gradually get into more altered ("higher and higher") states.[92] Since meditation is considered briefly and approached from a rather wide perspective in this dissertation, the focus is mostly on meditative states as altered states of consciousness. Further, the meditative state provides a comparable contrast with the pathologically altered states of consciousness that are present in depersonalization. Indeed, it would be strange to

---

[91] However, since the theme of meditation is broad, this chapter involves some simplifications. Meditation and its relation to self have been conceptualized in different ways in different theories. Of course, it would be highly interesting to study in more detail how different concepts of self and styles of meditation are related to notions of self-consciousness used here. However, this short dissertation cannot cover that elaboration. For interesting studies of meditation within neurophilosophy and modern mind sciences, see, e.g., Austin, 1999; Flanagan, 2011; Metzinger, 2020; Schmidt & Walach, 2014; Thompson, 2014; Wallace, 2007. For an introduction to conceptions of self in different meditative traditions, see, e.g., Siderits et al., 2010.

[92] Indeed, the experience of meditative practice is dissimilar for meditation beginners and expert meditators, and empirical research has also shown different neural activations between these groups (see Sec. 8.2.4., e.g., Tang et al., 2015).

approach pathological depersonalization[93] as a technique; pathological depersonalization is not a form of mental training that one can choose to exercise; patients suffer from the condition and remain in depersonalized states involuntarily. That is, the comparison in the sense of technique would not function the same way in this chapter.

Meditation is highly relevant to the study of self because meditation-induced alterations of mind appear to concern the very manifestations of different forms of self-consciousness. In a meditative state, an excessive evaluative RSC withdraws, making way for a centered and calm MSC (in more detail in Sec. 8.2.). Further, many philosophical traditions in the background of meditation explicitly concern 'self' or 'ego,' seeing that our ordinary sense of self causes suffering and psychological distress (see, e.g., (Siderits et al., 2010). For instance, Buddhist philosophy, which is the background of modern mindfulness meditation, aims at disidentification from a static concept of self. This disidentification results in the freedom to experience a more genuine way of being (see, e.g., Hölzel et al., 2011; Tang et al., 2015; Vago & Silbersweig, 2012). In other words, according to Buddhist philosophy, meditation assists in grasping the illusive character of the everyday self and can lead to eradication of all identification and dissolution of self. By contrast, for instance, yoga philosophy sees that the aim of meditation is to find one's "true self" or "the self in its purity," which lies at the background of consciousness, hidden by typical self-related contents of consciousness (see, e.g., Radhakrishnan & Moore, 1957). Thus, it is interesting that although the character of self and the ultimate aim of meditation is conceptualized differently in different traditions, the traditions agree that meditative practice

---

[93] In a sense, depersonalization *can* be considered a technique to overcome a highly traumatic situation. In this sense, depersonalization has an evolutionary function in life-threatening situations, when it is necessary to detach from the situation and act "as if it is not really happening," to not get frozen by the fear and panic. People undergoing catastrophic situations often feel momentary derealization and depersonalization, and in these situations these feelings aid in immediate survival. However, that kind of "acute" depersonalization is not analyzed here; the target depersonalization is pathological experience in DPD; see Sec. 6.3.

is beneficial since it reconstructs and reorganizes self-consciousness. That is, the purpose of meditative mental training is to realize something profound about one's self-consciousness and to let go of one's self in a fixed everyday sense.

8.1.1. Meditative state as the self-presence of experiencing itself

Wolfgang Fasching (2008) offers a general characterization of meditation that is directly linked to notions of self-consciousness. In addition, Fasching's description is relevant here since it gives a generic definition that covers many forms of meditation. Further, Fasching refers to phenomenological literature, thus sharing some ideas with Zahavi (e.g., 2005, 2014), whose conception of MSC has been central in this dissertation (see Sec. 2.2.1.).[94]

Encompassing a large number of meditation studies, Fasching (2008, 464) describes a distinction that is drawn between two meanings of 'self-consciousness.' These two meanings are:

> 1) the self-identification with certain configurations of what one experiences and
>
> 2) the self-presence of experiencing itself.

According to Fasching, meditation aims at a temporary inhibition of the former to uncover the latter, which is more fundamental and constitutes the very being as subjectivity. Thus, meditation is "the endeavor to withdraw from one's being occupied with the objects of consciousness in order to become aware of consciousness itself" (Fasching 2008, 469).

The distinction bears a strong resemblance to the distinction between minimal and reflective self-consciousness used in this dissertation. At least the distinctions are similar enough to be paralleled.[95] MSC was broadly defined as being the subject or

---

[94] The resemblances between the ideas of the structure of mind in phenomenological and meditation literature can be seen, for instance, in the characterizing of mental acts. Fasching (2008) explicitly refers, e.g., to Husserl, 1952; Sartre, 1956; Zahavi, 1999, 2005.

[95] As explained in Sec. 1.1.5., the distinction between MSC and RSC is the most general distinction within self-consciousness and holds up even if exact definitions of the two forms

subjectivity of consciousness (Sec. 2.2.1.), which seems to correspond to Fasching's notion of "the self-presence of experiencing itself." RSC was broadly defined as the capacity of taking oneself as the object of her cognition and to think of herself as herself (Sec. 3.1.); the "self-identification" in Fasching's definition seems clearly to be part of this kind of contemplating of oneself. In meditation, one refrains from everyday evaluative RSC and, instead, focuses on the simple MSC without active identification.

Fasching (2008, 464) sees that different forms of meditation[96] have something fundamental in common: "They all aim at a stilling of the mind, which ... implies a ... withdrawal from intentional activity dealing with objects. ... In meditation one attempts to inhibit the usual way of being actively occupied with the contents, of dealing with them." Although some forms of meditation practice involve concentrating upon an object, the point is not to examine the object. Rather, the meditative concentration is a way to fixate the mind in order to abstain from being carried away by thoughts and feelings that are aroused in one's mind and tempt her interest. At the same time, some forms of meditation are not inhibitory in nature at all but, instead, involve choiceless awareness of whatever comes to the meditator's mind. According to Fasching, this kind of being aware can also be considered a form of withdrawal from everyday mental activity since the meditator is in a mode of mere nonreactive noticing. The meditator does not let herself engage in any mental "occupation" with contents that arise in her mind but refuses to react to it cognitively or practically.

In terms of self-consciousness, Fasching (2008, 478) proposes that the "essence of the meditative process of becoming self-aware

---

vary in their details slightly. An elaborated analysis of alterations in both forms of self-consciousness in meditation will follow in this chapter.

[96] According to Fasching (2008, 464), meditation is a technique of "mental restriction which aims at a state of utmost stillness of the mind in which all mental activities ... are brought to a halt." Fasching refers, e.g., to Patañjali's definition of yoga as "nirodha": as a "bringing to a standstill" the activities of the mind (Yoga Sutra, I.2, Radhakrishnan & Moore, 1957, 454). In more detail, with 'meditation' Fasching (2008, 463) refers "exclusively to what David Fontana calls "nonideational meditation," in contrast to "ideational meditation," the latter meaning "that the meditator holds an idea or a group of ideas in the forefront of awareness, and uses them to stimulate a directed course of intellectual activity" (Fontana, 2007, 154).

is … a de-identification from what we normally ascribe to ourselves." Instead of investigations of (self-related) contents of consciousness, the meditator becomes conscious of consciousness itself. Usually the consciousness itself remains "hidden" behind the contents of which one is conscious. However, meditative self-realization does not mean to introspectively evaluate oneself but to "forget oneself" in the sense of stopping opposing objects to oneself. Fasching Fasching (2008, 479-80) emphasizes again that meditation brings mental activities to a halt. Normally, every mental "act" has the structure of "I direct myself to … ," but in meditation one withdraws from this activity and abides in a kind of mental stillness. Below, I very briefly present two different forms of meditation that can further illustrate the ideas of meditation, and elaborate how they are interesting for self-consciousness studies (in Sec. 8.2.).

8.1.1.1. "Pure consciousness" meditation

Some forms of meditation explicitly aim at a state of "pure consciousness" or "pure subjectivity."[97] This kind of state clearly coheres with Fasching's description of meditation as inhibition of everything but being conscious. Forman (1998, 185-186) describes this kind of state:

> One neither thinks nor perceives any mental or sensory content. Yet, despite this suspension of content, one emerges from such events confident that one had remained awake inside, fully conscious. … The pure consciousness event may be defined as a wakeful but contentless (nonintentional) consciousness.

---

[97] Metzinger (2020) regards this pure consciousness as "minimal phenomenal selfhood" and gives an interesting neurophilosophical analysis for it. Buddhist tradition has called pure consciousness "witness-consciousness"; see, e.g., Albahari, 2006, 2009; Fasching, 2010. Further, it can be noted that some theories claim that pure consciousness is *not* a form of self-consciousness at all, but selfless experience; see, e.g., Albahari (2006, 2009, 2010) for this kind of no-self theory. However, according to the concepts used in this dissertation, consciousness always involves MSC, i.e., pure consciousness is the minimum of consciousness and self-consciousness at the same time; see Sec. 2.2.1.; Zahavi, 2010.

That is, a meditative state is a state of simply being wide-awake, without the content of consciousness, or even spatiotemporal self-location (Fasching, 2008; Metzinger, 2020; see also, e.g., Sekida, 1985). Hence, meditation can be seen as a mental technique of "isolating" and becoming aware of consciousness as such (see, e.g., Fasching, 2008; Metzinger, 2020). Yet, as Fasching (2008) notices, this does not mean an introspective observation of mental processes; a meditator is simply aware of being conscious at the very moment, or experiences herself "as this very moment of consciousness" (Fasching, 2008, 465).

Fasching (2008) highlights that the task of meditative self-realization is not to gain something new. Instead, meditation removes the veil (of contents of consciousness) showing what is always already there without ever being an object. Accordingly, Fasching (2008, 476-7) calls meditation a movement of "unconstructing" rather than "construction." What has to be "unconstructed" is the identificatory reflective self-consciousness in the form of "I am this-and-that"[98]. That is, a meditator lets go of what she normally takes for her 'self,' including her body, mental capacities, and social roles. When the identificatory self-consciousness is suspended, the so-called pure subjectivity[99] remains. In other words, Fasching describes meditation as de-identification that involves a withdrawal of the "emotional investment" with which one appraises things according to their value with regard to herself. This emotional evaluation is driven by self-interest and involves the notions of 'mine' and 'not mine,' which keep one bounded to 'I.' Thus, "the de-identification from this inner-objective self implies a 'letting go' of the usual attachment to things" (Fasching, 2008, 477). In a meditative state, the meditator does not attribute consciousness to an 'ego,' but rather becomes consciousness. Instead of the self-related content of consciousness or even a perspective, the pure subjectivity is described as unbounded self-luminous wakefulness without boundaries, limits, or a horizon (Metzinger, 2020). The meditator can feel connection,

[98] In Sanskrit 'asmita': "I-am-ness" (Fasching, 2008; Naranjo & Ornstein, 1972).
[99] What Ramana Maharshi called "I-I"—an "I" without a "me" (Fasching, 2008).

relatedness, oneness and presence, but without an ego: while there is no self as an evaluator, neither is there need for evaluation.

The described "pure subjectivity" clearly is interesting for theories of self, since it is characterized by the absence of self-related contents. However, it is a very advanced (or "high") altered state of consciousness, which many times can be reached only through hard training. Alongside the pure consciousness style, there are also forms of meditation that are easier to get familiar with. These include mindfulness, to which we turn next.[100]

8.1.1.2. Mindfulness meditation

According to Fasching (2008), mindfulness shares the general aim of meditation: to uncover the fundamental (minimal) level of self-consciousness that is the self-presence of experiencing itself. Mindfulness does not "fixate" the mind, but a mindfulness meditator choicelessly and non-reactively observes the flow of her mental contents. Thus, a mindfulness meditator cultivates "bare attention" (Naranjo & Ornstein, 1972, 86), and the meditator becomes "an onlooker to his stream of consciousness" (Goleman, 1978, 4). In a way, this kind of observation of mental states might be seen as introspection.[101] However, Fasching remarks that the decisive point in mindfulness does not concern the observed content but rather the attitude of observing itself: one achieves "pure seeing" without reacting to observed things (neither seeking or fleeing some contents). Mindfulness is training in detachment (Goleman, 1978, 4; Austin, 1999, 127) since the meditator experiences herself non-observationally, not as an object but as the observing itself. As in the forms of pure consciousness meditation described above, also in mindfulness "one no longer sets oneself

---

[100] Mindfulness is probably the most popular and studied method of meditation in western countries (see, e.g., Gu et al., 2015; Tang et al., 2015).

[101] Fasching (2008, 470) also notes that "this is exactly the sense in which, e.g., authors like Varela, Thompson, and Rosch take meditation to be important for the investigation of the mind: as a sophisticated technique in observing its goings-on" (Thompson, 2008; Varela et al., 1991). That is, neurophenomenology considers meditation not only as an interesting target altered state of consciousness in the study of self but also as a method for carrying out the study; see Sec. 1.3.3.

up against what appears, through interest-driven occupation with it, but functions as the mere dimension of appearance itself. One simply lets what is there be there and is aware of oneself as this very taking place of being-there, which is the fundamental dimension of subjectivity" (Fasching, 2008, 470).

In current empirical research, mindfulness meditation is often described as nonjudgmental attention to experiences in the present moment (e.g., Hölzel et al., 2011; Kabat-Zinn, 1994; Tang et al., 2015). This definition implies that mindfulness involves both the regulation of attention and the manner in which to approach one's experiences with openness and acceptance. Or more specifically, it has been proposed that mindfulness involves three interacting components: 1) enhanced attention control, 2) improved emotion regulation and 3) altered self-awareness, which includes diminished self-referential processing and enhanced body awareness (Tang et al., 2015, 214; Hölzel et al., 2011). The idea in meditation studies has been that a close interaction of these components constitutes a process of enhanced self-regulation.

Of these three components of mindfulness meditation, the last mentioned is explicitly about altered self-consciousness. This general component of self-awareness also resembles Fasching's description of alteration in self-consciousness in meditation: the reflective self-identification with contents of consciousness is inhibited in favor of the self-presence of experiencing itself. Further, also the two other mentioned components of mindfulness meditation could be considered components of self-consciousness that are needed in producing and maintaining the altered self-awareness. Enhanced attention control involves RSC, which is linked to the general psychological-cognitive aspects of self. RSC seems to be employed in refraining from evaluative self-related thinking that easily arises in mind without focused attention on a task (see Sec. 4.1.1.). Improved emotion regulation involves both MSC and RSC; RSC regulates affectivity, which is a feature of MSC (see also Secs. 4.1.2., 5.3.2., and 6.5.2.). The alterations that self-consciousness undergoes in meditation are elaborated below.

8.2. Meditation alters the dynamics within self-consciousness

On the grounds of the above descriptions of meditation, altered self-consciousness is what a meditative state is all about. The whole purpose of meditation is to alter the typical functioning of self-consciousness: a meditator refrains from normal (self-)reflection in order to focus on minimally self-conscious experience. Thus, meditation is highly interesting in terms of the two forms of self-consciousness and their connections. However, at first sight, meditation seems to have a paradoxical character, especially for RSC. On the one hand, it aims at silencing or dampening of typical RSC in order to reach enhanced (or "purer") MSC. On the other hand, RSC is actively used in order to alter the normal functioning of self-consciousness. This seeming paradox can be resolved by taking a closer look into the operations of self-consciousness. Below, I consider the dynamics of self-consciousness in a meditative state in a little more detail. These considerations partly resemble the analysis of CS in the previous chapter; they show the complex structures in both MSC and RSC and also highlight the links holding the two forms of self-consciousness together.

8.2.1. Minimal self-consciousness in the focus

Since meditation generally shifts the locus of self-consciousness from reflective to minimal form, the features of MSC become accentuated in a meditative state. Actually, meditation provides an especially interesting tool for the examination of MSC, since it brings out tacit aspects that normally remain in the background of consciousness (see Sec. 6.1.). On the other hand, it can be noted that meditation is an altered state of consciousness; the MSC, which meditation brings to the surface, differs from everyday experience. The alterations of MSC are considered below in terms of three aspects of self.

First, meditation highlights the experiential aspects of MSC. Meditation can even be seen as a way to isolate pure consciousness or subjectivity (see, e.g., Fasching, 2008; Metzinger, 2020). In terms of the concepts of this dissertation, this means that meditation can be considered as a way to isolate the experiential aspect of self. In

other words, experience in a meditative state can lack all the typical features of self other than experiential aspects. Descriptions of pure consciousness often are of a wholly negative kind, bringing out the complete absence of content. In particular, high-level mental content, which involves reflective or conceptual thought, is absent. However, affective and sensorimotor content and related bodily self-consciousness are also absent. As Metzinger (2020) describes it, pure consciousness is nonsensory, nonmotor, atemporal, and aperspectival. The pure consciousness experience lacks representation of body and, instead, is described as simply wakefulness, without boundaries or a horizon. This is highly interesting for a theory of self-consciousness since it supports the idea that even if MSC typically involves bodily self-consciousness and a spatiotemporal perspective on the world, it should not be defined as a geometrical perspective (see Sec. 2.2.1.) but as a subjectivity of consciousness at bottom. That is, an experiential aspect is present in the experiences of bare consciousness since these experiences maintain subjectivity. Thus, pure consciousness states can reveal the minimum of self-consciousness and support the idea that experiential aspects of MSC are the necessary and most fundamental feature of self-consciousness. Self-consciousness can lack all other features but not subjectivity.

Second, the de-identification in meditation involves reduced affectivity—in other words, lessening of affective aspects. This also characterizes meditative states that are not instances of "pure consciousness." Typically, affectivity is an important part of experience and also colors self-related thoughts. However, the meditative training aims to withdraw from "emotional investment" and to bring in a state of de-identification that lacks self-interest-driven affects. A meditator turns her focus of attention from the everyday abundance of self-centered affectivity into an emotionally calm or emotionless stance.

Third, meditation involves altered bodily self-consciousness. Mindfulness meditation, at least, involves "enhanced body awareness"; the embodied aspects of self are taken as the hub of focus. Instead of concentrating on wandering self-related thoughts, the meditator focuses on the present moment, for instance, on

breathing, feelings in the body and bodily sensations from the environment (indeed, mindfulness has been described as "coming to our senses," Kabat-Zinn, 2005). By contrast, the "pure consciousness" form of meditation lacks conscious bodily features. In this kind of state, the meditator claims to have an experience of consciousness as such, without any content, not even one's body.[102]

Altogether, meditative experience is different from everyday MSC but seems, interestingly, to give a vantage point into the structures of consciousness. It assists in unconstructing the heavy affective aspects that often accompany self-related thinking, and even the embodied features of experience, and can be considered as a way to approach a bare experiential aspect of self. Thus, meditation is an intriguing example case of altered self-consciousness that can shed light on the most essential experiential characteristics of MSC.

8.2.2. Modification of reflective self-consciousness

Although a meditative state is focused on MSC, it can teach about the functioning of RSC at the same time. Meditation was just defined as a contemporary inhibition of self-identification and as a transition from RSC to MSC. In Chapter 3, RSC was defined as the capacity to take oneself as the object of consciousness and to think of oneself as oneself. Thus, the de-identification of meditation seems to run down both functions of RSC; it especially concerns withdrawing from thinking of oneself *as* oneself and, in pure consciousness forms of meditation, self is not taken as the content of consciousness at all. However, meditation seems to involve RSC at least as an object of modification—a starting point and a means to get into a mental state that is more MSC-oriented.

---

[102] Ciaunica et al. (2021, 14) argue that even the most complex process of meditation is "anchored in one's *bodily* experiences." That is, even when an advanced meditator manages to reach the 'pure presence,' her "basic bodily processing, such as breathing, tacitly ... sustains her subjectivity." It is also acknowledged in this dissertation that self-experience has an embodied basis, even when experience does not involve a representation of the body; see Sec. 2.5. However, the emphasis here is on the descriptions of experience in a pure consciousness state.

8.2.2.1. Narrative aspects abate

In terms of pattern theory (Sec. 1.1.3.; Gallagher, 2013; Gallagher & Daly, 2018), it can be said that the contribution of narrative aspects especially lessens in a meditative state. It has been argued that the present-centered awareness, which meditation aims at, assists in disengaging from narrative forms of self-identification (e.g., Thompson, 2014; Tang et al. 2015). Narrative thinking is important in order to understand oneself as a person with history and plans for the future and as a member of social groups (Sec. 3.6.). However, one can become too attached to some mental representation of herself or get stuck in her personal narratives, entering negative thought circles and rumination. For instance, one can excessively blame herself for some past actions or continuously think about how she should be smarter in order to achieve her future goals. Instead of making one more active, this kind of rumination can enfeeble and turn her passive and thus restrain her from progress. Clinical depression is especially characterized by repetitive negative self-centered thinking and rumination, and interventions based on mindfulness meditation (Gu et al., 2015; Kuyken et al., 2010; Teasdale et al., 2000) have been especially effective in treating depression—in terms of both relieving current depressive symptoms (Strauss et al., 2014) and preventing relapse (Kuyken et al., 2008; Teasdale et al., 2000). Thus, meditation gives one an opportunity to move away from excessive narrativity and evaluation of one's personal life, instead strengthening the more general reflective and cognitive/psychological aspects of self that are essential for regulating attention and maintaining the meditative state.

8.2.2.2. Reflective aspects

At first sight, it is not obvious which mode of self-reflection is involved in meditation, approaching oneself in first-person or third-person (see Sec. 3.2.). Actually, it can be questioned whether meditation can be considered as a form of self-reflection at all, given that it is defined as inhibition of self-identification. However, even if meditation is not self-reflection in the sense of everyday self-evaluation, it seems to essentially involve reflective aspects of self

since these aspects are needed in the modification of altered self-consciousness. In other words, meditation seems to involve almost a paradoxical character in self-reflection.

Roughly, it seems that meditation involves third-personal (3P) examination of oneself: one's self is observed as an object of consciousness among other objects, not as an active agent initiating actions. This seems to be interpretation, for instance, in the meta-analysis of Frewen and colleagues (2020; Sec. 4.1.1.), which associates mindfulness meditation with 'objective self' taking a deidentified allocentric third-person perspective to self-related thinking and emotional processing. The objective self is responsible for open and decentered monitoring of self-related features and also involved in regulating emotions. It just passively "observes" but lacks behavioral orientation. In other words, a meditator looks at her mental states in an open, nonjudging and de-identifying manner; instead of using I-thoughts, she observes the mental states more objectively. For instance, instead of thinking "I am angry," "I am miserable," or "I am maltreated," she can just watch how the feelings of anger, sorrow, and discomfort come and go in the mind. Thus, it could be said that the meditator approaches herself in 3P manner.

The emphasis on 3P mode is interesting from the viewpoint of self-consciousness theory since 1P mode to oneself is typically considered more essential. Self-conscious thoughts have unique epistemic and motivational features (see Sec. 3.2.), and strong 1P mode as a 'deliberative stance' on oneself is considered crucial to psychological well-being (Sec. 3.3.; Moran, 2001; see also Sec. 7.2.). To recap, it is important for psychological well-being that one has a special first-personal relation to herself and, by realizing this special characteristic, can actively direct her action. However, it seems the deliberative stance ceases to be beneficial for well-being if it is employed in excessive amounts or an unbalanced manner when it can lead to the exaggerated narrativity that was mentioned above. Thus, the rather 3P stance towards oneself that is employed in a meditative state can be beneficial in running down the immoderate narrativity and 1P thinking.

On the other hand, it might be argued that 1P or a more deliberative stance on oneself is needed in order to get into and sustain the meditative state. This 1P stance can be seen in the regulation of attention and emotion in meditation. It was discussed in Sections 4.1.2. and 5.3.2. that self as cognitive–affective agent is present in the cognitive control of attention in effortful emotion regulation. The case was interesting since it extended the notion of self-specific processing, coupled with MSC, to complex cognitive processes that explicitly involve modification of both evaluation and emotions. Christoff and colleagues (2011) proposed that self is specified in effortful emotion regulation as the cognitive–affective agent who reinterprets and thereby controls her emotional response. Deliberate emotion regulation involves reappraisal, which relies on the cognitive control mechanisms that are needed for attention-demanding tasks. Frewen and colleagues (2020; Sec. 4.1.1.) also find these mechanisms interesting and argue that they are involved in objective and executive selves. The executive self might be connected to deliberative stance, which has a 1P approach to oneself and is needed in the initiation of action. Thus, meditation also seems to involve first-personal elements of RSC.

Combining the ideas of 1P and 3P reflective elements, meditation could be seen as a process involving flexible adjustment of self-reflection. This is interesting for self-consciousness theories, since it is argued (Moran, 2001) that in addition to the emphasis on deliberative stance, psychological well-being requires alternation of theoretical and deliberative stances. Even if deliberative 1P is important for action and determining oneself, it is equally important that one is able to examine herself descriptively in 3P, in the same way as others, since this capacity enables taking an objective point of view. It is of utmost importance that one can shift between these stances, since each stance is needed in different cognitive tasks. In addition, for instance, Ciaunica and colleagues (2021) argue that a healthy self-consciousness involves the capacity to flexibly connect and switch between the I-as-subject and the I-as-object forms of experience. It seems that meditation involves this kind of flexible connectedness in reflective aspects. The meditator needs deliberative stance in order to get into the meditative state

and in order to keep her attention on MSC. The meditative state involves self-as-subject as cognitive–affective agent carrying through the modification of self-consciousness but also observation of self with a rather third-personal manner. Thus, meditation can be seen as a mix of elements in both modes of self-reflection and as an interesting interplay between them. This interplay cannot be further elaborated here, but for the purposes of this dissertation, it is sufficient to see the general wide-ranging changes in RSC.

8.2.2.3. Psychological cognitive aspects

Further, the complex modification of self-consciousness in reaching a meditative state indicates that RSC involves more features than just the taking of 1P or 3P stance. In Chapter 3, voluntariness and mode were also mentioned as factors that contribute to the whole of RSC, and these features play a key role in meditation.

First, the mode of RSC is crucial in meditation: the whole manner of mental activity is altered. The point is exactly not to enter an everyday mind-wandering or evaluating RSC but to adjust or enact the way of reflection. For instance, Thompson (2015, 350) refers to meditation studies with the subtitle "From mind wandering to mindfulness." Further, Thompson points out that a wandering mind is a less happy mind (referring to Killingsworth & Gilbert, 2010) and that the benefits of meditation include a reduction of mind-wandering and an increase of cognitive control. That is, the observed objects are not crucial, but the attitude or manner of observing is decisive in meditation. This nonjudgmental attitude enables detachment and non-evaluative observation of one's operations of mind and emotions. Indeed, meditative training aims at altering or "de-automatizing" a wandering mind, which is easily caught by self-referential thoughts, towards a more focused and cognitive-emotionally flexible one.

Second, it can be remarked that meditation is voluntarily initiated; it involves an effort of RSC to enter a meditative state. Meditative RSC deviates from everyday self-reflection and requires regulation in order to maintain the meditative state and not to "slide" to mind-wandering. This strengthens the conclusion made in the previous chapter that voluntariness and flexible modification

of the modes of reflection are important features of RSC. In other words, although the I-thoughts are important for self-understanding and action, RSC is not only about examining self as the object of consciousness but also about the manner of employing this capacity. The way by which self is taken to the locus of consciousness is essential since it influences the related experience.

8.2.3. Strengthened and widened connections between minimal and reflective self-consciousness

The key feature of meditation is the change in the dynamics of self-consciousness, which implies a reform of the connections between MSC and RSC. In a meditative state, everyday evaluative RSC is replaced by greater awareness of the present-moment experience. By adopting the experiential focus, the meditator avoids getting stuck in her self-narratives. In other words, the meditator can flexibly modulate and move between self-related evaluative-narrative thinking and present-centered embodied awareness. It has been suggested that exactly this shift in self-awareness is a major mechanism of the beneficial effects of meditation (e.g., Tang et al., 2015; Thompson, 2014).

The interplay of minimal and reflective self-consciousness in meditation can also be seen in the conception of self as a cognitive–emotional agent. This kind of cognitive–emotional agency has been considered as self-specific processing that is linked with MSC (Christoff et al., 2011). Christoff and colleagues (2011) propose that effortful emotion regulation can implement a functional self–non-self distinction in which the self initiates the effortful reappraisal process that targets the emotional scene (non-self). Meditation can be described in terms of this idea by Christoff and colleagues (2011); self is an object of consciousness and an overly emotional sense of self is withdrawn and thus, reappraised as non-self. At the same time, a self-as-subject is present as the cognitive–affective agent that regulates the narrative and emotional aspects of experience. Thus, also this conception seems to support the idea that meditation enhances the minimally self-conscious self-as-the-subject ('I') and reduces the dominance of reflectively self-conscious evaluation of self-as-the-object ('me').

Further, the capability to adaptively and appropriately use this kind of shifts seems to strengthen the links between MSC and RSC. In order to modulate and maintain the balance within the multiple aspects of self, the connections between the two forms of self-consciousness need to be strong and extensive. These strengthened connections form a wide and thick network, which can provide the basis for the cognitively demanding withdrawal from RSC to stripped-down MSC. Within another explanatory level, the changes in the connections between MSC and RSC can be considered in terms of empirical studies that show new kinds of connections in the neural networks of expert meditators' brains.

8.2.4. Neural base of meditation

Empirical results of meditation-related brain functions are briefly considered here in order to elaborate on the involved dynamics of self-consciousness and its underlying embodied features. Empirical results especially give strong support to the general idea that meditation produces a shift from a self-related "narrative" focus on oneself towards a more "experiential" focus in which the meditator is more aware of her present-moment experience (Tang et al., 2015; Thompson, 2014).[103] Meditators often report improved attention control and emotion regulation (which are also mentioned as two of the three components of mindfulness meditation by Tang et al., 2015), as well as enhanced awareness of bodily states and increased sensitivity to interoception (the third component of mindfulness meditation) after training.

The most obvious neuroscientific result concerning self-consciousness in meditation is changes in activity and connectivity of the default mode network (DMN), which is involved in self-related processing (and also e.g., mind-wandering, Sec. 4.1.1.). That

---

[103] Generally, empirical studies have established that practicing meditation induces changes in multiple neural systems and functions, suggesting that meditation affects large-scale brain networks (see, e.g., Fox et al., 2016; Gu et al., 2015; Tang et al., 2015). Meditation is shown to involve consistent changes at least in the ACC, PFC, PCC, insula, striatum and amygdala. This is not surprising from the viewpoint of self-consciousness, since (as presented in Sec. 4.1.) the basis of the complex whole of self-consciousness involves various neural networks and their interaction.

is, the typical DMN activity is reduced, and other neural processes are enhanced, in meditators' brains. For instance, a study by Farb and colleagues (2007) showed mindfulness training results in increased connectivity of the insula with dorsolateral PFC, and uncoupling of the right insula and medial PFC. This was interpreted as a shift in self-processing from affective or subjective self-referential evaluation (especially involving medial PFC) towards a more objective and detached and observation (dorsolateral PFC) of interoceptive and exteroceptive sensory information (insula).

It is also interesting that the neural profiles of beginning and expert meditators differ significantly, which further supports the idea that meditation practice can bring about structural changes in the functioning of self-consciousness. For instance, Tang and colleagues (2015) remark that beginners of mindfulness meditation show greater prefrontal activation, which can be associated with active cognitive regulation in order to overcome habitual ways of reacting to emotions. Instead, expert meditators have automated the accepting stance toward experience and do not employ this kind of prefrontal control. Thus, Tang and colleagues (2015, 218) propose that expert meditators do not have to engage in top-down control efforts anymore but instead have enhanced bottom-up processing. Accordingly, it has been found that experienced meditators require less neural activity from executive regions in order to disengage from mind-wandering (Hasenkamp et al., 2012; Tang et al., 2015).

Further, it has been proposed that experienced meditators have increased cognitive control over the function of the DMN (Brewer et al., 2011; see also, Brefczynski-Lewis et al., 2007; Tang et al., 2015; Thompson, 2014). In other words, self-related processing, which DMN supports, is more accessible to monitoring and control in expert meditators than in novices. This can be seen in results that show stronger coupling between the PCC, dorsal ACC and dorsolateral PFC in experienced meditators. In other words, when experienced meditators' DMN is activated, it is coactivated with different brain regions than in novice meditators' brains. Interestingly, this functional connectivity can be seen both at

baseline and during meditation, which arguably indicates that the accessibility of mind-wandering and cognitive control over it might eventually bring about a new default mode of brain activity (Brewer et al., 2011). Since this new default is also present during passive resting states, long-term meditation really seems to lead to less uncontrolled mind-wandering and, by implication, less-negative mood states (e.g., Thompson, 2014). In addition, the widespread neural changes in meditation can rather plainly be seen, for instance, in the studies that show a strong gamma synchrony in expert meditators' brains (Berkovich-Ohana, 2017; Lutz et al., 2004; F. Varela et al., 2001). That is, distant brain regions show communication by coordinated firing rhythms during the meditation practice, thus forming temporarily a massive, interconnected network. This has been described by saying that practicing meditation affects the architecture of the "brainweb" (Varela et al., 2001).

However, more empirical research is still needed in order to establish the changes that meditation can bring about in self-consciousness and its neural base. As Tang and colleagues (2015, 220) point out, "interpretations are built on a still-fragmentary understanding of the function of the involved brain regions," "many experiments are not yet based on elaborated theories," and "Interpretations of study outcomes remain tentative until they are clearly linked to subjective reports."[104] Thus, neurophilosophy can be useful for empirical studies of meditation; it can provide a theory of self-consciousness that acknowledges different dimensions of self and is interested in links between concepts of self, self-experience, and its underlying neural mechanisms (Secs. 1.3.3. and 4.2.1.).

---

[104] It is important to also note other limitations of current empirical studies of meditation (see, e.g., Fox et al., 2016; Tang et al., 2015). For instance, many findings have not yet been replicated, the direction of effects has not been consistent across all studies, and there are individual differences in responding to meditation practice. In addition, the methodological quality of many meditation studies should be improved; longitudinal studies rarely are actively controlled, sample sizes are small and research might be biased by an enthusiasm to "prove" the positive effects of meditation. Thus, some critics caution against "mindfulness hype" (Van Dam et al., 2018).

On the other hand, empirical studies are useful for theory of self since empirical results can give a concrete illustration of the dynamics of self-consciousness. Generally, the results support the idea that meditation involves changes in the connections of self-consciousness; neural processes that are associated with self-consciousness change during and as a result of meditation. Especially, typical activity in DMN that is linked to self-related RSC decreases. At the same time, connectivity increases, for instance, between frontal and insular areas, the former related to attention and the latter to bodily processes, arguably enhancing MSC. These changes indicate a shift from self-centered and evaluative RSC towards focused and emotionally tranquil MSC alongside alteration in the dynamics between these two forms of self-consciousness. Meditation seems to strengthen these connections and also generate new ones. This can be seen in that beginner meditators seem to use top-down control processes of self-consciousness, whereas expert meditators can focus on bottom-up experiencing. The profound alterations in expert meditators' neural functions suggests that they are more also able to flexibly switch between different modes of self-consciousness when they are not in a meditative state.

8.3. Similarities between meditation and depersonalization

It has been proposed (e.g., by Metzinger, 2003; Ciaunica et al., 2021) that depersonalization and meditation resemble each other as altered states of consciousness. Both states involve withdrawal from self, and this withdrawal is associated with diminished self-affection and changes in self-reflection. In depersonalization, self feels unreal, and the idea of meditation is to grasp the potentially illusive character self and refrain from typical self-centered thinking. In contemplating the similarities between meditation and depersonalization, Ciaunica and colleagues (2021, 14) point out that "one may argue that depersonalization-like states are inevitable steps on the challenging path towards self-exploration to the ultimate no-self or 'pure subjectivity' state targeted by Buddhist-

based meditative practices." However, Ciaunica and colleagues (2021, 14) present this as an open question, which requires further examination and systematic comparison between i) meditation-induced forms of DPD (depersonalization-derealization disorder) and ii) trauma, drugs and anxiety-induced forms of DPD.

Here I conduct this kind of systematic comparison between meditation and depersonalization. However, I find it misleading to talk about differently "induced forms of DPD." More accurately, different experiences of depersonalization are present in the listed different states of consciousness, but 'DPD' refers to a pathological depersonalization-derealization disorder in which the feelings of depersonalization have become chronic and reach the threshold for diagnosis (see Sec. 6.3; e.g., Reutens et al., 2010; Sierra & Berrios, 1998). Although meditation and drug-induced altered states of consciousness involve some kind of experience of depersonalization, the depersonalized experience in DPD has a distinct pathological character. In the terminology of this chapter, 'depersonalization' refers to the pathological form of depersonalization in particular, and this pathological experience is contrasted with the experience of meditation.

Further, I want to highlight that there are also significant dissimilarities between the experiential profiles of these two altered states of consciousness. These dissimilarities are linked at least to the extent of RSC, and changes in MSC, including the lack of affectivity. In addition, the comparison between meditative and depersonalized states can illustrate the different ways by which minimal and reflective self-consciousness are connected to each other and clarify the major role of these connections. In this subchapter, I consider the similarities between the states, and in the next subchapter I examine their differences.

8.3.1. Detachment from identification

Based on the descriptions above, it is rather clear that the altered self-consciousness in meditation and depersonalization share at least some characteristics. Generally, their common feature could be called de-identification: in both states, a subject desists from everyday self-identification. Ciaunica and colleagues (2021, 11)

describe this shared aspect as "the experience of a change or a 'detachment' from how we normally perceive ourselves, and how we relate to our ordinary sense of self." They argue that depersonalization involves a "split between an observing and observed self" (Ciaunica et al., 2021, 115) and that this kind of split is also experienced in some meditative practices, which allow a meditator to modulate her basic pre-reflective forms of self-consciousness. In other words, both conditions involve an experience of increased distance between the self-as-object observed by RSC and the minimally self-conscious self-as-the-subject that is observing.

Meditation involves a change in the dynamics of self-consciousness in which RSC makes way for MSC. A meditator de-identifies with what she typically takes her self to be and instead absorbs into subjective presence. This accessed MSC is essentially detached from everyday affectivity and emotional load. In addition, the MSC in meditation involves altered bodily self-consciousness. Mindfulness meditation has been described in terms of enhanced bodily consciousness, in which immediate bodily sensations can be the focus of attention, whereas pure consciousness meditation is more about the "bare" experiential aspect of self.

Further, the meditative detachment concerns RSC. A meditator can let her mind go through all the qualities she typically associates with herself, but instead of entangling in them, she looks at them rather neutrally, without holding any of them essential. Ciaunica and colleagues (2021, 11) highlight that a meditator modifies the familiar way in which she perceives herself and the world, "leaving behind" first and foremost her former self (or ego). In other words, the meditator actively aims to withdraw from being occupied with contents of consciousness of which many are self-referential. Instead of self-conscious thoughts, the open and nonjudgmental manner of the exercise is the determining factor of meditation. Since self is seen as an object among other objects, meditation seems to employ the 3P mode of RSC. Thus, it can be said that a meditator de-identifies from her normal evaluative RSC and self-related thoughts.

Depersonalization is also characterized by alterations in self-consciousness. In this pathological state, one feels estranged from herself: the features of self, which one normally feels familiar with, do not have their ordinary intimacy. Although a depersonalization patient can think of herself and focus on her qualities, she feels herself to be unreal and is somehow detached from her thoughts and feelings. Thus, depersonalization is characterized by de-identification of what one typically considers her self to be.

As analyzed in the previous chapter, depersonalization arises from diminished MSC, which results in distorted RSC and in the loop between altered MSC and exaggerated RSC. Within MSC, a patient suffers especially from diminished self-affectivity and self-presence, and also from derealization (in addition to one's self, the world also feels unfamiliar or unreal) and anomalous bodily self-awareness. That is, depersonalized MSC involves a lack of affectivity and atypical bodily feelings, described as feelings of detachment from one's self and world.

Within RSC, a depersonalization patient employs a 3P mode of self-reflection especially and finds this distressing, while her 1P approach to herself is deficient in the normal sense that enables cognitive control over MSC. This implies that the relations between minimal and reflective forms of self-consciousness are also atypical in depersonalization. The balance of these relations is lost, and instead, RSC is compulsively turned towards minimally self-conscious feelings of unfamiliarity. RSC cannot correct or alter those feelings but remains captured by them, and this creates a vicious circle of de-identification, running from alienated feelings to alienated thoughts and then back to feelings.

Thus, detachment from identification can be seen in both meditation and depersonalization, and it manifests itself in all levels of self-consciousness. First, altered MSC: the de-identification is characterized by lack of typical self-affectivity and involves alterations in bodily self-consciousness too. Second, the de-identification concerns refraining from emotionally loaded self-reflection. Detachment involves a rather third-personal examination of oneself, in which self is considered more like an object among other objects instead of taking a first-personal mode

that is more directly linked to action. Further, the usual balance relations between MSC and RSC alter in de-identification: altered MSC is accentuated and RSC is repeatedly employed in focusing on MSC. Altogether, diminished self-affectivity and stepping back from one's typical self-centered thoughts are present in both of these altered states of consciousness; a subject de-identifies from what she typically regards as her self.

8.4. Differences between meditation and depersonalization

Although they do share some common features, it is evident that experiences in depersonalization and meditation are not the same. Instead, there are major differences between what it is like to be in a depersonalized versus a meditative state. Ciaunica and colleagues (2021, 13-14) briefly remark that a trained meditator can detach from her ordinary self by modulating her pre-reflective self-consciousness, and "becoming self-aware in this elevated sense may lead" her to reach a level of experiencing as 'pure subjectivity'—"as opposed to 'object-like' feelings in DPD." This characterization can be considered a start for tracking the dissimilarities between meditation and depersonalization. These differences are clear in the descriptions of experience by depersonalization patients and meditators. Pathological depersonalization is a cause of anxiety and suffering; generally, it is a "negative" state. In contrast, meditation is often described as a pleasant "positive" state.

Thus, it is highly interesting *why* the de-identification in one case leads to "elevated sense of being self-aware," "pure subjectivity" and positive feelings, whereas in the other case it leads to anxiety, "object-like feelings" and pathological distress. My short answer is that the fundamental difference between depersonalization and meditation is that the former is a pathology, in which a patient has not chosen to be, whereas the latter is actively self-induced by a meditator. This difference concerns not only the mechanism by which the state is induced but also the experienced state of self-consciousness. This tentative answer is elaborated

below by analyzing the differences between meditative and depersonalized states in terms of MSC, RSC and their connections.

As an introductory remark, it can be noted that the incongruity between meditation and depersonalization seems evident already in terms of the three components of meditation. First, meditation involves enhanced attention control. In DPD too, focused attention is increased; however, the involved attention is involuntary and thus it is not an instance of enhanced attention *control*. Second, emotion regulation is improved in meditation, whereas depersonalization lacks this kind of regulation. Instead, a patient suffers from feelings of unfamiliarity and unreality that she cannot modify even if she wanted to. Third, a meditator's self-awareness alters towards diminished self-referential processing and enhanced body awareness. Depersonalization patients' self-awareness alters as well, but almost in the opposite way; she cannot get rid of the thoughts of unrealness of self. Further, although she is thinking of her body repeatedly, her body awareness is not enhanced, but rather the body feels remote and object-like.

8.4.1. An alteration versus a disturbance of MSC

A clear difference between meditation and depersonalization is that the former is described in terms of enhanced MSC, but the latter in terms of distorted MSC. Even if a meditator's MSC is not enhanced, it is not distorted in a meditative state, but altered as a result of mental practice. In the sense used here, distortion is a subclass of alteration (in the same way as pathological states of consciousness can be considered a subclass of altered states of consciousness; see Sec. 6.1.). The term 'distorted' refers in the first place to experience; the patients suffering from depersonalization describe their experience as distorted in the sense that it lacks something and this lack feels unpleasant. The experienced distortion is connected to distortions or dysfunctions in the (non-conscious) structures of self and associated with changes in neural processing.

Although both meditation and depersonalization involve reduced affectivity, a closer examination shows that the lack of affectivity is not of the same kind. A meditator regulates her MSC,

268

and due to this mental practice, she can abate affectivity and instead focus on the present moment and her bodily sensations. By contrast, depersonalization arises from distorted MSC that is disturbed by inexplicable feelings of diminished self-affection. A depersonalized patient cannot change this strange experience even if she wants to but is captured in the diminution of self-affection. Even though de-identification in depersonalization bears some resemblance to an actively produced meditative state, the alterations of MSC in depersonalization are pathological, a corollary of mental malfunctioning and beyond the patient's control. The voluntariness can be linked to the process that brings in the meditative or depersonalized state (see also Sec. 3.4.); however, I propose that it also plays a role in the experience of the state. The diminution of affectivity feels different in these states; in depersonalization, it is felt as disruptive, whereas a meditative state lacks this kind of feeling. I want to point out that a possible explanation for the disruptive feeling of the lack of affectivity in depersonalization could be a related lack of sense of agency. As presented in Section 2.4.1., experiential aspects of self involve a sense of agency, which many times coincides with subjectivity but can also be absent from experience (Gallagher, 2000). In terms of this concept, depersonalization involves the lack of a typical sense of agency, whereas a meditator's tacit sense of agency is intact. This is also consistent with empirical studies of depersonalization, which indicate abnormalities in neural functions that are associated with experiencing a lack of agency (e.g., Sierra & David, 2011; see Sec. 8.4.4.). In other words, since the depersonalized lack of affection is not chosen, it results in negative feelings of anxiety. This negative feeling is not present in meditation, in which the lessened affectivity is volitional.

In addition, bodily self-consciousness is altered in both depersonalization and meditation, but the alterations differ from each other. While some forms of meditation involve enhanced bodily-consciousness, the body is experienced as unreal and unfamiliar in depersonalization. In depersonalization, these features are not voluntarily induced, and one's body feels more object-like. The anomalous bodily awareness in depersonalization can also be

viewed through the distinction between the senses of a body as Leib and Körper (discussed in Sec. 2.1.). To recap, in the German language, there are two senses for the term 'body': 'Leib' refers to the lived, feeling, and expressive body, whereas 'Körper' refers to the body as it appears when examined like any other extended object. For instance, Colombetti and Ratcliffe (2012, 148) argue that in depersonalization, a body "loses its character of Leib and appears more like a Körper. In other words, its merely physical and 'thing-like' features become a primary focus of awareness.... In depersonalization this sense appears exaggerated, and the Körper rather than the Leib comes to predominate in awareness." By contrast, while a meditator turns attention to her bodily self-consciousness, the body is felt as lived and feeling, and thus, rather the sense of Leib seems to be in use. Or even if the body was examined as Körper in a meditative state, this examination does not originate it the anomalous feeling of unfamiliarity, as in depersonalization, but is actively and voluntarily induced.

8.4.2. Flexible and nonjudgmental versus involuntary and objectifying use of RSC

In this section, I examine how the operations of RSC in meditative and depersonalized states are different in terms of voluntariness of the third-personal stance, mode of reflection and contribution of narrative elements.

Both meditation and depersonalization seem to involve a third-personal stance toward oneself. However, phenomenology is not the same. A meditator strives to "look at" her mental states with a nonjudgmental attitude, thus taking distance and withdrawing from her self-attachments. In order to maintain the nonjudgmental focus, she actively modifies her RSC, possibly involving an interplay between 3P and 1P stance to herself. Instead, depersonalization involves hyper-reflectivity—that is, an exaggerated, non-voluntary and objectifying self-reflection. A patient suffering from depersonalization employs a third-personal viewpoint on herself and continually has thoughts in which her self appears as remote and object-like. However, her RSC seems to be partly "impotent": it cannot restore affectivity, and she fails to take the first-personal

deliberative stance by which she could feel herself as an active whole.

Thus, the difference of the voluntariness in meditation and depersonalization can also be seen within RSC. Hyper-reflectivity in depersonalization is involuntary; the depersonalization patient does not have an intention for the third-personal self-reflection, and her thoughts do not just flow by in her stream of consciousness but are striking. By contrast, meditation is voluntarily initiated and involves maintaining a nonjudgmental mode. In other words, in meditation, the use of RSC modifies MSC by taking a mode of detachment, and this requires flexible modulation of RSC. In the case of depersonalization, the voluntariness and power of RSC are failing, and the patient is caught in an involuntary objectifying reflection of her pathological experience of detachment.

These differences are also evident in terms of the distinctions made within RSC in Chapter 3—that is, the distinctions between deliberative and theoretical stance, voluntary and unvoluntary self-reflection, and identification and non-identification. Depersonalization is characterized by a prolonged theoretical stance toward oneself, whereas meditation involves interplay between theoretical and deliberative stances. Both meditative and depersonalized states involve self-reflection without identification; however, it seems that the case of depersonalization would not count as an instance of non-identification in the sense discussed in Chapter 3, which considered non-identification as a voluntary capacity or mode of RSC. The voluntariness was presented as an initiation of the process of self-reflection, but the comparison between meditation and depersonalization here indicates that it can also bear an effect to the experience of self-reflection. The same point was made above in examining the voluntariness of lack of affection as a difference between meditation and depersonalization. The involuntariness is associated with feelings of distressing disruption in the pathological prolonged case of depersonalization (which is not present in more everyday cases of involuntary self-reflection such as orientation by a change in environment or mind-wandering). Instead, meditation does not involve these kinds of feelings of disruption or distress. As I presented above, the

distressing feeling might result from a lack of the sense of agency in depersonalization. This interpretation is also consistent with the idea that depersonalization lacks (at least partly) the deliberative stance toward oneself, which is linked to feelings of active agency.

Further, the contribution of narrative aspects of self seems to be different in meditation and depersonalization. While meditation can be considered a way to effortfully lessen especially the self-related thinking and narratives, depersonalization seems to involve feelings of disrupted self-narrative. Instead of a chosen withdrawal of the narrative, a depersonalization patient suffers from being cut out of it and seeks to get hold of the narrative again. Thus, although depersonalization can be considered a form of de-identification in one sense, it does not meet the characteristics of meditation. That is, depersonalization does *not* involve forgetting or letting go of one's self at all in the final analysis. On the contrary, one's self is painfully present as if unreal, and this feeling is in the locus of attention. That is, actually a patient's thinking remains self-referential, since her missing self is searched for all the time. Thus, the pathological experience is tied to self, not lacking it.

8.4.3 Extended versus narrowed connections between MSC and RSC

Furthermore, the connectivity between MSC and RSC differs in meditation and depersonalization. In meditation, RSC is used to regulate MSC and the stream of consciousness. There is no malfunctioning in MSC, but the connections between MSC are voluntarily driven, and their character is actively modified. A meditator can go through her thoughts and feelings, but by maintaining an approving attitude, she can withdraw these thoughts and feelings and proceed towards a present-moment experience. In order to succeed in this and not be carried away by some self-related aspect of self-consciousness, she needs to exercise the very connections between MSC and RSC. Thus, it can be said that meditation involves extending and tightening these connections and develops the art of making them more flexible.

In contrast, depersonalization seems to involve pruned connections between MSC and RSC. Depersonalization is

characterized by hyper-reflectivity; diminished MSC is constantly taken as the object of third-personal RSC in an involuntary or passive manner. Since the lack of affectivity in MSC is the source of anxiety originally, paying RSC to it just sustains the vicious circle between the diminished self-affectivity and the reflection on the unreality of self. This vicious circle can rule the focus on consciousness so completely that it can be hard for the patients to concentrate on anything else. Both MSC and RSC lack features that they usually involve; MSC especially lacks affectivity, and RSC lacks a first-personal stance toward oneself. In addition, the whole of self-consciousness lacks connections that usually hold the two aspects together. The connections are stiffened and narrowed compared to normal; instead of an extensive network of connections between the aspects, only limited connections prevail between limited MSC and limited RSC. The connections have lost their normal flexibility and cannot be extensively modified, but diminished MSC compulsorily remains as the object of third-personal RSC.

Thus, importantly, the comparison between meditation and depersonalization indicates that the connections between MSC and RSC matter. In both states, MSC and RSC are different from typical. However, examination of them separately does not suffice to account for the altered self-consciousness in question; the way in which they are connected is also crucial. Depersonalization involves the vicious circle between hyper-reflectivity and diminished self-affection; the connection is involuntary and narrowed from normal. By contrast, meditation involves voluntary and active modification of the connections of MSC and RSC; RSC is employed in order to adjust MSC in a special manner. In order to maintain the focus in the present moment, the connections within self-consciousness need to be strong and extend from normal circles of mind-wandering.

The differences in alterations of self-consciousness in meditation and depersonalization are also rather clear in terms of the recent predictive processing model of mind (see Sec. 6.5.2.). For instance, Gerrans (2015, 2019) argues that the involuntary lack of affectivity in MSC, which characterizes depersonalization, is a

prediction error. This unpredicted and surprising error causes anxiety and RSC is used to correct the error. However, although RSC constantly allocates attention to the error, it cannot be corrected top-down, and the distressing vicious circle keeps on operating. By contrast, in meditation, the transition to the altered MSC is voluntarily induced; it is in accordance with predictive models. Further, it might be proposed that the practice of meditation can develop and reinforce the predictive models and self-regulation by bringing more control and visibility to the models. In other words, meditation could enable fine adjustment of a predictive coding system.

8.4.4. Differences in neural activation

The differences between depersonalization and meditation can also be seen in differences in their neural realizations. As on the experiential level, the neural activations related these two altered states of consciousness have some common features but also differ significantly from each other. The common features include frontal activation that can be connected to cognitive-attentional self-related RSC and activation in the parietal cortex and deeper brain structures that can be connected to bodily-affective self-specific MSC. However, the exact activations and related neural networks are different in meditation and depersonalization.

As a very brief characterization, depersonalization involves dysfunctions in two neural networks: fronto-limbic and parietal (proposed, e.g., by Sierra & David, 2011). The latter network involves areas that are relevant to the experience of embodiment and feelings of agency (Sierra & David, 2011). For instance, depersonalization involves abnormally increased activation in the angular gyrus of the right parietal lobe, and this abnormality has been associated with experiencing a lack of agency (Farrer et al., 2004; Frith et al., 2000). In other words, the changes in parietal processing seem to be related to self-specific processing that underlies MSC. However, the first-mentioned fronto-limbic mechanisms have been discussed more in studies of depersonalization (see, e.g., Reutens et al., 2010; Sierra & Berrios, 1998; Sierra & David, 2011). As presented in Sec. 6.5.2., for

instance, Gerrans (2015; 2019) argues that this mechanism is crucial in explaining the experience of depersonalization. Gerrans highlights abnormal activity especially in two neural structures. First, hypoactivity in the Anterior Insular Cortex (AIC), which has a primary role in higher-order representation of interoceptive states. And secondly, hyperactivity in the Ventrolateral Prefrontal Cortex (VLPFC), which plays a crucial role in the regulation of affective feeling, since it enables redirecting attention and diverting cognitive resources to alternative interpretations of self-relevance. According to Gerrans, hyperactivity in the VLPFC leads to hypoactivity in the AIC; a depersonalization patient suffers from involuntary deactivation of mechanisms that produce the bodily experience of emotion and thus, experiences the loss of a sense of presence. It was proposed in Section 6.6.1. that this dysfunctional mechanism could (at least partly) underlie the vicious circle between MSC and RSC.

Thus, both meditation and depersonalization involve changes in neural functioning but several differences in these changes can be seen directly. An obvious difference related to self-specific processing (connected to MSC) concerns insula; depersonalization is characterized especially by hypoactivity of AIC whereas meditation involves enhanced insular activity. That is, they have rather opposite patterns in insular activity. This significant difference in insular activity matches also enhanced bodily-awareness in meditation and experiences of lack of affectivity towards the body in depersonalization. In terms of self-related processing (connected to RSC), both altered states of consciousness involve changes in DMN and frontal areas; both are associated with increased activation in prefrontal areas but the activated areas are not entirely the same. In depersonalization, especially VLPFC is hyperactive, whereas in meditation, for instance, DLPFC and VMPFC show increased connectivity.

It is remarkable that meditation and depersonalization involve different kinds of changes in neural connectivity in general, and these changes could be considered changes in the connections between MSC and RSC. Meditation is characterized by enhanced connectivity between several neural systems and also diminution in

typical DMN; in other words, it induces extensive changes in neural patterns. By contrast, increased activity in DPD seems to be more restricted, occurring mostly in some parietal areas and the downregulation process from VLPFC to AIC. That is, depersonalization is characterized by more fixed and limited neural patterns. More detailed comparison of the neural mechanisms of meditation and depersonalization would be interesting but cannot be elaborated here. For the purposes of this dissertation, it is sufficient to remark that empirical research also supports the claim that meditation and depersonalization differ significantly from each other, even if they share the experience of detachment to some extent.

8.4.5. Self-induced versus suffered alteration in a pattern of self

Summarizing, meditation and depersonalization are not the same kind of altered states of consciousness, although they have some common features. The similarity between the two states is detachment or de-identification, which involves withdrawal from typical emotional self-related contents of consciousness, attentional focus to this diminished affectivity, and a third-personal stance toward oneself. However, in meditation the withdrawal is voluntarily induced: it helps one to organize and calm her mind and to see more "clearly" into her MSC while slackening self-evaluation and narrativity. By contrast, in depersonalization, one feels herself as if unreal due to distorted MSC, and the withdrawal is involuntary: although one examines it, she cannot change it but is captured in hyper-reflective self-reflection. Further, the connections between MSC and RSC are different in meditation and depersonalization, the former involving active, flexible and extensive modification, whereas the latter involves a passive and narrow vicious circle.

In terms of the pattern theory of self, both altered states of consciousness can be considered altered patterns of the network of features of self. It is interesting that although the altered states share the experience of detachment, they display almost opposite overall patterns of self. Indeed, depersonalization is characterized by

diminished MSC, hyper-reflectivity and narrowed overall connectivity between the two. Meditation, on the other hand, is characterized by enhanced MSC, diminished self-referential RSC and an extensive connectivity between the two.

These different patterns can also be considered differences in the balance within self-consciousness. Depersonalization involves pathological and involuntary unbalance; the vicious circle dominates the self-consciousness while parts of MSC, RSC and their connections outside the vicious circle are withered. More precisely, embodied, affective and experiential aspects of MSC are diminished; the lack of affectivity is especially striking. This altered MSC is scrutinized by RSC, which lacks typical 1P elements and employs a 3P objectifying stance towards oneself. The narrative aspects are not completely lacking, but a patient feels a disconnection from her memories, thoughts and future plans. Hyper-reflectivity involves narrowed connections between MSC and RSC; the connections keep the feelings of unreality and disconnectedness in the center of experience but do not reach other features of self in the way they are present in typical everyday experience. While the vicious circle obtains, the balance of self-consciousness is lost for the third-personal experience and approach without strong first-personal elements.

By contrast, meditation does not involve this kind of pathological unbalance, although it involves a change in dynamics of self-consciousness. Instead, meditation could be seen as a state that *brings* balance to self-consciousness. In one sense, meditation is a way to mend an unbalance in which RSC is overemphasized by turning the attention towards MSC. In a mindfulness meditative state, none of the typically active aspects of self, including affective and narrative aspects, dominates but can be (partly) "shut down" to make room for experiential present-moment focus. In other words, meditation can be seen as a state in which balance within self-consciousness moves or is returned from excessive RSC to calm MSC.

In the other sense, meditation can be seen as building balance in self-consciousness by strengthening the connections between MSC and RSC that enable a flexible modulation of the features of

self-consciousness. The balancing highlights the links between MSC and RSC, since the links are needed to alter the dynamics of self-consciousness and maintain the sought-after state. In modulating the balance within the multiple aspects of self, the connections form an extensive network within aspects of self, since only strong and flexible connections can ensure the cognitively demanding withdrawal from RSC to stripped-down MSC. In pure consciousness style meditation, it seems that only experiential aspects of self are underway, while all other aspects are put out, and this putting out requires considerable alterations in the typical dynamics of self-consciousness. Thus, meditative practice can be seen as a way to develop the skills to reach balance in and modify the whole of self-consciousness.

8.4.6. Meditation as a treatment?

Recapping, meditation is considered positive and beneficial for mental well-being, whereas pathological depersonalization is distressing. On the other hand, these states partly resemble each other, and thus, it is interesting to study whether meditation could be used as a treatment for depersonalization. In that case, meditation could be seen as means to modify distressing detachment towards a liberating withdrawal. All possible treatments of depersonalization cannot be considered here, but I want to complete the comparison between meditation and depersonalization by noting that they are different to the extent that one can be considered a treatment for the other. I end this chapter with a brief consideration of this idea of meditation as a treatment.

Generally, empirical research supports the therapeutic potential of meditation for various psychopathologies. For instance, Tang and colleagues (2015, 222) propose that:

> Convergent findings indicate that mindfulness meditation could ameliorate negative outcomes resulting from deficits in self-regulation and could consequently help patient populations suffering from diseases and behavioural abnormalities...the practice of mindfulness meditation might be promising for the treatment of

clinical disorders and might facilitate the cultivation of a healthy mind and increased well-being.

Currently, for instance, mindfulness-based interventions have been shown to be effective in improving a range of clinical conditions, such as current depressive symptoms (Strauss et al., 2014), the risk of relapse for depression (Kuyken et al., 2008; Teasdale et al., 2000), chronic pain (Grossman et al., 2007), stress (Chiesa & Serretti, 2009), and quality of life (Godfrin & van Heeringen, 2010; Kuyken et al., 2008).

Since meditation can assist in coping with distressing experience, it seems worth trying to apply it in treating depersonalization. Meditation could release anxiety by regulating attention and adopting an approving attitude, and this paves the way for regaining mental balance and well-being. Meditation aids in stopping negative mind-wandering and freeing the mind from automatic and unhelpful reactions; it is a change in the manner in which one sees herself. Thus, it seems very relevant for relieving the distress in depersonalization; it does not immediately delete the symptoms, but gives means to overcome the anxiety and to be more open to further steps to find balance in self-consciousness.

One relevant study for the use of meditation in treating depersonalization was conducted by King and colleagues (2016). It studied combat veteran posttraumatic stress disorder (PTSD) patients in 16-week group psychotherapy that involved daily mindfulness training (MBET). PTSD is a highly debilitating disorder affecting 20% of combat veterans, and depersonalization is one of its symptoms. Generally, King and colleagues (2016) found that PTSD patients who completed MBET showed changes in DMN connectivity in resting state and increased DMN connectivity to DLPFC regions within the central executive network. The changes in neural activation were significantly correlated with improvement in avoidant and hyperarousal PTSD symptoms. That is, MBET did have a positive effect on PTSD patients, who suffer from depersonalization.

More theoretically, Sass and colleagues (2013) suggest that it could be useful to develop psychotherapeutic interventions that particularly target the diminished self-affection of

depersonalization, for instance, by enhancing the sense of minimal self. In meditation, the ongoing MSC is in the center while typical RSC weakens, and thus, it might be considered as enhancing the MSC and dampening the RSC. However, since MSC is also altered in the first place in depersonalization, it is not fully clear whether meditation has the effects it typically does, or if it could make the anxiety worse. This is acknowledged in mindfulness guides, which mention that mindful attention is directed to challenging sides of experience, and thus, one should consult professionals if she is currently in therapy or medical care (e.g., Silverton, 2012). One factor that could be elaborated in future studies is individual differences (in general in meditation-based intervention and also in treating depersonalization); is there a form of depersonalization that could benefit from meditation-based interventions particularly, and are there forms of depersonalization that need more careful consideration for the usefulness or potential harm of these interventions? In addition, some depersonalization patients could find meditation satisfactory, while others would find more relief from embodied or social interventions (e.g., Ciaunica et al., 2021, below).

Ciaunica and colleagues (2021, 14) point out that "observation of the similarity between some aspects of meditation and DPD symptomology may have important implications for potential therapy and interventions." Especially, meditation-based interventions could help in overcoming the mentalistic overscrutinization and feelings of being disconnected from one's body and the world that are present in depersonalization. However, Ciaunica and colleagues (2021) assert that meditation lacks "the key dynamic and embodied engagement component which would allow patients to override the static feelings of 'living in a bubble' or in one's head." Instead of traditional meditative practice, Ciaunica and colleagues (2021) recommend development of dynamic, body-based therapies that involve "moving with others" as a treatment for DPD, since this kind of treatment could dissolve the feelings of a fractured self that is isolated from the world. Also analyses in this dissertation support that these kinds of therapies, which acknowledge the multidimensional character of self, are

promising (see also the discussion of the treatment of CS in Sec. 7.5.1.).

Altogether, the theoretical considerations in this chapter support the idea that meditation-based therapies could be useful in treating depersonalization and balancing the fractured self-consciousness. However, the role and effect of meditation as a treatment for depersonalization is an empirical question, which still needs more study.

8.5. Summary

The analyses in this chapter have supported the usefulness of a multidimensional viewpoint on the self. The analyses especially strengthened the ideas that a theory of self-consciousness needs to recognize the connections between MSC and RSC and that both forms of self-consciousness involve several features and variations among them.

The relevance of meditation for self-consciousness studies is obvious in meditation literature that explicitly proposes that meditation aims at alterations in one's self-realization. Meditation is detachment from identification and alters the dynamics of self-consciousness: self-referential evaluative and narrative RSC is superseded by MSC and focus on being present in this moment. In mindfulness meditation, RSC is used to bring objects to consciousness that one then refrains from, striving for a conscious state that is free from identifications. The locus of consciousness is not on its objects but on the concentration of attention to an approving and emotionally calm attitude and on the present moment. In pure consciousness forms of meditation, a meditator can refrain from typical contents of consciousness altogether and reach a bare subjective character of consciousness. In other words, in meditation, one withdraws from her everyday evaluative RSC and reaches a mental state that is focused on the present moment and the presence of subjectivity of MSC. Empirical results support that the practice of meditation can transform the dynamics of self-consciousness. The empirical studies have established that meditation modifies especially the functioning of the default mode

network that is activated in self-related processing; meditation is a way to "switch off" the default approach to self and change it.

Lessons from meditation support the Zahavi-Gallagher-style multidimensional conception of self that has been used in this dissertation. The multidimensional approach emphasizes experience instead of, for instance, neural models or self-conscious thinking alone. Especially, MSC is highlighted in meditation; it is the most fundamental level and a necessary feature of self-consciousness that is involved even in a pure consciousness state. The concept of MSC is useful in explaining meditation because (in the minimalist sense) it can be considered pure subjectivity. On the other hand, the (more robust) concept of MSC can be used as a unit that includes alterations of affectivity and bodily self-consciousness. Thus, meditation research elicits varieties of MSC. In addition, meditation highlights the active modification of RSC, since RSC is needed in getting into and maintaining a meditative state. That is, meditation brings out that the mode and voluntariness are significant features of self-reflection. The specific manner in which RSC is employed can make a decisive difference in self-consciousness, not only the difference in approaching oneself from either a first-personal or third-personal stance.

Further, meditation studies illustrate that the connections holding MSC and RSC together are a key element of self-consciousness. Dynamics between the forms of self-consciousness is crucial for self-experience, and meditation alters this dynamic by shifting the focus from excessive self-evaluation to present-moment bodily self-consciousness and subjectivity. Thus, meditation is a concrete example of a means to volitionally alter the balance in self-consciousness, providing a way to develop flexibility in self-consciousness.

The comparison between meditation and depersonalization highlights these theoretical points. Although both altered states involve an experience of detachment, the experiences spring from very different patterns of self. The patterns differ from each other in terms of the extent of both MSC and RSC and their connections. Roughly, depersonalization involves an excessive connection between diminished MSC and hyper-reflective objectifying RSC,

and thus, the connections are restricted compared to typical ones. Instead, meditation involves detachment from self-evaluation and enhanced MSC and their extensive interconnections that are needed in enabling flexible modulation of different aspects of self-consciousness. Meditation can even be considered as a way to reach balance between different forms of self-consciousness and to cultivate self-consciousness that is free from domination of some particular aspect or pattern of self, since meditators can volitionally direct attention and induce changes in the pattern. Altogether, this chapter revealed that ideas of multidimensionality and subjectivity of self-consciousness are needed to explain the meditative state and the difference between the experiences in meditation and depersonalization. This strengthens these ideas and gives support to the notion that they are needed in formulating a complete theory on self-consciousness. Further, the analysis in this chapter fortifies the conclusion that self-consciousness is an interconnected whole that in practice cannot be fully divided into independent components; rather, it is a holistic entity for which the cooperation of its different parts is crucial.

9. Concluding remarks: self-consciousness as a whole

The nature of self is one of the major questions in the philosophy of mind. In this dissertation, I have studied selfhood in terms of self-consciousness, through which we are familiar with our selves. At the beginning of the dissertation, I proposed that the generic distinction between minimal and reflective self-consciousness is a useful entry point for analyzing self-consciousness. To clarify the different features in the theory of self-consciousness, I posed the following research question: How are the two forms of self-consciousness connected to each other? Although a distinction between them is frequently drawn, their connections have not yet gained the attention they deserve, as the majority of studies have focused only on one form or the other. I sought to answer this research question by clarifying the concepts of self-consciousness, unpacking the subfeatures of both forms, and examining their interconnections. I proposed that a multidimensional approach to self as well as neurophilosophical methods were useful for this examination. A multidimensional pattern theory provided a basis for connecting the different concepts of the self and for highlighting its experiential side. Neurophilosophy kept the analysis empirically informed, offered concrete real-world cases for the study, and set the ground for multidisciplinary cooperation.

Part One of the dissertation focused on laying the conceptual basis for the analysis of self-consciousness. I argued that to understand self-consciousness and the connections between its two forms, having a detailed picture of the involved components is essential. In Chapter 2, I analyzed MSC from a phenomenological viewpoint, in which MSC is a constitutive feature of all conscious states. MSC does not refer to the contents ("what") of consciousness but to the subjective manner ("how") of experiencing. I proposed that MSC is the most elementary form of self-consciousness but it involves variety, as was evident from the comparison between minimalist and robust readings of MSC. In the robust sense, MSC involves several aspects of self, including at least experiential, embodied, and affective aspects, and can be considered a

subjective perspective. MSC is based on non-conscious embodied aspects of self; however, some bodily features are typically also present in our experiences, of which they are critical parts. In addition, MSC often involves affective aspects of the self and the sense of agency. These features are typically present in MSC, but not strictly necessary for it since they may be lacking in some states of consciousness. In the minimalist reading, MSC is subjectivity or the subjective character of consciousness, which is always present in experience and can be associated with the experiential aspect of the self. I proposed that the experiential aspect is special in pattern theory since it is necessary in a way that other aspects are not.

In Chapter 3, I analyzed RSC, which is the capacity to take oneself the object of one's cognition and to think of oneself as oneself. Through RSC, one can act according to reason as well as to weave a complex narrative of oneself. RSC has often been discussed in terms of the distinction between first-personal and third-personal reference to oneself: self-conscious thoughts have epistemically and motivationally unique features that distinguish them from all other thoughts. I argued that to recognize the wide range of variants of self-reflection, the distinction between first- and third-personal self-reference is not sufficient and the concepts describing RSC must be elaborated. First, I highlighted that the distinction between two stances toward oneself, namely deliberative and theoretical, is useful in the examination of the forms of self-reflection. From the deliberative stance, the self is experienced as an active agent of practical reasoning that makes decisions and shapes its attitudes, whereas from the theoretical stance, the self is considered an object of description and explanation. Second, RSC is typically a voluntary capacity through which one can consider and direct one's actions. However, RSC sometimes operates involuntarily, as in the pathological cases discussed in this dissertation. Third, the extent or mode of identification with self-reflection varies: one can feel a strong identification with a self-conscious thought or withdraw oneself from it. In addition, I highlighted that the manner of self-reflection is connected to mental well-being. In particular, the deliberative

stance and the capacity to flexibly and voluntarily shift between the stances of self-reflection are crucial for well-being.

In Chapter 4, I began to examine the amalgamation of the two forms of self-consciousness and the application of neurophilosophy. Self-consciousness refers to personal-level experiences, whereas the results of neuroscience are placed at the subpersonal level underlying experience. I proposed that to keep the difference between these levels in mind, it is useful to elaborate a multidisciplinary conception of selfhood in which fine-grained concepts of self-consciousness and data from cognitive and neurophysiological processes can enrich each other. The elaborated concepts of self are required in the interpretation of results and the development of a paradigm in empirical sciences studying the self. On the other hand, empirical studies are interesting for philosophers since they can elucidate the mechanism of self-consciousness. I proposed that the distinction between minimal and reflective self-consciousness is empirically relevant, since it can assist in grasping the wide-scale neural functions involved in processing the self. MSC can be connected to self-specific neural processing, which identifies the self as the subject or agent and can be associated with sensory-motor processes and cognitive–affective agency. By contrast, RSC can be connected to self-related neural processing that employs the E-network, which is the default network of the brain involving the frontal and temporal cortical areas. In addition, the overlapping of these processes is evident in neural studies, and future studies would benefit from a more detailed theoretical picture of the links within self-consciousness.

The major finding of this dissertation is that the interconnectedness between the two forms is crucial for the apprehension of self-consciousness. This was determined in Chapter 5. First, as a constitutive feature of experience, MSC is a constituent of RSC: whenever the self is the object of reflective consciousness, it is simultaneously the minimally self-conscious subject of reflection. This is significant since it implies that the study of RSC cannot be wholly independent of the study of MSC; that is, a full understanding of RSC requires the consideration of MSC. In

addition, this constitutive character of MSC is significant for theories of consciousness and undermines anonymity theories of phenomenal consciousness, which deny the existence of MSC. Second, I analyzed the relations between MSC and RSC. This viewpoint has been dismissed numerous times in the study of self, whereas the first point regarding the fundamental character of MSC is made more often. A simple picture includes bottom-up and top-down relations in self-consciousness: a change in one form of self-consciousness can induce changes in the other. In the bottom-up direction, a minimally self-conscious experience triggers a closer scrutiny of RSC. In the top-down direction, in turn, self-related cognition in RSC is used to regulate or modify MSC. I argued that in addition to this simple picture of the relations, the dynamics of the two forms of self-consciousness should be elaborated to ensure a more extensive picture. In everyday life, the forms amalgamate into a whole: the top-down and bottom-up processes occur in loops and form a diverse network of relations between MSC and RSC. Alterations in this network induce alterations in self-experience. This indicates that a mere list of aspects is insufficient for understanding self-consciousness; rather, it is crucial to examine the dynamicity in the interactions of these aspects. Furthermore, I argued that connections in self-consciousness are significant for mental well-being. Flexible connections and a balance between the two forms of self-consciousness are connected to well-being, whereas mental disorders involve imbalance and distorted relations within self-consciousness.

Part Two of the dissertation applied the ideas elaborated in Part One to real-word cases of altered self-consciousness. Altered states of consciousness were used to clarify the dimensions of self, since they offer a contrast condition that can elicit the tacit structure of typical self-consciousness. Chapters 6 and 7 focused on the altered self-experience in the pathologies of Cotard syndrome and depersonalization. Depersonalization refers to alienating feelings of estrangement from one's self, one's body, and/or the world. Cotard syndrome is characterized by the same kind of symptoms but is a more severe condition, in which patients actually claim that they do not have body parts, mind, or self, and may even say that they are

dead. At first glance, the experiences of patients with Cotard syndrome seem to be a counterexample to theories that consider self-consciousness a necessary feature of experience. However, I proposed that it is more fruitful to ask which kind of self-consciousness is lacking in the syndrome. I started my analysis by studying previous explanations that have indicated that Cotard syndrome involves dysfunctions in several aspects of self. The explanations that recognize MSC were the most promising since they indicate that highly abnormal first-order (minimal) self-experience gives rise to (reflectively) self-conscious beliefs about non-existence in CS. However, I argued that all previous accounts have only provided a partial explanation since they have ignored the connections between minimal and reflective self-consciousness. I indicated the significance of these connections in two ways: the vicious circle involving parts of them and the decreased connectivity in the normally wide connections between them. Overall, Cotard syndrome involves diminished minimal self-consciousness, reflective self-consciousness in 3P mode, and the distorted coupling of the two. Together, these factors form a distorted pattern of self. Moreover, I highlighted that this analysis of CS indicates that a theory of self-consciousness must acknowledge MSC and the dynamics between MSC and RSC. In empirical fields, this invites an extensive psychiatric perspective on mental disorders and the targeting of patients' overall experiential orientation.

In Chapter 8, I discussed meditation, which is an altered state of consciousness aimed at producing changes in self-consciousness and is associated with improvements in mental well-being. Meditation is characterized by diminished evaluative self-reflection and calm, centered minimal self-consciousness. I proposed that meditation involves a specific manner of using RSC to induce the enhancement of MSC: MSC becomes more prevalent due to this interplay between the two forms of self-consciousness. Furthermore, I deepened the analysis of the dynamics of self-consciousness by comparing altered self-consciousness in meditation and depersonalization. Noteworthily, these states share de-identification from typical self-experience, especially in terms of the diminution of affectivity and the holding of a third-personal

stance toward oneself. However, I argued that the states also significantly differ from each other: a meditator actively induces herself into a calm meditative state and the withdrawal of normal self-consciousness. By contrast, a patient with depersonalization disorder suffers from an involuntary feeling of diminished self-affection and, due to hyper-reflectivity, she is captured in the processing of her anomalous self-experience. The analysis of these two cases of altered consciousness strengthened the general conclusions of self-consciousness presented in this dissertation. First, minimal self-consciousness is the most fundamental form of self-consciousness, but it can involve variance in terms of its affective and embodied features and the sense of agency. Second, reflective self-consciousness involves variance in terms of the first- or third-personal stance of reflection, voluntariness, and the mode of identification. Third, minimal and reflective self-consciousness form a whole, in which their connections have a crucial role.

In general, the examination demonstrated the validity of the multidimensional and conceptual framework of self that was introduced at the beginning of this dissertation. This framework is beneficial for analyzing self-consciousness since it is sufficiently fine-grained to account for even exceptional cases. Since the framework manages to embrace the whole phenomenon, its use and elaboration are encouraged in further studies. Future studies of self-consciousness offer several interesting research questions that cannot be considered here in their entirety, but include at least the following general themes:

An obvious research subject is the formation of a complete picture of self-consciousness. This dissertation indicated the links within self-consciousness by focusing on a few basic aspects of self; however, other aspects certainly also contribute to self-consciousness, including behavioral, interpersonal, and extended/situational aspects. These aspects are especially interesting since they seem relevant for both forms of self-consciousness; thus, studying them is likely to assist in revealing further relations within self-consciousness. For instance, extended aspects can be considered significant for MSC when a device outside of the biological body is incorporated into the bodily

identity and experience of a self (see, e.g., Thompson & Stapleton, 2009); a blind person's cane is a classic example of this kind of incorporation. RSC can be considered to "extend" in another sense, such as in the classic example of a notebook (Clark & Chalmers, 1998), which functions as a memory. In future studies, the relations in self-consciousness in an "extended self" could be targeted in more detail; for instance, a blind person's cane is also important in her self-narratives.

Another obvious research line is to continue to use altered states of consciousness as a methodological tool for studying the self, which was fruitful for detecting the subtleties of self-consciousness in this dissertation. For instance, on a very broad scale, future studies could address varieties of self-consciousness in different forms of meditation, dreaming, and mind-wandering, and also how the pattern of self-consciousness changes in these states. Related to altered states of consciousness, an obvious useful area of application of the study of self-consciousness is pathologically altered self-experience and the philosophy of psychiatry. It would be interesting to compare the similarities and differences between altered states of consciousness (including to what extent pathological cases resemble non-pathological ASCs), and to determine what these similarities or differences reveal about the structures of self.

Concerning the theoretical frameworks in the philosophy of mind, it would be interesting to apply the phenomenologically oriented neurophilosophy of self used here to other recent trends. These trends include enactivism and the 4E approach to cognition, which seem to share the multidimensional conception of self. The analytical framework of this dissertation could assist in conceptual analysis, which is still required in these novel approaches, and the enactivist or 4E perspective could enrich the neurophilosophical picture of the self. In addition, the relationship between phenomenologically oriented neurophilosophy and predictive processing theory requires elaboration. Specifically, studies could address the following question: Are they complementary or does predictive modeling lack some experiential layers of the self?

Furthermore, neurophilosophy promotes a multidisciplinary application of the concepts of self-consciousness. In terms of

neurosciences, this means a fine-grained conceptual repertoire for describing the neural processes associated with the self. For instance, this repertoire involves MSC, differences in the mode of RSC, and the interplay between the two forms. One important research topic concerns self-consciousness and mental well-being, which can be studied in the fields of psychology and psychiatry. Generally, the analyses in this dissertation have endorsed the multidimensional picture of the self which brings out the dynamics in self-consciousness. This means, for instance, that the treatment of mental disorders should not be restricted to medicine but can also involve therapies aimed at changes in the experiential focus, such as in meditation. This picture can be extended by including embodied, social, and environmental dimensions of the self, such as physical exercise and intersubjective communication. In addition, the dynamics of self-consciousness have a major role in psychological well-being in non-pathological cases. Thus, the development of a model of balanced self-consciousness is a significant objective, not only for theoretical interest but also for practical purposes to increase mental well-being.

Altogether, this dissertation has explored self-consciousness to contribute to a more complete picture of the human mind. Drawing this picture is challenging since self-consciousness is colored by many shades that go beyond the mere contents of consciousness, and these shades can vary considerably between people and situations. The first step in the analysis was to draw a distinction between minimal and reflective self-consciousness, following by recognizing the different aspects of self that they involve. However, a mere list of components is not sufficient—we also need to examine their dynamics in accounting for self-consciousness. Based on the examinations in this dissertation, I have defended the view that the connections between the minimal and reflective forms are crucial for self-consciousness. Without noticing these connections, cases of self-consciousness remain only partially explained. When the complexity of self is embraced and delineated into fine-grained concepts, the concepts can be used to elaborate more exhaustive philosophical theories and more detailed research of the self in mind sciences.

Epilog: What did Sofia find out in the forest?

We began the quest for the shades of self-consciousness with Sofia, who wandered in the forest and had vivid experiences of entering deeper in the woods and in her self. Now we can see that while walking in the forest, Sofia lived through an extensive journey into her self-consciousness. The spectrum of her self-experience ranged from fundamental minimal self-consciousness to more sophisticated reflective forms of self-consciousness and to different mixtures of these two.[105] Although Sofia mostly was fully immersed in the vivid forest environment and without any self-reflection, she was minimally self-conscious all the time simply by being conscious. She was undergoing experiences essentially characterized by 1PP or subjectivity; her self was all the time present as a subject of experience, and there was a variety of what-it-is-likeness in her experience. This minimal self-consciousness involved immediate sensations such as seeing the bright colors of nature, hearing the sounds of wind and animals, and scenting the flowers and leaves. Further, her minimal self-consciousness also importantly included bodily and affective features: Sofia was moving up and down along the shape of the terrain, feeling the warmth of the sun on her skin, and experiencing the joy of being there and then. When Sofia came to the pond and leaned to touch the water, she used her hand to examine the temperature of the water but still, her body was not an object of her consciousness but a part of an experience through which she was sensing.

---

[105] Albeit, following the general scope of this dissertation, it can be noted that some aspects of self are not elaborated here. During the walking episode Sofia is alone, which excludes intersubjective aspects. In addition, extended aspects of self are rather left out of the story because Sofia did not pay especial attention to her material possessions or used tools. Further, neither situational aspects of self are dealt with in detail but significance of environment in general is included in experiential aspects, and some cultural practices can be seen in Sofia's memories invoked by narrative aspects.

During her outing in the forest, Sofia did also undergo episodes of reflective self-consciousness. Importantly, her minimal self-consciousness was a constituent in these episodes too, and her first-order experience triggered the self-conscious thoughts. In reflective self-consciousness, by means of psychological-cognitive and reflective aspects of self, Sofia took herself as the object of her cognition and was thinking about herself as herself. When she remembered herself as a child and recalled events, atmosphere and her moods back then and felt the distance of those years, Sofia was also exercising the narrative aspects of self. In other words, in addition to plain immediate experience, she viewed herself as a narrative entity with a wider time perspective: she framed the narrative structure of her life going back to her early childhood and reaching forward to her plans for the future. The affectivity in her memories, for instance, the warmth with which she was thinking about her grandmother, also indicates how aspects of minimal self-consciousness are embraced in reflective self-consciousness.

When Sofia underwent the short episode of depersonalization, the functioning of her self-consciousness altered strikingly. Her minimal self-consciousness was diminished; the typical self-affection was missing for a moment and replaced by a sudden feeling of unfamiliarity. At the same time, the mode of Sofia's reflective self-consciousness changed; Sofia lost feeling herself as an active agent, an 'I' that directs action. By contrast, Sofia felt as if she were looking at herself from outside, in a third-person perspective, as somebody else or as an inanimate object. Sofia did not identify with the mirror image in the surface of the pond, but the reflection on the water could have been an image of someone else. Thus, during the transient period of depersonalization Sofia's self-consciousness was altered both at minimal and reflective levels. In addition, the dynamics of these levels altered; Sofia's self-reflection was captured by the striking, alienated, minimally self-conscious feelings, but she was unable to change those feelings by self-reflection. While being overcome by the alienated feelings, the normal balanced coupling of minimal and reflective form turned anomalous. Although the episode of depersonalization was brief, it had a strong effect on Sofia; the sudden breakdown of everyday self-consciousness made

her meditate on the complexity of being a self and on what it is to be a person.

Altogether, Sofia found not only blueberries but also many aspects of her self during the walk in the forest, and her experiences offer an illustration of the many shades of self-consciousness. It is important to note that Sofia's self-consciousness did not disappear at any point, although it went through many forms and was even altered. Minimal self-consciousness is something into which she is tied in her every experience, whereas reflective self-consciousness gives her further capacities to deliberate on and direct the course of her actions. It is also remarkable how smoothly the different forms of self-consciousness followed each other in Sofia's mind. This brings out the significance of the connections between minimal and reflective self-consciousness and flexible modulation of them. Minimal and reflective self-consciousness are essentially connected together and form a unity. This unity of self-consciousness is of utmost importance since the balance between its aspects is crucial for mental well-being. Although Sofia lost the balance for a moment, the balance was also quickly restored. The search for this balance in mental well-being and the ubiquity of self-consciousness provides reasons for examining self-consciousness in future studies as well. Furthermore, it shows why the shades that self-consciousness takes matter.

References

Adams, F. & Aizawa, K. (2001). The bounds of cognition. *Philosophical Psychology*, *14*(1), 43–64. https://doi.org/10.1080/09515080120033571

Adams, F, & Aizawa, K. (2010). Defending the Bounds of Cognition. In R. Menary (Ed.), *The Extended Mind* (pp. 67–80). The MIT Press.

Adams, F. & Aizawa, K. (2010). *The Bounds of Cognition*. Wiley-Blackwell. https://doi.org/10.1002/9781444391718

Ainley, V., & Tsakiris, M. (2013). Body Conscious? Interoceptive Awareness, Measured by Heartbeat Perception, Is Negatively Correlated with Self-Objectification. *PLoS ONE*, *8*(2), e55568. https://doi.org/10.1371/journal.pone.0055568

Albahari, M. (2006). *Analytical Buddhism: The Two-tiered Illusion of Self*. Palgrave Macmillan UK. https://doi.org/10.1057/9780230800540

Albahari, M. (2009). Witness-Consciousness: Its Definition, Appearance and Reality. *Journal of Consciousness Studies*, *16*(1), 62–84.

Albahari, M. (2010). Nirvana and Ownerless Consciousness. In M. Siderits, E. Thompson, & D. Zahavi (Eds.), *Self, No Self?: Perspectives from Analytical, Phenomenological, and Indian Traditions* (pp. 1–57). Oxford University Press.

Amsterdam, B. (1972). Mirror self-image reactions before age two. *Developmental Psychobiology*, *5*(4), 297–305. https://doi.org/10.1002/dev.420050403

Anscombe, G. E. M. (1975). The First Person. In S. D. Guttenplan (Ed.), *Mind and Language* (pp. 45–65). Oxford University Press.

Araujo, H. F., Kaplan, J., & Damasio, A. (2013). Cortical Midline Structures and Autobiographical-Self Processes: An Activation-Likelihood Estimation Meta-Analysis. *Frontiers in Human Neuroscience*, *7*.

Araujo, H. F., Kaplan, J., Damasio, H., & Damasio, A. (2015). Neural correlates of different self domains. *Brain and Behavior*, *5*(12), 1–5. https://doi.org/10.1002/brb3.409

Arpaly, N., & Schroeder, T. (1999). Praise, Blame and the Whole Self. *Philosophical Studies*, *93*(2), 161–188.

Austin, J. H. (1999). *Zen and the brain. Toward an understanding of meditation and consciousness*. MIT Press.

Baker, Hunter, Lawrence, David, Medford, & Sierra. (2012). Depersonalization and Detachment from One's Body. *Emotion Review Radovic & Radovic Sierra Sierra & Berrios*, *4*(2), 145–150. https://doi.org/10.1177/1754073911430131

Barandiaran, X. E., Di Paolo, E., & Rohde, M. (2009). Defining agency: Individuality, normativity, asymmetry, and spatio-temporality in action. *Adaptive Behavior*, *17*(5), 367–386.

Barresi, J., & Martin, R. (2011). History as Prologue: Western Theories of the Self. In S Gallagher (Ed.), *The Oxford Handbook of the Self*. Oxford University Press.

Barwise, J., & Perry, J. (1981). Situations and Attitudes. *The Journal of Philosophy*, *78*(11), 668. https://doi.org/10.2307/2026578

Bayne, T. (2004). Self-consciousness and the unity of consciousness. *The Monist*, *87*(2), 219–236.

Bayne, T., & Pacherie, E. (2007). Narrators and comparators: the architecture of agentive self-awareness. *Synthese*, *159*(3), 475–491. https://doi.org/10.1007/s11229-007-9239-9

Beaney, M. (2014). Analysis. In Edward N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy* (Summer 2021 ed.). https://plato.stanford.edu/archives/sum2021/entries/analysis/

Bechtel, William, Mandik, P., & Mundale, J. (2001). Philosophy Meets the Neurosciences. In W. Bechtel, P. Mandik, J. Mundale, & R. S. Stufflebeam (Eds.), *Philosophy and the neurosciences: A reader* (pp. 4–22). Basil Blackwell.

Belt, J. (2019). Between Minimal Self and Narrative Self: A Husserlian Analysis of Person. *Journal of the British Society for Phenomenology*, *50*(4), 305–323.

Beni, M. D. (2016). Structural realist account of the self. *Synthese*, *193*(12), 3727–3740. https://doi.org/10.1007/S11229-016-1098-9

Bennett, M. R., & Hacker, P. M. S. (2003). *Philosophical foundations of neuroscience*. Blackwell.

Berkman, E. T., & Lieberman, M. D. (2009). Using Neuroscience to Broaden Emotion Regulation: Theoretical and Methodological Considerations. *Social and Personality Psychology Compass*, *3*(4), 475–493.

Berkovich-Ohana, A. (2017). A case study of a meditation-induced altered state: increased overall gamma synchronization. *Phenomenology and the Cognitive Sciences*, *16*(1), 91–106. https://doi.org/10.1007/s11097-015-9435-x

Berkovich-Ohana, A., Dor-Ziderman, Y., Trautwein, F.-M., Schweitzer, Y., Nave, O., Fulder, S., & Ataria, Y. (2020). The Hitchhiker's Guide to Neurophenomenology – The Case of Studying Self Boundaries With Meditators. *Frontiers in Psychology*, *11*.

Bermúdez, J. L. (2001). Nonconceptual self-consciousness and cognitive science. *Synthese*, *129*(1), 129–149.

Bermúdez, J. L. (2007). Self-Consciousness. In S. Velmans, Max & Scheineider (Ed.), *The Blackwell Companion to Consciousness* (pp. 456–467). Blackwell.

Bermúdez, J. L. (2014). *Cognitive Science* (2nd ed.). Cambridge University Press.

Berrios, G. E. (1991). Delusions as "Wrong Beliefs": A Conceptual History. *British Journal of Psychiatry*, *159*(S14), 6–13.

Berrios, G. E., & Luque, R. (1995a). Cotard's delusion or syndrome?: A conceptual history. *Comprehensive Psychiatry*, *36*(3), 218–223.

Berrios, G. E., & Luque, R. (1995b). Cotard's syndrome: analysis of 100 cases. *Acta Psychiatrica Scandinavica*, *91*(3), 185–188. https://doi.org/10.1111/j.1600-0447.1995.tb09764.x

Berrios, G. E., & Luque, R. (1999). Cotard's "On hypochondriacal delusions in a severe form of anxious melancholia." *History of Psychiatry*, *10*(38), 269–274.

Bickle, J. (Ed.). (2009). *The Oxford Handbook of Philosophy and Neuroscience* (Vol. 1). Oxford University Press.

Bilgrami, A. (2006). *Self-Knowledge and Resentment*. Harvard University Press.

Billon, A. (2016). Making Sense of the Cotard Syndrome: Insights from the Study of Depersonalisation. *Mind and Language*, *31*(3), 356–391. https://doi.org/10.1111/mila.12110

Billon, A. (2017). Basic Self-Awareness: Lessons from the Real World. *European Journal of Philosophy*, *25*(3), 732–763. https://doi.org/10.1111/ejop.12168

Billon, A., & Kriegel, U. (2015). Jaspers' Dilemma: The Psychopathological Challenge to Subjectivity Theories of Consciousness. In R. J. Gennaro (Ed.), *Disturbed Consciousness* (pp. 29–54). The MIT Press.

Blanke, O., & Metzinger, T. (2009). Full-body illusions and minimal phenomenal selfhood. *Trends in Cognitive Sciences*, *13*(1), 7–13.

Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences*, *18*(2), 227–287.

Block, N. (2009). Comparing the major theories of consciousness. In M. S. Gazzaniga & E. Al. (Eds.), *The cognitive neurosciences* (pp. 1111–1122). MIT Press.

Block, N. (2011). The higher order approach to consciousness is defunct. *Analysis*, *71*(3), 419–431. https://doi.org/10.1093/ANALYS/ANR037

Bortolan, A. (2020). Affectivity and the distinction between minimal and narrative self. *Continental Philosophy Review*, *53*(1), 67–84.

Bortolotti, L. (2009). *Delusions and Other Irrational Beliefs*. Oxford University

Bortolotti, L. (2018). Delusion. In Edward N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy* (Spring 2018 ed.). https://plato.stanford.edu/archives/spr2018/entries/delusion/

Boyle, M. (2009). Two Kinds of Self-Knowledge. *Philosophy and Phenomenological Research*, *78*(1), 133–164. https://doi.org/10.1111/j.1933-1592.2008.00235.x

Brefczynski-Lewis, J. A., Lutz, A., Schaefer, H. S., Levinson, D. B., & Davidson, R. J. (2007). Neural correlates of attentional expertise in long-term meditation practitioners. *Proceedings of the National Academy of Sciences*, *104*(27), 11483–11488. https://doi.org/10.1073/pnas.0606552104

Brewer, J. A., Worhunsky, P. D., Gray, J. R., Tang, Y. Y., Weber, J., & Kober, H. (2011). Meditation experience is associated with differences in default mode network activity and connectivity. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(50), 20254–20259. https://doi.org/10.1073/pnas.1112029108

Broesch, T., Callaghan, T., Henrich, J., Murphy, C., & Rochat, P. (2011). Cultural Variations in Children's Mirror Self-Recognition. *Journal of Cross-Cultural Psychology*, *42*(6), 1018–1029. https://doi.org/10.1177/0022022110381114

Brook, A. (2001). Kant, self-awareness and self-reference. In R. Brook, A. & DeVidi (Ed.), *Self-reference and self-awareness* (pp. 9–30). John Benjamins.

Brook, A., & DeVidi, R. C. (Eds.). (2001). *Self-Reference and Self-Awareness*. John Benjamins Publishing Company. https://doi.org/10.1075/aicr.30

Brown, R., Lau, H., & LeDoux, J. E. (2019). Understanding the Higher-Order Approach to Consciousness. *Trends in Cognitive Sciences*, *23*(9), 754–768. https://doi.org/10.1016/j.tics.2019.06.009

Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The Brain's Default Network. *Annals of the New York Academy of Sciences*, *1124*(1), 1–38. https://doi.org/10.1196/annals.1440.011

Burge, T. (1996). Our Entitlement to Self-Knowledge. *Proceedings of the Aristotelian Society*, *96*(1), 91–116. https://doi.org/10.1093/aristotelian/96.1.91

Butterworth, G. (1998). A developmental-ecological perspective on Strawson's 'The self.' *Journal of Consciousness Studies*, *5*(2), 132–40.

Byrne, A. (2005). Introspection. *Philosophical Topics*, *33*(1), 79–104.

Campbell, J. (2001). Rationality, Meaning, and the Analysis of Delusion. *Philosophy, Psychiatry & Psychology*, *8*(2), 89–100.

Carhart-Harris, R. L., Erritzoe, D., Williams, T., Stone, J. M., Reed, L. J., Colasanti, A., Tyacke, R. J., Leech, R., Malizia, A. L., Murphy, K., Hobden, P., Evans, J., Feilding, A., Wise, R. G., & Nutt, D. J. (2012). Neural correlates of the psychedelic state as determined by fMRI studies with psilocybin. *Proceedings of the National Academy of Sciences*, *109*(6), 2138–2143.

Carruthers, P. (1996). *Language, Thoughts and Consciousness: An Essay in Philosophical Psychology*. Cambridge University Press.

Castañeda, H.-N. (1966). 'He': A Study in the Logic of Self-Consciousness. *Ratio*, *8*, 130–157.

Castañeda, H.-N. (1967). Indicators and Quasi-Indicators. *American Philosophical Quarterly*, *4*, 85–100.

Castillo, R. (1990). Depersonalisation and Meditation. *Psychiatry*, *53*(2), 158–168.

Chalmers, D. J. (1996). *The conscious mind*. Oxford University Press.

Chalmers, D. J. (2007). Consciousness and its Place in Nature. In S. P. Stich & T. A. Warfield (Eds.), *The Blackwell Guide to Philosophy of Mind* (pp. 102–142). Blackwell. https://doi.org/10.1002/9780470998762.ch5

Charland-Verville, V., Bruno, M. A., Bahri, M. A., Demertzi, A., Desseilles, M., Chatelle, C., Vanhaudenhuyse, A., Hustinx, R., Bernard, C., Tshibanda, L., Laureys, S., & Zeman, A. (2013). Brain dead yet mind alive: A positron emission tomography case study of brain metabolism in cotard's syndrome. In *Cortex*. https://doi.org/10.1016/j.cortex.2013.03.003

Chiesa, A., & Serretti, A. (2009). Mindfulness-Based Stress Reduction for Stress Management in Healthy People: A Review and Meta-Analysis. *The Journal of Alternative and Complementary Medicine*, *15*(5), 593–600. '

Christoff, K., Cosmelli, D., Legrand, D., & Thompson, E. (2011a). Specifying the self for cognitive neuroscience. In *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2011.01.001

Christoff, K., Cosmelli, D., Legrand, D., & Thompson, E. (2011b). Clarifying the self: Response to Northoff. *Trends in Cognitive Sciences*, *15*(5), 187–188. https://doi.org/10.1016/j.tics.2011.03.004

Christoff, K., & Fox, K. C. R. (Eds.). (2018). *The Oxford Handbook of Spontaneous Thought*. Oxford University Press.

Churchland, P. S. (1986). *Neurophilosophy: Toward a unified science of the mind-brain*. MIT Press.

Ciaunica, A. (2015). Basic forms of pre-reflective self-consciousness: A developmental perspective. In S. Miguens, G. Preyer, & C. Morando (Eds.), *Pre-reflective Consciousness: Sartre and Contemporary Philosophy of Mind* (pp. 422–438). Routledge.

Ciaunica, A. (2019). The ' Meeting of Bodies ' – Empathy and Basic Forms of Shared Experiences. *Topoi*, *38*, 185–195.

Ciaunica, A., Charlton, J., & Farmer, H. (2021). When the Window Cracks: Transparency and the Fractured Self in Depersonalisation. *Phenomenology and the Cognitive Sciences*, *20*(1). https://doi.org/10.1007/s11097-020-09677-z

Ciaunica, A., & Crucianelli, L. (2019). Minimal self-awareness from within: A developmental perspective. *Journal of Consciousness Studies*, *26*(3–4), 207–226.

Ciaunica, A., & Fotopoulou, A. (2017). The touched self: Psychological and philosophical perspectives on proximal intersubjectivity and the self. *Embodiment, Enaction, and Culture: Investigating the Constitution of the Shared World*, 173–192. https://doi.org/10.7551/mitpress/9780262035552.003.0009

Clark, A., & Chalmers, D. (1998). The Extended Mind. *Analysis*, *58*(1), 7–19. https://doi.org/10.1093/analys/58.1.7

Clark, A. (2003). *Natural-Born Cyborgs – Minds, Technologies, and the Future of Human Intelligence*. Oxford University Press.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204.

Clark, A. (2016). *Surfing Uncertainty*. Oxford University Press.

Cole, J. (1997). On "being faceless": selfhood and facial embodiment. *Journal of Consciousness Studies*, *4*(5–6), 467–84.

Colombetti, G. (2007). Enactive appraisal. *Phenomenology and the Cognitive Sciences*, *6*(4), 527–546. https://doi.org/10.1007/s11097-007-9077-8

Colombetti, G. (2011). Varieties of pre-reflective self-awareness: Foreground and background bodily feelings in emotion experience. *Inquiry*, *54*(3), 293–313. https://doi.org/10.1080/0020174X.2011.575003

Colombetti, G., & Ratcliffe, M. (2012). Bodily feeling in depersonalization: A phenomenological account. In *Emotion Review* (Vol. 4, Issue 2, pp. 145–150). https://doi.org/10.1177/1754073911430131

Colombetti, G., & Thompson, E. (2008). The feeling body: Toward an enactive approach to emotion. In W. Overton, U. Muller, & J. Newman (Eds.), *Developmental Perspectives on Embodiment and Consciousness* (pp. 45–68). Lawrence Erlbaum. https://doi.org/10.4324/9780203809778

Cotard, J. (1882). Nihilistic delusions. In S. R. Hirsh & M. Shepherd (Eds.), *Themes and variations in European psychiatry. An Anthology (1974)*. John Wright & Sons LTD.

Craig, A. D. (2009). How do you feel — now? The anterior insula and human awareness. *Nature Reviews Neuroscience*, *10*(1), 59–70. https://doi.org/10.1038/nrn2555

Craig, A. D. (2010). The sentient self. *Brain Structure & Function*, *214*(5–6), 563–577. https://doi.org/10.1007/s00429-010-0248-y

Critchley, M. (1964). Psychiatric symptoms and parietal disease: differential diagnosis. *Proc R Soc Med*, *57*, 422–428.

Dahl, C. J., Lutz, A., & Davidson, R. J. (2015). Reconstructing and deconstructing the self: cognitive mechanisms in meditation practice. *Trends in Cognitive Sciences*, *19*(9), 515–523. https://doi.org/10.1016/j.tics.2015.07.001

Daly, A., & Gallagher, S. (2019). Towards a Phenomenology of Self-Patterns in Psychopathological Diagnosis and Therapy. *Psychopathology*, *52*(1), 33–49. https://doi.org/10.1159/000499315

Damasio, A. (1999). *The feeling of what happens: Body and emotion in the making of consciousness*. Harcourt College Publishers.

Dambrun, M. (2016). When the dissolution of perceived body boundaries elicits happiness: The effect of selflessness induced by a body scan meditation. *Consciousness and Cognition*, *46*, 89–98. https://doi.org/10.1016/j.concog.2016.09.013

Debruyne, H., Portzky, M., Peremans, K., & Audenaert, K. (2011). Cotard 's Syndrome. *Mind & Brain, The Journal of Psychiatry*, *2*(1), 67–72.

Debruyne, H., Portzky, M., Van Den Eynde, F., & Audenaert, K. (2009). Cotard's Syndrome: A Review. *Current Psychiatry Reports Current Medicine Group LLC ISSN*, *11*, 197–202.

Dehaene, Stanislas, Changeux, J.-P., & Naccache, L. (2011). The Global Neuronal Workspace Model of Conscious Access: From Neuronal Architectures to Clinical Applications. In S. Dehaene & Y. Christen (Eds.), *Characterizing Consciousness: From Cognition to the Clinic?. Research and Perspectives in Neurosciences* (pp. 55–84). Springer.

Dennett, D. (1991). *Consciousness Explained*. Little, Brown.

Dennett, D. (1992). The self as a center of narrative gravity. In F. Kessel, P. Cole, & D. Johnson (Eds.), *Self and consciousness: Multiple perspectives*. Psychology Press

Denny, B. T., Kober, H., Wager, T. D., & Ochsner, K. N. (2012). A Meta-analysis of Functional Neuroimaging Studies of Self- and Other Judgments Reveals a Spatial Gradient for Mentalizing in Medial Prefrontal Cortex. *Journal of Cognitive Neuroscience*, *24*(8), 1742–1752. https://doi.org/10.1162/jocn_a_00233

Di Paolo, E. (2009). Extended Life. *Topoi*, *28*(1), 9–21.

Di Paolo, E., & Thompson, E. (2014). The Enactive Approach. In L. Shapiro (Ed.), *The Routledge Handbook of Embodied Cognition*. Routledge Press. https://doi.org/10.31231/osf.io/3vraf

Dörfel, D., Lamke, J.-P., Hummel, F., Wagner, U., Erk, S., & Walter, H. (2014). Common and differential neural networks of emotion regulation by

Detachment, Reinterpretation, Distraction, and Expressive Suppression: A comparative fMRI investigation. *NeuroImage*, *101*, 298–309.

Dretske, F. (1999). The Mind's Awareness of Itself. *Philosophical Studies*, *95*(1), 103–124. https://doi.org/10.1023/A:1004515508042

Dretske, F. I. (1995). *Naturalizing the mind*. MIT Press.

Dugas, L. (1898). Un cas de dépersonnalisation. *Revue Philosophique de La France et de l'Étranger*, *45*, 500–507.

Ebisch, S. J. H., & Aleman, A. (2016). The fragmented self: imbalance between intrinsic and extrinsic self-networks in psychotic disorders. In *The Lancet Psychiatry* (Vol. 3, Issue 8, pp. 784–790). Elsevier Ltd. https://doi.org/10.1016/S2215-0366(16)00045-6

Enoch, M. D., & Ball, H. N. (2001). *Uncommon psychiatric syndromes* (4th ed.). Arnold.

Enoch, M. D., & Trethowan, S. W. H. (1991). *Uncommon psychiatric syndromes* (3rd ed.). Buttenworh/Heinemann.

Eronen, M. I., & Bringmann, L. F. (2021). The Theory Crisis in Psychology: How to Move Forward. *Perspectives on Psychological Science*, *16*(4), 779–788. https://doi.org/10.1177/1745691620970586

Farb, N. A. S., Segal, Z. V., Mayberg, H., Bean, J., Mckeon, D., Fatima, Z., & Anderson, A. K. (2007). Attending to the present: Mindfulness meditation reveals distinct neural modes of self-reference. *Social Cognitive and Affective Neuroscience*, *2*(4), 313–322. https://doi.org/10.1093/scan/nsm030

Farrer, C., Franck, N., Frith, C. D., Decety, J., Georgieff, N., D'Amato, T., & Jeannerod, M. (2004). Neural correlates of action attribution in schizophrenia. *Psychiatry Research: Neuroimaging*, *131*(1), 31–44.

Farrer, C., & Frith, C. D. (2002). Experiencing Oneself vs Another Person as Being the Cause of an Action: The Neural Correlates of the Experience of Agency. *NeuroImage*, *15*(3), 596–603. https://doi.org/10.1006/nimg.2001.1009

Farthing, G. W. (1992). *The Psychology of Consciousness*. Prentice-Hall.

Fasching, W. (2008). Consciousness, self-consciousness, and meditation. *Phenomenology and the Cognitive Sciences*, *7*(4), 463–483. https://doi.org/10.1007/s11097-008-9090-6

Fasching, W. (2010). "I Am of the Nature of Seeing": Phenomenological Reflections on the Indian Notion of Witness-Consciousness. In M. Siderits, E. Thompson, & D. Zahavi (Eds.), *Self, No Self?: Perspectives from Analytical, Phenomenological, and Indian Traditions*. Oxford University Press.

Fazekas, P., & Overgaard, M. (2018). Perceptual consciousness and cognitive access: an introduction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1755), 20170340. https://doi.org/10.1098/rstb.2017.0340

Feyaerts, J., Henriksen, M. G., Vanheule, S., Myin-Germeys, I., & Sass, L. A. (2021). Delusions beyond beliefs: a critical overview of diagnostic, aetiological, and therapeutic schizophrenia research from a clinical-phenomenological perspective. In *The Lancet Psychiatry* (Vol. 8, Issue 3, pp. 237–249). Elsevier Ltd. https://doi.org/10.1016/S2215-0366(20)30460-0

Flanagan, O. (2011). *The Bodhisattva's Brain : Buddhism Naturalized*. MIT Press.

Fontana, D. (2007). Meditation. In Max Velmans & S. Schneider (Eds.), *The Blackwell companion to consciousness* (pp. 154–162). Blackwell.

Forman, R. K. C. (1998). What does mysticism have to teach us about consciousness? *Journal of Consciousness Studies*, *5*(2), 185–201.

Fox, K. C. R., Dixon, M. L., Nijeboer, S., Girn, M., Floman, J. L., Lifshitz, M., Ellamil, M., Sedlmeier, P., & Christoff, K. (2016). Functional neuroanatomy of meditation: A review and meta-analysis of 78 functional neuroimaging investigations. In *Neuroscience and Biobehavioral Reviews*.

Fox, K. C. R., Kang, Y., Lifshitz, M., & Christoff, K. (2016). Increasing cognitive-emotional flexibility with meditation and hypnosis: The cognitive neuroscience of de-automatization. In A. Raz & M. Lifshitz (Eds.), *Hypnosis and meditation: Towards an integrative science of conscious planes*. Oxford University Press.

Frankfurt, H. G. (1971). Freedom of the Will and the Concept of a Person. *Journal of Philosophy*, *68*(1), 5–20.

Frankfurt, H. G. (1982). The importance of what we care about. *Synthese*, *53*(2), 257–272. https://doi.org/10.1007/BF00484902

Frankfurt, H. G. (1987). Identification and wholeheartedness. In F. Schoeman (Ed.), *Responsibility, character, and the emotions: New essays in moral psychology* (pp. 27–45). Cambridge University Press.

Frankfurt, H. G. (1988). *The importance of what we care about*. Cambridge University Press.

Freud, S. (1961). *The Ego and the Id (1923)*. W. W. Norton & Co.

Frewen, P., Schroeter, M. L., Riva, G., Cipresso, P., Fairfield, B., Padulo, C., Kemp, A. H., Palaniyappan, L., Owolabi, M., Kusi-Mensah, K., Polyakova, M., Fehertoi, N., D'Andrea, W., Lowe, L., & Northoff, G. (2020). Neuroimaging the consciousness of self: Review, and conceptual-methodological framework. In *Neuroscience and Biobehavioral Reviews* (Vol. 112, pp. 164–212). Elsevier Ltd. https://doi.org/10.1016/j.neubiorev.2020.01.023

Friston, K. (2003). Learning and inference in the brain. *Neural Networks*, *16*(9), 1325–1352. https://doi.org/10.1016/j.neunet.2003.06.005

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. https://doi.org/10.1038/NRN2787

Frith, C. D. (1992). *The Cognitive Neuropsychology of Schizophrenia*. Lawrence Erlbaum.

Frith, C. D., Blakemore, S.-J., & Wolpert, D. M. (2000). Abnormalities in the awareness and control of action. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *355*(1404), 1771–1788.

Fuchs, T. (2015). From Self-Disorders to Ego Disorders. *Psychopathology*, *48*(5), 324–331. https://doi.org/10.1159/000432404

Fulford, K. W. M., Davies, M., Gipps, R. G. T., Graham, G., Sadler, J. Z., Stanghellini, G., & Thornton, T. (Eds.). (2013). *The Oxford Handbook of Philosophy and Psychiatry* (Vol. 1). Oxford University Press.

Gallagher, S. (1997). Mutual enlightenment: recent phenomenology in cognitive science. *Journal of Consciousness Studies*, *4*(3), 195–214.

Gallagher, S. (2000). Philosophical conceptions of the self: Implications for cognitive science. *Trends in Cognitive Sciences*, *4*(1), 14–21. https://doi.org/10.1016/S1364-6613(99)01417-5

Gallagher, S. (2003). Phenomenology and Experimental Design: Toward a Phenomenologically Enlightened Experimental Science. *Journal of Consciousness Studies*, *10*(9–10), 85–99.

Gallagher, S. (2006). How the Body Shapes the Mind. In *How the Body Shapes the Mind*. Clarendon Press. https://doi.org/10.1093/0199271941.001.0001

Gallagher, S. (Ed.). (2011). *The Oxford Handbook of the Self*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199548019.001.0001

Gallagher, S. (2012). Multiple aspects in the sense of agency. *New Ideas in Psychology*, *30*(1), 15–31.

Gallagher, S. (2013). A pattern theory of self. *Frontiers in Human Neuroscience*, *7*(August), 443. https://doi.org/10.3389/fnhum.2013.00443

Gallagher, S. (2017a). *Enactivist interventions: Rethinking the mind*. Oxford University Press.

Gallagher, S. (2017b). Self-defense: Deflecting deflationary and eliminativist critiques of the sense of ownership. *Frontiers in Psychology*.

Gallagher, S. & Daly, A. (2018). Dynamical relations in the self-pattern. *Frontiers in Psychology*, *9*(MAY), 1–13. https://doi.org/10.3389/fpsyg.2018.00664

Gallagher, S. & Marcel, A. J. (1999). The self in contextualized action. *Journal of Consciousness Studies*, *6*(4), 4–30.

Gallagher, S. & Zahavi, D. (2007). *The Phenomenological Mind*. Routledge.

Gallese, V. (2014). Bodily selves in relation: embodied simulation as second-person perspective on intersubjectivity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1644). https://doi.org/10.1098/RSTB.2013.0177

Gallup, G. G. (1970). Chimpanzees: Self-Recognition. *Science*, *167*(3914), 86–87.

Gallup, G. G. (1977). Self recognition in primates: A comparative approach to the bidirectional properties of consciousness. *American Psychologist*, *32*(5), 329–

Gallup, G. G. (1982). Self-awareness and the emergence of mind in primates. *American Journal of Primatology*, *2*(3), 237–248.

Gallup, G. G., Anderson, J. R., & Platek, S. M. (2011). Self-Recognition. In S. Gallagher (Ed.), *The Oxford Handbook of the Philosophy of Consciousness*. Oxford University Press.

Ganeri, J. (2012). *The Self: Naturalism, Consciousness, and the First-Person Stance*. Oxford University Press.

Garfield, J. L. (2015). *Engaging Buddhism*. Oxford University Press. 1

Gazzaniga, Michael S. (1998). *The Mind's Past*. University of California Press.

Gennaro, R. J. (2020). Cotard syndrome, self-awareness, and I-concepts. *Philosophy and the Mind Sciences*, *1*(I), 1–20. https://doi.org/10.33735/phimisci.2020.i.41

Gerrans, P. (2002). A One-Stage Explanation of the Cotard Delusion. *Philosophy, Psychiatry & Psychology*, *9*(1), 47–53. https://doi.org/10.1353/ppp.2003.0007

Gerrans, P. (2013). Delusional Attitudes and Default Thinking. *Mind & Language*, *28*(1), 83–102.

Gerrans, P. (2015). All the Self We Need. *Open MIND*, *15*. https://doi.org/10.15502/9783958570078

Gerrans, P. (2019). Depersonalization Disorder, Affective Processing and Predictive Coding. *Review of Philosophy and Psychology*, *10*(2), 401–418.

Gertler, B. (2011). *Self-knowledge*. Routledge.

Gibson, E. J. (1994). Ontogenesis of the perceived self. In U. Neisser (Ed.), *The Perceived Self* (pp. 25–42). Cambridge University Press.

Gibson, J. J. (1977). The Theory of Affordances. In R. Shaw & J. Bransford (Eds.), *Perceiving, acting, and knowing. Toward an ecological psychology* (pp. 67–82). Hillsdale.

Gillihan, S. J., & Farah, M. J. (2005). Is self special? A critical review of evidence from experimental psychology and cognitive neuroscience. *Psychological Bulletin*, *131*(1), 76–97. https://doi.org/10.1037/0033-2909.131.1.76

Godfrin, K. A., & van Heeringen, C. (2010). The effects of mindfulness-based cognitive therapy on recurrence of depressive episodes, mental health and quality of life: A randomized controlled study. *Behaviour Research and Therapy*, *48*(8), 738–746. https://doi.org/10.1016/j.brat.2010.04.006

Goldman, A. I. (1970). *A Theory of Human Action*. Princeton University Press.

Goleman, D. (1978). *The varieties of the meditative experience*. Rider.

Gross, J. J., & Thomspon, R. A. (2007). Emotion Regulation: Conceptual Foundations. In J. J. Gross (Ed.), *Handbook of Emotion Regulation*. Guilford Press.

Grossman, P., Tiefenthaler-Gilmer, U., Raysz, A., & Kesper, U. (2007). Mindfulness Training as an Intervention for Fibromyalgia: Evidence of Postintervention and 3-Year Follow-Up Benefits in Well-Being. *Psychotherapy and Psychosomatics*, *76*(4), 226–233. https://doi.org/10.1159/000101501

Grünbaum, T., & Zahavi, D. (2013). Varieties of Self-Awareness. In K. W. M. Fulford, M. Davies, R. G. T. Gipps, G. Graham, J. Z. Sadler, G. Stanghellini, & T. Thornton (Eds.), *The Oxford Handbook of Philosophy and Psychiatry* (Vol. 1, pp. 221–239). Oxford University Press.

Gu, J., Strauss, C., Bond, R., & Cavanagh, K. (2015). How do mindfulness-based cognitive therapy and mindfulness-based stress reduction improve mental health and wellbeing? A systematic review and meta-analysis of mediation studies. In *Clinical Psychology Review* (Vol. 37, pp. 1–12). Elsevier Inc.

Gunn, R. (2016). On Thought Insertion. *Review of Philosophy and Psychology*, *7*(3), 559–575. https://doi.org/10.1007/s13164-015-0271-2

Haanila, H. (2021). The Dream Self and the Waking Self. In T. Koskinen, I., & Toppinen (Ed.), *Reasons and Responsibilities: Proceedings of the Philosophical Society of Finland Colloquium 2020 (Acta Philosophica Fennica; Vol. 97),* pp. 79–103.

Hasenkamp, W., Wilson-Mendenhall, C. D., Duncan, E., & Barsalou, L. W. (2012). Mind wandering and attention during focused meditation: A fine-

grained temporal analysis of fluctuating cognitive states. *NeuroImage*, *59*(1), 750–760. https://doi.org/10.1016/j.neuroimage.2011.07.008

Hayes, S. C., & Hofmann, S. G. (2017). The third wave of cognitive behavioral therapy and the rise of process-based care. *World Psychiatry*, *16*(3), 245–246.

Henriksen, M. G., Parnas, J., & Zahavi, D. (2019). Thought insertion and disturbed for-me-ness (minimal selfhood) in schizophrenia, *Consciousness and Cognition* Sep; 74: 102770. https://doi.org/10.1016/j.concog.2019.102770

Hohwy, J. (2013). *The Predictive Mind*. Oxford University Press.

Hohwy, J., Roepstorff, A., & Friston, K. (2008). Predictive coding explains binocular rivalry: An epistemological review. *Cognition*, *108*(3), 687–701.

Hölzel, B. K., Lazar, S. W., Gard, T., Schuman-Olivier, Z., Vago, D. R., & Ott, U. (2011). How does mindfulness meditation work? Proposing mechanisms of action from a conceptual and neural perspective. In *Perspectives on Psychological Science* (Vol. 6, Issue 6, pp. 537–559).

Huarcaya-Victoria, J., la Torre, J. B.-D., & De la Cruz-Oré, J. (2020). Factor structure of Cotard's syndrome: Systematic review of case reports. *Revista Colombiana de Psiquiatría (English Ed.)*, *49*(3), 187–193.

Huarcaya-Victoria, J., Ledesma-Gastañadui, M., & Huete-Cordova, M. (2016). Cotard's Syndrome in a Patient with Schizophrenia: Case Report and Review of the Literature. *Case Reports in Psychiatry*, *2016*, 1–7.

Hunter, E. C. M., Phillips, M. L., Chalder, T., Sierra, M., & David, A. S. (2003). Depersonalisation disorder: A cognitive-behavioural conceptualisation. *Behaviour Research and Therapy*, *41*(12), 1451–1467. https://doi.org/10.1016/S0005-7967(03)00066-4

Hunter, E. C. M., Sierra, M., & David, A. S. (2004). The epidemiology of depersonalisation and derealisation - A systematic review. In *Social Psychiatry and Psychiatric Epidemiology* (Vol. 39, Issue 1, pp. 9–18). https://doi.org/10.1007/s00127-004-0701-4

Hurley, S. L. (1998). *Consciousness in action*. Harvard University Press.

Husserl, E. (1952). *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie II. Husserliana IV*. Nijhoff.

James, W. (1890). *The principles of psychology, Vol I*. Henry Holt and Co. https://doi.org/10.1037/10538-000

Jaspers, K. (1997). *General psychopathology (1913)* (7th ed). University of Chicago Press.

Jáuregui Renaud, K. (2015). Vestibular Function and Depersonalization/ Derealization Symptoms. *Multisensory Research*, *28*(5–6), 637–651. https://doi.org/10.1163/22134808-00002480

Jones, M., & Love, B. C. (2011). Bayesian Fundamentalism or Enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *Behavioral and Brain Sciences*, *34*(4), 169–188.

Kabat-Zinn, J. (1994). *Wherever You Go, There You Are: Mindfulness Meditation in Everyday Life*. Hyperion.

Kabat-Zinn, J. (2005). *Coming to Our Senses: Healing Ourselves and the World through Mindfulness*. Hyperion.

Kahneman, D. (2011). *Thinking, Fast and Slow*. Macmillan.

Kaplan, D. (1989). Demonstratives: An Essay on the Semantics, Logic, Metaphysics and Epistemology of Demonstratives and other Indexicals. In J. Almog, J. Perry, & H. Wettstein (Eds.), *Themes From Kaplan* (pp. 481–563). Oxford University Press.

Killingsworth, M. A., & Gilbert, D. T. (2010). A Wandering Mind Is an Unhappy Mind. *Science*, *330*(6006), 932–932. https://doi.org/10.1126/science.1192439

King, A. P., Block, S. R., Sripada, R. K., Rauch, S., Giardino, N., Favorite, T., Angstadt, M., Kessler, D., Welsh, R., & Liberzon, I. (2016). Altered default mode network (DMN) resting state functional connectivity following a mindfulness-based exposure therapy for posttraumatic stress disorder (PTSD) in combat veterans of Afghanistan and Iraq. *Depression and Anxiety*.

Kiverstein, J. (2010). Minimal sense of self, temporality and the brain. *Psyche*, *15*(1), 59–74.

Kiverstein, J. (2020). Free Energy and the Self: An Ecological–Enactive Interpretation. *Topoi*, *39*(3), 559–574. https://doi.org/10.1007/s11245-018-9561-5

Kiverstein, J., & Clark, A. (2009). Introduction: Mind Embodied, Embedded, Enacted: One Church or Many? *Topoi*, *28*(1), 1–7. https://doi.org/10.1007/s11245-008-9041-4

Knyazev, G. G. (2013). EEG Correlates of Self-Referential Processing. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00264

Korsgaard, C. M. (1996). *The Sources of Normativity*. Cambridge University Press. https://doi.org/10.1017/CBO9780511554476

Kriegel, U. (2004). Consciousness and Self-Consciousness. *Monist*, *87*(2), 182–205. https://doi.org/10.5840/monist20048725

Kriegel, U. (2006). Consciousness: phenomenal consciousness, access consciousness, and scientific practise. In P. Thagarg (Ed.), *Handbook of Philosophy of Psychology and Cognitive Science* (pp. 195–217). Amsterdam: North-Holland.

Kriegel, U. (2009). *Subjective Consciousness*. Oxford University Press. 1

Kuyken, W., Byford, S., Taylor, R. S., Watkins, E., Holden, E., White, K., Barrett, B., Byng, R., Evans, A., Mullan, E., & Teasdale, J. D. (2008). Mindfulness-based cognitive therapy to prevent relapse in recurrent depression. *Journal of Consulting and Clinical Psychology*, *76*(6), 966–978.

Kuyken, W., Watkins, E., Holden, E., White, K., Taylor, R. S., Byford, S., Evans, A., Radford, S., Teasdale, J. D., & Dalgleish, T. (2010). How does mindfulness-based cognitive therapy work? *Behaviour Research and Therapy*, *48*(11),

Kyselo, M. (2014). The body social: An enactive approach to the self. *Frontiers in Psychology*, *5*(SEP). https://doi.org/10.3389/FPSYG.2014.00986

Lamme, V. A. F. (2010). How neuroscience will change our view on consciousness. *Cognitive Neuroscience*, *1*(3), 204–220.

Lane, R. D. (2008). Neural Substrates of Implicit and Explicit Emotional Processes: A Unifying Framework for Psychosomatic Medicine. *Psychosomatic Medicine*, *70*(2), 214–231. https://doi.org/10.1097/PSY.0b013e3181647e44

Lane, T. (2012). Toward an explanatory framework for mental ownership. *Phenomenology and the Cognitive Sciences*, *11*(2), 251–286. https://doi.org/10.1007/s11097-012-9252-4

Legerstee, M. (1998). Mental and bodily awareness in infancy: Consciousness of self-existence. *Journal of Consciousness Studies*, *5*(5–6), 627–44.

Legrand, D. (2005). Transparently oneself. *Psyche*, *11*(1995), 1–19.

Legrand, D. (2006). The bodily self: The sensorimotor roots of pre-reflective self-consciousness. *Phenomenology and the Cognitive Sciences*, *5*(1), 89–118. https://doi.org/10.1007/s11097-005-9015-6

Legrand, D. (2007). Pre-reflective self-as-subject from experiential and empirical perspectives. *Consciousness and Cognition*, *16*(3), 583–599. https://doi.org/10.1016/j.concog.2007.04.002

Legrand, D., & Grammont, F. (2005). A matter of facts. *Phenomenology and the Cognitive Sciences 2005 4:3*, *4*(3), 249–257. https://doi.org/10.1007/S11097-005-4072-4

Legrand, D., & Ravn, S. (2009). Perceiving subjectivity in bodily movement: The case of dancers. *Phenomenology and the Cognitive Sciences 2009 8:3*, *8*(3), 389–408. https://doi.org/10.1007/S11097-009-9135-5

Legrand, D., & Ruby, P. (2009). What is self-specific? Theoretical investigation and critical review of neuroimaging results. *Psychological Review*, *116*(1), 252–282. https://doi.org/10.1037/a0014172

Lemche, E., Brammer, M. J., David, A. S., Surguladze, S. A., Phillips, M. L., Sierra, M., Williams, S. C. R., & Giampietro, V. P. (2013). Interoceptive–reflective regions differentiate alexithymia traits in depersonalization disorder. *Psychiatry Research: Neuroimaging*, *214*(1), 66–72.

Lenggenhager, B., Tadi, T., Metzinger, T., & Blanke, O. (2007). Video Ergo Sum: Manipulating Bodily Self-Consciousness. *Science*, *317*(5841), 1096–1099.

Letheby, C., & Gerrans, P. (2017). Self unbound: ego dissolution in psychedelic experience. *Neuroscience of Consciousness*, *2017*(1). https://doi.org/10.1093/nc/nix016

Levine, J. (1983). Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly*, *64*(4), 354–361. https://doi.org/10.1111/j.1468-0114.1983.tb00207.x

Levine, J. (2001). *Purple Haze*. Oxford University Press. https://doi.org/10.1093/0195132351.001.0001

Limanowski, J., & Friston, K. (2020). Attenuating oneself. *Philosophy and the Mind Sciences*, *1*(I), 1–16. https://doi.org/10.33735/phimisci.2020.i.35

Lindahl, J. R., & Britton, W. B. (2019). Buddhist meditation and changes in the Sense of Self. *Journal of Consciousness Studies*, *26*(7–8), 157–183.

López-Silva, P. (2019). Me and I Are Not Friends, Just Aquaintances: on Thought Insertion and Self-Awareness. *Review of Philosophy and Psychology*, *10*(2), 319–335. https://doi.org/10.1007/s13164-017-0366-z

Lou, H. C., Changeux, J. P., & Rosenstand, A. (2017). Towards a cognitive neuroscience of self-awareness. In *Neuroscience and Biobehavioral Reviews* (Vol. 83, pp. 765–773). Elsevier Ltd. https://doi.org/10.1016/j.neubiorev.2016.04.004

Ludwig, A. M. (1966). Altered States of Consciousness. *Archives of General Psychiatry*, *15*(3), 225.

Lutz, A. (2002). Toward a neurophenomenology as an account of generative passages: a first empirical case study. *Phenomenology and the Cognitive Sciences*, *1*(2), 133–167. https://doi.org/10.1023/A:1020320221083

Lutz, A., Greischar, L. L., Rawlings, N. B., Ricard, M., & Davidson, R. J. (2004). Long-term meditators self-induce high-amplitude gamma synchrony during mental practice. *Proceedings of the National Academy of Sciences*, *101*(46), 16369–16373. https://doi.org/10.1073/pnas.0407401101

Lutz, A., & Thompson, E. (2003). Neurophenomenology: integrating subjective experience and brain dynamics in the neuroscience of consciousness. *Journal of Consciousness Studies*, *12*(3), 379–415.

Lycan, W. G. (1996). *Consciousness and Experience*. MIT Press.

MacIntyre, A. C. (1981). *After virtue: A study in moral theory*. University of Notre Dame Press.

Mackenzie, C. (2014). Embodied agents, narrative selves. In *Philosophical Explorations* (Vol. 17, Issue 2, pp. 154–171). Routledge. https://doi.org/10.1080/13869795.2014.886363

Malloy, P. F., & Richardson, E. D. (1994). The frontal lobes and content-specific delusions. *The Journal of Neuropsychiatry and Clinical Neurosciences*, *6*(4), 455–466. https://doi.org/10.1176/jnp.6.4.455

Mandik, P. (2007). The neurophilosophy of consciousness. In M. Velmans & S. Schneider (Eds.), *The Blackwell Companion to Consciousness* (pp. 458–471). John Wiley & Sons Ltd. https://doi.org/10.1002/9781119132363.ch33

Mandik, P. (2009). The Neurophilosophy of Subjectivity. In J. Bickle (Ed.), *The Oxford Handbook of Philosophy and Neuroscience*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195304787.003.0025

Martin, R., & Barresi, J. (2006). *The Rise and Fall of Soul and Self: An Intellectual History of Personal Identity*. Columbia University Press.

Martiny, K. M., Toro, J., & Høffding, S. (2021). Framing a Phenomenological Mixed Method: From Inspiration to Guidance. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.602081

Matthews, G. B. (1991). *Thought's Ego in Augustine and Descartes*. Cornell University Press.

Maturana, H., & Varela, F. J. (1980). *Autopoiesis and Cognition: The Realization of the Living*. D. Reidel.

McGinn, C. (1983). *The Subjective View: Secondary Qualities and Indexical Thoughts*. Oxford University Press. https://www.jstor.org/stable/2185441?origin=crossref

Mead, G. H. (1913). The Social Self. *The Journal of Philosophy, Psychology and Scientific Methods*, *10*(14), 374. https://doi.org/10.2307/2012910

Medford, N., Sierra, M., Stringaris, A., Giampietro, V., Brammer, M. J., & David, A. S. (2016). Emotional Experience and Awareness of Self: Functional MRI Studies of Depersonalization Disorder. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00432

Menary, R. (Ed.). (2010a). *The Extended Mind*. The MIT Press. https://doi.org/10.7551/mitpress/9780262014038.001.0001

Menary, R. (2010b). Introduction to the special issue on 4E cognition. *Phenomenology and the Cognitive Sciences*, *9*(4), 459–463. https://doi.org/10.1007/s11097-010-9187-6

Merleau-Ponty, M. (1962). *Phenomenology of Perception (Phénoménologie de la perception, 1945)*. Routledge.

Metzinger, T. (2003). *Being No One: The Self-Model Theory of Subjectivity*. The MIT Press. https://doi.org/10.7551/mitpress/1551.001.0001

Metzinger, T. (2004). Why Are Identity Disorders Interesting for Philosophers? In T. Schramme & J. Thome (Eds.), *Philosophy and Psychiatry* (pp. 311–325). De Gruyter. https://doi.org/10.1515/9783110905762.311

Metzinger, T. (2013). Why are dreams interesting for philosophers? The example of minimal phenomenal selfhood, plus an agenda for future research. *Frontiers in Psychology*, *4*(October), 746. https://doi.org/10.3389/fpsyg.2013.00746

Metzinger, T. (2020). Minimal phenomenal experience Meditation, tonic alertness, and the phenomenology of "pure" consciousness. *Philosophy and the Mind Sciences*, *1*(2020), 1–44.

Michal, M., Wiltink, J., Subic-Wrana, C., Zwerenz, R., Tuin, I., Lichy, M., Brähler, E., & Beutel, M. E. (2009). Prevalence, Correlates, and Predictors of Depersonalization Experiences in the German General Population. *Journal of Nervous & Mental Disease*, *197*(7), 499–506. https://doi.org/10.1097/NMD.0b013e3181aacd94

Millière, R., Carhart-Harris, R. L., Roseman, L., Trautwein, F.-M., & Berkovich-Ohana, A. (2018). Psychedelics, Meditation, and Self-Consciousness. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.01475

Minkowski, E. (1997). Du symptome au trouble générateur. In *Au-delà du rationalisme morbide (1928)* (pp. 93–124). Éditions l'Harmattan.

Moran, R. (2001). *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton University Press.

Murphy, D. (2020). Philosophy of Psychiatry. In Edward N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy* (Fall 2020). https://plato.stanford.edu/archives/fall2020/entries/psychiatry/

Musholt, K. (2013). A Philosophical Perspective on the Relation between Cortical Midline Structures and the Self. *Frontiers in Human Neuroscience*, *7*(SEP). https://doi.org/10.3389/fnhum.2013.00536

Naccache, L. (2018). Why and how access consciousness can account for phenomenal consciousness. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1755). https://doi.org/10.1098/rstb.2017.0357

Nagel, T. (1974). What is it like to be a bat? *The Philosophical Review*, *83*(4), 435–450. http://www.jstor.org/stable/2183914

Nagel, T. (1996). Universality and the reflective self. In C. M. Korsgaard (Ed.), *The Sources of Normativity* (pp. 200–209). Cambridge University Press. https://doi.org/10.1017/CBO9780511554476.009

Naranjo, C., & Ornstein, R. E. (1972). *On the psychology of meditation*. George Allen & Unwin.

Neisser, U. (1988). Five Kinds of Self-knowledge. *Philosophical Psychology*, *1*(1), 35–59. https://doi.org/10.1080/09515088808572924

Neisser, U. (1994). Self-narratives: True and false. In U. Neisser & R. Fivush (Eds.), *The Remembering Self* (pp. 1–18). Cambridge University Press.

Nelson, B., Fornito, A., Harrison, B. J., Yücel, M., Sass, L. a, Yung, a R., Thompson, A., Wood, S. J., Pantelis, C., & McGorry, P. D. (2009). A disturbed sense of self in the psychosis prodrome: linking phenomenology and neurobiology. *Neuroscience and Biobehavioral Reviews*, *33*(6), 807–817. https://doi.org/10.1016/j.neubiorev.2009.01.002

Newen, A. (2018). The embodied self, the pattern theory of self, and the predictive mind. *Frontiers in Psychology*, *9*(NOV). https://doi.org/10.3389/fpsyg.2018.02270

Newen, A., Gallagher, S., & De Bruin, L. (2018). 4E Cognition. In A. Newen, L. De Bruin, & S. Gallagher (Eds.), *The Oxford Handbook of 4E Cognition* (pp. 2–16). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198735410.013.1

Newen, A., Welpinghus, A., & Juckel, G. (2015). Emotion Recognition as Pattern Recognition: The Relevance of Perception. *Mind & Language*, *30*(2), 187–208. https://doi.org/10.1111/mila.12077

Norman, D. A., & Shallice, T. (1986). Attention to Action: Willed and Automatic Control of Behaviour. In R. J. Davidson, G. E. Schwartz, & D. E. Shapiro (Eds.), *Consciousness and Self-Regulation* (pp. 1–18). Plenum Press.

Northoff, G. (2005). Is emotion regulation self-regulation? *Trends in Cognitive Sciences*, *9*(9), 408–409. https://doi.org/10.1016/j.tics.2005.06.005

Northoff, G. (2011). Self and brain: what is self-related processing? *Trends in Cognitive Sciences*, *15*(5), 186–187. https://doi.org/10.1016/j.tics.2011.03.001

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.

Northoff, G., & Panksepp, J. (2008). The trans-species concept of self and the subcortical–cortical midline system. *Trends in Cognitive Sciences*, *12*(7), 259–264. https://doi.org/10.1016/j.tics.2008.04.007

O'Brien, L. (2007). *Self-Knowing Agents*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199261482.001.0001

Ochsner, K., & Gross, J. (2005). The cognitive control of emotion. *Trends in Cognitive Sciences*, *9*(5), 242–249. https://doi.org/10.1016/j.tics.2005.03.010

Parfit, D. (1971). Personal Identity. *The Philosophical Review*, *80*(1), 3.

Parnas, J., & Handest, P. (2003). Phenomenology of anomalous self-experience in early schizophrenia. *Comprehensive Psychiatry*, *44*(2), 121–134. https://doi.org/10.1053/comp.2003.50017

Parnas, J., Møller, P., Kircher, T., Thalbitzer, J., Jansson, L., Handest, P., & Zahavi, D. (2005). EASE: Examination of Anomalous Self-Experience. *Psychopathology*, *38*(5), 236–258. https://doi.org/10.1159/000088441

Parnas, J., Sass, L. A., & Zahavi, D. (2013). Rediscovering psychopathology: The epistemology and phenomenology of the psychiatric object. *Schizophrenia Bulletin*. https://doi.org/10.1093/schbul/sbs153

Parnas, J., & Zahavi, D. (2002). The Role of Phenomenology in Psychiatric Diagnosis and Classification. In M. Maj, W. Gaebel, J. J. López-Ibor, & N. Sartorius (Eds.), *Psychiatric Diagnosis and Classification* (pp. 137–162). https://doi.org/10.1002/047084647x.ch6

Parvizi, J., & Damasio, A. (2001). Consciousness and the brainstem. *Cognition*, *79*(1–2), 135–160. https://doi.org/10.1016/S0010-0277(00)00127-X

Paul, S. K. (2020). *Philosophy of Action*. Routledge.

Perry, J. (1977). Frege on Demonstratives. *The Philosophical Review*, *86*(4), 474. https://doi.org/10.2307/2184564

Perry, J. (1979). The Problem of the Essential Indexical. *Noûs*, *13*(1), 3–21. https://doi.org/10.2307/2214792

Perry, J. (1993). Perception, Action, and the Structure of Believing. *The Problem of the Essential Indexical and Other Essays*, 121–149.

Philippi, C. L., Feinstein, J. S., Khalsa, S. S., Damasio, A., Tranel, D., Landini, G., Williford, K., & Rudrauf, D. (2012). Preserved self-awareness following extensive bilateral brain damage to the insula, anterior cingulate, and medial prefrontal cortices. *PloS One*, *7*(8), e38413. https://doi.org/10.1371/journal.pone.0038413

Phillips, M. L., Ladouceur, C. D., & Drevets, W. C. (2008). A neural model of voluntary and automatic emotion regulation: implications for understanding the pathophysiology and neurodevelopment of bipolar disorder. *Molecular Psychiatry*, *13*(9), 833–857. https://doi.org/10.1038/mp.2008.65

Pickering, J. (1999). The self is a semiotic process. *Journal of Consciousness Studies*, *6*(4), 31–47.

Plotnik, J. M., de Waal, F. B. M., & Reiss, D. (2006). Self-recognition in an Asian elephant. *Proceedings of the National Academy of Sciences*, *103*(45), 17053–17057. https://doi.org/10.1073/pnas.0608062103

Posner, M. I., & Rothbart, M. K. (1998). Attention, self–regulation and consciousness. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *353*(1377), 1915–1927. https://doi.org/10.1098/rstb.1998.0344

Qin, P., Liu, Y., Shi, J., Wang, Y., Duncan, N., Gong, Q., Weng, X., & Northoff, G. (2012). Dissociation between anterior and posterior cortical regions during self-specificity and familiarity: A combined fMRI-meta-analytic study. *Human Brain Mapping*, *33*(1), 154–164. https://doi.org/10.1002/hbm.21201

Radhakrishnan, S., & Moore, C. A. (1957). *A source book in Indian philosophy*. Princeton University Press.

Radovic, F. (2017). The sense of death and non-existence in nihilistic delusions. *Phenomenology and the Cognitive Sciences*, *16*(4), 679–699. https://doi.org/10.1007/s11097-016-9467-x

Radovic, F., & Radovic, S. (2002). Feelings of Unreality: A Conceptual and Phenomenological Analysis of the Language of Depersonalization. *Philosophy, Psychiatry & Psychology*, *9*(3), 271–279. https://doi.org/10.1353/ppp.2003.0048

Räikkä, J. (2020). On the Nontechnical Limits of Brain Imaging. *Cambridge Quarterly of Healthcare Ethics*, *29*(4), 527–541. https://doi.org/10.1017/S0963180120000298

Ramachandran, V. S., & Blakeslee, S. (1998). *Phantoms in the Brain. Human nature and the architecture of the mind*. Harper Perennial.

Ramirez-Bermudez, J., Aguilar-Venegas, L. C., Crail-Melendez, D., Espinola-Nadurille, M., Nente, F., & Mendez, M. F. (2010). Cotard Syndrome in Neurological and Psychiatric Patients. *The Journal of Neuropsychiatry and Clinical Neurosciences*, *22*, 409–416.

Ratcliffe, M. (2008). *Feelings of Being: Phenomenology, psychiatry and the sense of reality*. Oxford University Press. https://doi.org/10.1093/med/9780199206469.001.0001

Régis, E. (1893). Note historique et clinique sur le délire des négations. *Gazette Médicale de Paris*, *2*, 61–64.

Reiss, D., & Marino, L. (2001). Mirror self-recognition in the bottlenose dolphin: A case of cognitive convergence. *Proceedings of the National Academy of Sciences*, *98*(10), 5937–5942. https://doi.org/10.1073/pnas.101086398

Reutens, S., Nielsen, O., & Sachdev, P. (2010). Depersonalization disorder. In *Current Opinion in Psychiatry* (Vol. 23, Issue 3, pp. 278–283). https://doi.org/10.1097/YCO.0b013e3283387ab4

Revonsuo, A. (2005). The Self in Dreams. In *The Lost Self* (pp. 206–219). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195173413.003.0014

Revonsuo, A. (2006). *Inner presence: Consciousness as a biological phenomenon.* MIT Press.

Revonsuo, A. (2010). *Consciousness: the science of subjectivity*. Psychology Press.

Revonsuo, A., Kallio, S., & Sikka, P. (2009). What is an altered state of consciousness? *Philosophical Psychology*, *22*(2), 187–204. https://doi.org/10.1080/09515080902802850

Rochat, P. (2001). *The Infant's World*. Harvard University Press.

Rochat, P. (2011). What is it Like to be a Newborn? In S Gallagher (Ed.), *The Oxford Handbook of the Self*. Oxford University Press.

Rochat, P. (2018). The Ontogeny of Human Self-Consciousness. *Current Directions in Psychological Science*, *27*(5), 345–350.

Rochat, P., & Zahavi, D. (2011). The uncanny mirror: A re-framing of mirror self-experience. *Consciousness and Cognition*, *20*(2), 204–213. https://doi.org/10.1016/J.CONCOG.2010.06.007

Rosenthal, D. M. (1991). The Independence of Consciousness and Sensory Quality. *Philosophical Issues*, *1*, 15. https://doi.org/10.2307/1522921

Rosenthal, D. M. (2005). *Consciousness and Mind*. Clarendon Press.

Rosenthal, D. M. (2009). Higher-Order Theories of Consciousness. In *The Oxford Handbook of Philosophy of Mind*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199262618.003.0014

Rosenthal, D. M. (2010). How to think about mental qualities. *Philosophical Issues*, *20*, 365–393. https://doi.org/10.1007/s11098-013-0172-0

Ruby, P., & Decety, J. (2003). What you believe versus what you think they believe: a neuroimaging study of conceptual perspective-taking. *European Journal of Neuroscience*, *17*(11), 2475–2480. https://doi.org/10.1046/j.1460-9568.2003.02673.x

Ruby, P., & Decety, J. (2004). How Would You Feel versus How Do You Think She Would Feel? A Neuroimaging Study of Perspective-Taking with Social Emotions. *Journal of Cognitive Neuroscience*, *16*(6), 988–999. https://doi.org/10.1162/0898929041502661

Sandman, N. (2017). *Nightmares: Epidemiological studies of subjective experiences*. University of Turku.

Sartre, J.-P. (1956). *Being and nothingness. An essay on phenomenological ontology (1943)*. Philosophical Library.

Sass, L. A. (1994). Civilized madness: schizophrenia, self-consciousness and the modern mind. *History of the Human Sciences*, *7*(2), 83–120. https://doi.org/10.1177/095269519400700206

Sass, L. A., Borda, J. P., Madeira, L., Pienkos, E., & Nelson, B. (2018). Varieties of Self Disorder: A Bio-Pheno-Social Model of Schizophrenia. *Schizophrenia Bulletin*, *44*(4), 720–727. https://doi.org/10.1093/schbul/sby001

Sass, L. A., & Parnas, J. (2003). Schizophrenia, Consciousness, and the Self. *Schizophrenia Bulletin*, *29*(3), 427–444. https://doi.org/10.1093/oxfordjournals.schbul.a007017

Sass, L. A., & Parnas, J. (2007). Explaining schizophrenia: the relevance of phenomenology. In M. C. Chung, K. W. M. Fulford, & G. Graham (Eds.), *Reconceiving Schizophrenia* (pp. 63–95). Oxford University Press. https://doi.org/10.1093/med/9780198526131.003.0004

Sass, L. A., Pienkos, E., Nelson, B., & Medford, N. (2013). Anomalous self-experience in depersonalization and schizophrenia: a comparative investigation. *Consciousness and Cognition*, *22*(2), 430–441. https://doi.org/10.1016/j.concog.2013.01.009

Schechtman, M. (2011). The Narrative Self. In S Gallagher (Ed.), *The Oxford Handbook of the Self*. Oxford University Press.

Schmidt, S., & Walach, H. (Eds.). (2014). *Meditation – Neuroscientific Approaches and Philosophical Implications* (Vol. 2). Springer. https://doi.org/10.1007/978-3-319-01634-4

Schneider, F., Bermpohl, F., Heinzel, A., Rotte, M., Walter, M., Tempelmann, C., Wiebking, C., Dobrowolny, H., Heinze, H. J., & Northoff, G. (2008). The resting brain and our self: Self-relatedness modulates resting state neural activity in cortical midline structures. *Neuroscience*, *157*(1), 120–131.

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, *3*(3), 417–424. https://doi.org/10.1017/S0140525X00005756

Searle, J. R. (2005). The Self as a Problem in Philosophy and Neurobiology. In *The Lost Self* (pp. 7–19). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195173413.003.0002

Seeley, W. W., Menon, V., Schatzberg, A. F., Keller, J., Glover, G. H., Kenna, H., Reiss, A. L., & Greicius, M. D. (2007). Dissociable Intrinsic Connectivity Networks for Salience Processing and Executive Control. *Journal of Neuroscience*, *27*(9), 2349–2356.

Sèglas, J. (1897). *Le délire des négations: séméiologie et diagnostic*. Masson, Gauthier-Villars.

Seigel, J. (2005). *The Idea of the Self*. Cambridge University Press.

Sekida, K. (1985). *Zen training: Methods and philosophy*. Weatherhill.

Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. In *Trends in Cognitive Sciences* (Vol. 17, Issue 11, pp. 565–573). https://doi.org/10.1016/j.tics.2013.09.007

Seth, A. K., Suzuki, K., & Critchley, H. D. (2012). An Interoceptive Predictive Coding Model of Conscious Presence. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00395

Seth, A. K., & Tsakiris, M. (2018). Being a Beast Machine: The Somatic Basis of Selfhood. In *Trends in Cognitive Sciences* (Vol. 22, Issue 11, pp. 969–981). https://doi.org/10.1016/j.tics.2018.08.008

Sheets-Johnstone, M. (1999). *The Primacy of Movement* (Vol. 14). John Benjamins Publishing Company. https://doi.org/10.1075/aicr.14

Shoemaker, S. (1968). Self-Reference and Self-Awareness. *The Journal of Philosophy*, *65*(19), 555. https://doi.org/10.2307/2024121

Shoemaker, S. (1988). On Knowing One's Own Mind. *Philosophical Perspectives*, *2*, 183–209. https://doi.org/10.2307/2214074

Shoemaker, S. (2011). On What we are. In S Gallagher (Ed.), *The Oxford Handbook of the Self*. Oxford University Press.

Shusterman, R. (2008). *Body Consciousness*. Cambridge University Press.

Siderits, M., Thompson, E., & Zahavi, D. (Eds.). (2010). *Self, No Self?: Perspectives from Analytical, Phenomenological, and Indian Traditions*. Oxford University Press.

Sierra, M. (2009). *Depersonalization: A New Look at a Neglected Syndrome*. Cambridge University Press. https://doi.org/10.1017/CBO9780511730023

Sierra, M., Baker, D., Medford, N., & David, A. S. (2005). Unpacking the depersonalization syndrome: an exploratory factor analysis on the Cambridge Depersonalization Scale. *Psychological Medicine*, *35*(10), 1523–1532. https://doi.org/10.1017/S0033291705005325

Sierra, M., & Berrios, G. E. (1998). Depersonalization: Neurobiological perspectives. *Biological Psychiatry*, *44*(9), 898–908. https://doi.org/10.1016/S0006-3223(98)00015-8

Sierra, M., & Berrios, G. E. (2000). The Cambridge Depersonalisation Scale: a new instrument for the measurement of depersonalisation. *Psychiatry Research*, *93*(2), 153–164. https://doi.org/10.1016/S0165-1781(00)00100-1

Sierra, M., & David, A. S. (2011). Depersonalization: A selective impairment of self-awareness. *Consciousness and Cognition*, *20*(1), 99–108. https://doi.org/10.1016/j.concog.2010.10.018

Sikka, P. (2020). *Dream affect: Conceptual and methodological issues in the study of emotions and moods experienced in dreams*. University of Turku.

Silva, J. A., Leong, G. B., Weinstock, R., & Gonzales, C. L. (2000). A case of Cotard's syndrome associated with self- starvation. *Journal of Forensic Sciences*, *45*(1), 188–190.

Silverton, S. (2012). *The Mindfulness Breakthrough*. Watkins Publishing.

Simeon, D., & Abugel, J. (2006). *Feeling Unreal: Depersonalization Disorder and the Loss of the Self*. Oxford University Press.

Smith, J. (2020). Self-Consciousness. In E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy* (Summer 2020 ed.). https://plato.stanford.edu/archives/sum2020/entries/self-consciousness/

Solomon, R. C. (1988). *Continental philosophy since 1750 : the rise and fall of the self*. Oxford University Press.

Sorabji, R. (2006). *Self: Ancient and Modern Insights about Individuality, Life, and Death*. Clarendon Press.

Stein, D. J., & Simeon, D. (2009). Cognitive-affective neuroscience of depersonalization. *CNS Spectrums*, *14*(9), 467–471.

Stern, D. N. (1985). *The interpersonal world of the infant: A view from Psychoanalysis and Developmental Psychology*. Basic Books.

Stewart, J., Gapenne, O., & Di Paolo, E. A. (2010). *Enaction: Towards a New Paradigm for Cognitive Science*. MIT Press.

Strauss, C., Cavanagh, K., Oliver, A., & Pettman, D. (2014). Mindfulness-Based Interventions for People Diagnosed with a Current Episode of an Anxiety or Depressive Disorder: A Meta-Analysis of Randomised Controlled Trials. *PLoS ONE*, *9*(4), e96110. https://doi.org/10.1371/journal.pone.0096110

Strawson, G. (1999). The Self and the SESMET. *Journal of Consciousness Studies*, *6*(4), 99–135. https://doi.org/10.1007/BF03342659

Strawson, G. (2000). The phenomenology and ontology of the self. In D. Zahavi (Ed.), *Exploring the Self* (p. 39). John Benjamins. https://doi.org/10.1075/aicr.23.05str

Strawson, G. (2004). Against Narrativity. *Ratio*, *17*(4), 428–452. https://doi.org/10.1111/j.1467-9329.2004.00264.x

Strawson, G. (2015). 'Self-intimation.' *Phenomenology and the Cognitive Sciences*, *14*(1), 1–31. https://doi.org/10.1007/s11097-013-9339-6

Suarez, S. D., & Gallup, G. G. (1981). Self-recognition in chimpanzees and orangutans, but not gorillas. *Journal of Human Evolution*, *10*(2), 175–188.

Sui, J., & Gu, X. (2017). Self as Object: Emerging Trends in Self Research. *Trends in Neurosciences*, *40*(11), 643–653. https://doi.org/10.1016/j.tins.2017.09.002

Swamy, K. N. C., Sanju, G., & Jaimon, M. (2007). An overview of the neurological correlates of Cotard syndrome. *European Journal of Psychiatry*, *21*(2), 99–116. https://doi.org/10.4321/s0213-61632007000200002

Talbert, M. (2019). Moral Responsibility. In Edward N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy* (Winter 2019 ed.). https://plato.stanford.edu/archives/win2019/entries/moral-responsibility/

Tang, Y.-Y., Hölzel, B. K., & Posner, M. I. (2015). The neuroscience of mindfulness meditation. *Nature Reviews Neuroscience*, *16*(4), 213–225. https://doi.org/10.1038/nrn3916

Tart, C. T. (Ed.). (1990). *Altered states of consciousness*. HarperCollins.

Taylor, C. (1989). *Sources of the Self*. Harvard University Press.

Teasdale, J. D., Segal, Z. V., Williams, J. M. G., Ridgeway, V. A., Soulsby, J. M., & Lau, M. A. (2000). Prevention of relapse/recurrence in major depression by mindfulness-based cognitive therapy. *Journal of Consulting and Clinical Psychology*, *68*(4), 615–623. https://doi.org/10.1037/0022-006X.68.4.615

Thagard, P. (2013). Cognitive sciences. In B. Kaldis (Ed.), *Encyclopedia of philosophy and the social sciences (Vol. 1)* (pp. 95–99). Sage.

Thiel, U. (2011). *The Early Modern SubjectSelf-Consciousness and Personal Identity from Descartes to Hume*. Oxford University Press.

Thompson, E. (2007). *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*. Harvard University Press.

Thompson, E. (2008). Neurophenomenology and Contemplative Experience. In P. Clayton (Ed.), *The Oxford Handbook of Religion and Science*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199543656.003.0015

Thompson, E. (2014). *Waking, Dreaming, Being*. Columbia University Press. https://doi.org/10.7312/thom13709

Thompson, E., Lutz, A., & Cosmelli, D. (2005). Neurophenomenology: An introduction for neurophilosophers. In *Cognition and the Brain: The Philosophy and Neuroscience Movement* (pp. 40–97). https://doi.org/10.1017/CBO9780511610608.003

Thompson, E., & Stapleton, M. (2009). Making Sense of Sense-Making: Reflections on Enactive and Extended Mind Theories. *Topoi*, *28*, 23–30. https://doi.org/10.1007/s11245-008-9043-2

Thompson, E., & Zahavi, D. (2007). Philosophical Issues: Phenomenology. In P. D. Zelazo, M. Moscovitch, & E. Thompson (Eds.), *The Cambridge Handbook of Consciousness* (pp. 67–88). Cambridge University Press. https://doi.org/10.1017/CBO9780511816789.005

Tibubos, A. N., Grammes, J., Beutel, M. E., Michal, M., Schmutzer, G., & Brähler, E. (2018). Emotion regulation strategies moderate the relationship of fatigue with depersonalization and derealization symptoms. *Journal of Affective Disorders*, *227*, 571–579. https://doi.org/10.1016/j.jad.2017.11.079

Tononi, G., Boly, M., Massimini, M., & Koch, C. (2016). Integrated information theory: from consciousness to its physical substrate. *Nature Reviews Neuroscience*, *17*(7), 450–461. https://doi.org/10.1038/nrn.2016.44

Tsakiris, M. (2017). The multisensory basis of the self: From body to identity to others. *Quarterly Journal of Experimental Psychology*, *70*(4), 597–609. https://doi.org/10.1080/17470218.2016.1181768

Tye, M. (2003). *Consciousness and Persons*. MIT Press.

Vago, D. R., & Silbersweig, D. A. (2012). Self-awareness, self-regulation, and self-transcendence (S-ART): a framework for understanding the neurobiological mechanisms of mindfulness. *Frontiers in Human Neuroscience*, *6*. https://doi.org/10.3389/fnhum.2012.00296

Vaitl, D., Birbaumer, N., Gruzelier, J., Jamieson, G. A., Kotchoubey, B., Kübler, A., Lehmann, D., Miltner, W. H. R., Ott, U., Pütz, P., Sammer, G., Strauch, I., Strehl, U., Wackermann, J., & Weiss, T. (2005). Psychobiology of Altered States of Consciousness. *Psychological Bulletin*, *131*(1), 98–127. https://doi.org/10.1037/0033-2909.131.1.98

Van Dam, N. T., van Vugt, M. K., Vago, D. R., Schmalzl, L., Saron, C. D., Olendzki, A., Meissner, T., Lazar, S. W., Kerr, C. E., Gorchov, J., Fox, K. C. R., Field, B. A., Britton, W. B., Brefczynski-Lewis, J. A., & Meyer, D. E. (2018). Mind the Hype: A Critical Evaluation and Prescriptive Agenda for Research on Mindfulness and Meditation. *Perspectives on Psychological Science*, *13*(1), 36–61. https://doi.org/10.1177/1745691617709589

Van Gulick, R. (2014). Consciousness. In Edward N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy* (Spring2018 ed.). https://plato.stanford.edu/archives/spr2018/entries/consciousness/

Varela, F. J. (1996). Neurophenomenology: a methodological remedy to the hard problem. *Journal of Consciousness Studies*, *3*(4), 330–350.

Varela, F. J., Rosch, E., & Thompson, E. (1991). *The Embodied Mind*. The MIT Press. https://doi.org/10.7551/mitpress/6730.001.0001

Varela, F., Lachaux, J. P., Rodriguez, E., & Martinerie, J. (2001). The brainweb: Phase synchronization and large-scale integration. *Nature Reviews Neuroscience*, *2*(4), 229–239. https://doi.org/10.1038/35067550

Vogeley, K., Bussfeld, P., Newen, A., Herrmann, S., Happé, F., Falkai, P., Maier, W., Shah, N. J., Fink, G. R., & Zilles, K. (2001). Mind Reading: Neural Mechanisms of Theory of Mind and Self-Perspective. *NeuroImage*, *14*(1), 170–181. https://doi.org/10.1006/nimg.2001.0789

Wallace, A. (2007). *Contemplative Science: Where Buddhism and Neuroscience Converge*. Columbia University Press.

Ward, D., Silverman, D., & Villalobos, M. (2017). Introduction: The Varieties of Enactivism. *Topoi*, *36*(3), 365–375. https://doi.org/10.1007/s11245-017-9484-6

Ward, D., & Stapleton, M. (2012). Es are good. Cognition as enacted, embodied, embedded, affective and extended. In F. Paglieri (Ed.), *Consciousness in Interaction. The Role of the Natural and Social Environment in Shaping Consciousness* (pp. 89–104). John Benjamins.

Weiskrantz, L. (1986). *Blindsight*. Oxford University Press.

Whiting, J. E. (2002). Personal Identity: The Non-Branching Form of "What Matters." In R. M. Gale (Ed.), *The Blackwell Guide to Metaphysics* (pp. 190–218). Blackwell.

Wicker, B., Ruby, P., Royet, J.-P., & Fonlupt, P. (2003). A relation between rest and the self in the brain? *Brain Research Reviews*, *43*(2), 224–230. https://doi.org/10.1016/j.brainresrev.2003.08.003

Wilson, G., & Shpall, S. (2016). Action. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016 ed.). https://plato.stanford.edu/archives/win2016/entries/action/

Windt, J. M. (2010). The immersive spatiotemporal hallucination model of dreaming. *Phenomenology and the Cognitive Sciences*, *9*(2), 295–316. https://doi.org/10.1007/s11097-010-9163-1

Windt, J. M. (2015). *Dreaming: A Conceptual Framework for Philosophy of Mind and Empirical Research*. MIT Press.

Wittgenstein, L. (1958). *The Blue and Brown Books*. Blackwell.

Wittmann, M. (2018). *Altered States of Consciousness*. The MIT Press. https://doi.org/10.7551/mitpress/11468.001.0001

Yamada, K., Katsuragi, S., & Fujii, I. (2007). A case study of Cotard's syndrome: stages and diagnosis. *Acta Psychiatrica Scandinavica*, *100*(5), 396–398.

Young, A. W., & Leafhead, K. (1996). Betwixt Life and Death: Case Studies of the Cotard Delusion. In P. W. Halligan & J. C. Marshall (Eds.), *Method in madness: case studies in cognitive neuropsychiatry* (pp. 147– 171). Psychology Press.

Young, G. (2012). Delusions of death and immortality: a consequence of being in cotard patients. *Philosophy, Psychiatry & Psychology*, *19*(2), 127–140.

Zahavi, D. (1999). *Self-Awareness and Alterity: A Phenomenological Investigation*. Northwestern University Press.

Zahavi, D. (Ed.). (2000). *Exploring the Self*. John Benjamins.

Zahavi, D. (2005a). Being someone. *Psyche*, *11*(5), 699.

Zahavi, D. (2005b). *Subjectivity and Selfhood*. The MIT Press.

Zahavi, D. (2010a). Minimal self and narrative self: A distinction in need of refinement. In T. Fuchs, H. Sattel, & P. Heningnsen (Eds.), *The Embodied Self: Dimensions, Coherence, and Disorders* (pp. 3–11). Schattauer.

Zahavi, D. (2010b). Zahavi - Naturalized Phenomenology (2010). In S. Gallagher & D. Schmicking (Ed.), *Handbook of Phenomenology and Cognitive Science* (pp. 2–19). Springer.

Zahavi, D. (2010c). The Experiential Self: Objections and Clarifications. In M. Siderits, E. Thompson, & D. Zahavi (Eds.), *Self, No Self? Perspectives from Analytical, Phenomenological, and Indian Traditions* (pp. 56–78). Oxford University Press.

Zahavi, D. (2014). *Self and Other: Exploring Subjectivity, Empathy, and Shame*. Oxford University Press.

Zahavi, D. (2017). Thin, Thinner, Thinnest: Defining the Minimal Self. In C. Durt, T. Fuchs, & C. Tewes (Eds.), *Embodiment, Enaction, and Culture* (pp. 193–199). The MIT Press.

Zahavi, D. (2018). Consciousness, Self-Consciousness, Selfhood: a Reply to some Critics. In *Review of Philosophy and Psychology* (Vol. 9, Issue 3). https://doi.org/10.1007/s13164-018-0403-6

Zahavi, D. (2020). Consciousness and Selfhood: Getting clearer on for-me-ness and mineness. In U. Kriegel (Ed.), *The Oxford Handbook of the Philosophy of Consciousness* (pp. 634–653). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198749677.013.29

Zahavi, D., & Kriegel, U. (2016). For-me-ness: What it is and what it is not. In D. O. Dahlstrom, A. Elpidorou, & W. Hopp (Eds.), *Philosophy of Mind and Phenomenology: Conceptual and Empirical Approaches*. Routledge.

Zelazo, P. D., Gao, H. H., & Todd, R. (2007). The Development of Consciousness. In P. D. Zelazo, M. Moscovitch, & E. Thompson (Eds.), *The Cambridge Handbook of Consciousness* (pp. 405–432). Cambridge University Press. https://doi.org/10.1017/CBO9780511816789.016

Zhao, W., Luo, L., Li, Q., & Kendrick, K. M. (2013). What Can Psychiatric Disorders Tell Us about Neural Processing of the Self? *Frontiers in Human Neuroscience*, *7*(August), 485. https://doi.org/10.3389/fnhum.2013.00485

The self plays a principal role in an individual's life and has been a central theme in the philosophy of mind. In this dissertation, selfhood is investigated within a neurophilosophical framework, which highlights cooperation between the philosophy of mind and empirical sciences. The study focuses on self-consciousness, and the author argues for the importance of the dynamic interaction between minimal and reflective forms of self-consciousness. Although a distinction between these two forms is often drawn, their interconnections have been under-studied. The significance of these connections is indicated in this study through a detailed analysis of constituents of both forms of self-consciousness and their relations. These relations are also considered in terms of the neural mechanisms underlying experiences of the self. In addition to theoretical insight, the idea of the dynamics of self-consciousness is found to be useful in practice to account for concrete cases of altered states of consciousness that involve changes in the experience of the self. The appeal to the connections in self-consciousness provides a more accurate picture of these states, which strengthens the idea that self-consciousness as a whole can be understood only by acknowledging the mutual influence of minimal and reflective self-consciousness on each other.