EXTIMATE TECHNOLOGY

SELF-FORMATION IN A TECHNOLOGICAL WORLD

Ciano Aydin

Extimate Technology

This book investigates how we should form ourselves in a world saturated with technologies that are profoundly intruding in the very fabric of our selfhood.

New and emerging technologies, such as smart technological environments, imaging technologies and smart drugs, are increasingly shaping who and what we are and influencing who we ought to be. How should we adequately understand, evaluate and appreciate this development? Tackling this question requires going beyond the persistent and stubborn inside-outside dualism and recognizing that what we consider our "inside" self is to a great extent shaped by our "outside" world. Inspired by various philosophers – especially Nietzsche, Peirce and Lacan – this book shows how the values, goals and ideals that humans encounter in their environments not only shape their identities but also can enable them to critically relate to their present state. The author argues against understanding technological self-formation in terms of making ourselves better, stronger and smarter. Rather, we should conceive it in terms of technological sublimation, which redefines the very notion of human enhancement. In this respect the author introduces an alternative, more suitable theory, namely Technological Sublimation Theory (TST).

Extimate Technology will be of interest to scholars and advanced students working in philosophy of technology, philosophy of the self, phenomenology, pragmatism and history of philosophy.

Ciano Aydin is Full Professor of Philosophy of Technology, Head of the Department of Philosophy and Vice-Dean of the Faculty of Behavioural, Management and Social Sciences (BMS) at the University of Twente, The Netherlands. He has published in *Transactions of the Charles S. Peirce Society, Phenomenology and the Cognitive Sciences, Journal of Medicine and Philosophy, Philosophy and Technology,* and other journals. See www.cianoaydin.nl for more information about his current research.

Routledge Studies in Contemporary Philosophy

For more information about this series, please visit: https://www.routledge.com/Routledge-Studies-in-Contemporary-Philosophy/book-series/SE0720

Extimate Technology

Self-Formation in a Technological World

Ciano Aydin

Routledge
Taylor & Francis Group

NEW YORK AND LONDON

To Nur, Zeno, Daniël and Helena

Contents

Acknowledgments

This book has a long history and originated from a seemingly simple idea that became an obsession: in Western culture and philosophy, the

notes are often mistaken for the melody (after Bergson). Different forms of dualism, essentialism and reductionism, which are also reflected in our views of technology, seem to be manifestations of this misconception. This book is an attempt to uncover and disentangle this confusion and to provide a more appropriate framework for understanding how technologies shape the human self.

Introduction

From Inside to Outside and from Outside to Inside

Shelter, clothing, the wheel, stone axes, furnaces, steam engines, roadways, canals, plumbing, clocks, furniture, automobiles, the telegraph, the telephone, cutlery, water taps, electric kettles, refrigerators, plains, guns, newspapers, computers, mobile phones: a relatively random list of technologies that have profoundly shaped our world. Characteristic of these technologies is that they are aimed toward the outside world. They serve as protection from the elements, enable interaction with outside objects and allow us to employ, manipulate and regulate the environment. They make it possible, directly or indirectly, to control, from an inside realm, the outside world.

This movement from inside to outside is also articulated by Ernst Kapp (1877), who can be considered the founding father of philosophy of technology. Kapp attempts to illustrate that technologies are essentially projections or exteriorizations of bodily organs in order to control and use the outside world. The movement from inside to outside is expressed by the human desire to tame and manipulate the obstinate environment. Technological objects extend the human organism by replicating or amplifying bodily and mental abilities: a hammer, for instance, is basically a strengthened version of a human fist, clothing is a duplication of the skin, telegraph cables are an extension of human nerves and railway systems are a copy of the vascular system. One hundred years later, Marshall McLuhan (1966) interprets technologies as amplifications or accelerations of functions originally performed by the unaided human organism; technologies, he claims, take over or supplement these functions. McLuhan has focused on media as extensions of the senses, particularly those of sight and sound; radio and telephone are seen as extensions of ears, while visual media, including writing and print, are understood as extensions of eyes (see also Brey 2000).

Over the past decades, we witnessed the emergence of different types of technologies: pacemakers, artificial heart valves, artificial hips, EEG, fMRI, cochlear implants, retinal chips, Deep Brain Stimulation, Google Glass, smart mirrors, smart bracelets, nanopills, RFID chips, smart

drugs, eyeborgs, and movement, health and emotion sensors and actu-ators. In general, these technologies are small, smart and more person-alized (Van Est et al. 2014). Characteristic of these technologies is that instead of being aimed toward the outside world, they are moving from the outside toward the inside, intruding into the human's body and psy-chology. Now the human being seems to be the target of control, change and modification.

   This movement from outside to inside is accelerating and intensify-ing. Since 1958, artificial cardiac pacemakers aid the functioning of the heart, generating electrical impulses delivered by electrodes to contract the heart muscles and regulate the electrical conduction system of the heart. Since the 1990s, patients with Parkinson's disease are being treated with Deep Brain Stimulation that reduces severe tremor by send-ing electrical pulses to an electrode that has been surgically inserted in the brain. Some studies show that people with obsessive-compulsive disorder or major depression also seem to benefit from Deep Brain Stim-ulation, instigating ethical debate about the aptness of this technology to treat mental illness. At my own university (the University of Twente), an electrical pill has been developed that contains sensitive nanowires that detect deviant DNA fragments in the intestines that may indicate early stages of cancer. If suitable, the patient could receive the measurements done by this pill on her smartphone. Other smart pills at nanoscale that contain medication have been developed; after being injected in the body, these pills, instead of contaminating the whole body, are able to find and release the medication near or in the targeted diseased tissue. A cochlear implant allows deaf babies to develop significant hearing and some adults to improve their hearing. A retinal photovoltaic chip enables blind people with a slightly damaged optic nerve to gain some sight. fMRI images are potentially tamper-proofing lie detection, providing insight into various psychological disorders. Smart mirrors confront people with, for example, bad eating habits, and persuade them to live healthier.

   These are just a few examples of intrusive technologies that are increasingly getting closer to and penetrating our corporeal surfaces, giv-ing rise to an intimate interplay between bodies, organs and technologies (Lettow 2011; Dalibert 2014; Zwart 2015, 2017). Technologies that for millennia have been developed and used to control our outside world, now, as Garreau (2005, 6) puts it, "have started to merge with our minds, our memories, our metabolisms, our personalities, our progeny and perhaps our souls." Techno-optimists, sometimes calling them-selves transhumanists or posthumanists (or something else), face these developments with confidence and great expectations. They predict that nanotechnology, biotechnology, information technologies, and cognitive science (the so-called NBIC technological disciplines and approaches)

will evermore intertwine and strengthen one another (known as NBIC convergence), which, they believe, will lead to numerous, cross-border technological achievements that we currently cannot even imagine. These technological developments will enable the engineering of our evolution and the radical enhancement of humanity. NBIC-convergence will, they predict, reinforce and accelerate technological developments to such a degree that a completely new human will ascend. They anticipate extraordinary posthumans equipped with smart skins, meta-brains, X-ray and infrared vision, and super intelligence. Upcoming technologies will, they believe, optimize cognitive performance, sensory functioning and even mood management. Their optimism seems to be limitless: technological developments will bring us, among other great things, ultra-intelligent nanobots that continuously detect, diagnose and cure diseases in our bodies, regenerative medicine that prevents biological decline and extends lifespan and genetic modification that not only makes us stronger and smarter but also limitless creative and permanently happy. Critics of these technologies, sometimes characterized as "bio-conservatives," are much less positive. They predict a shift from humans using these intimate technologies to control and regulate their lives to being controlled and determined by them. They also believe that these technologies, by changing our bodily and cognitive features and capacities, will violate our human nature and form a threat to our basic values (Kass 2002; Fukuyama 2002; Habermas 2003).

In addition, it can be observed that these new devices do not work independently but are increasingly connected and communicate with one another. This connectedness enables that the collected data and measurements regarding our behavior can be shared, which makes it possible to more successfully monitor and influence our thoughts, actions and habits. This development is currently bringing about the so-called Internet of Things and Ambient Technology. The influence of these new technologies is particularly effective because they are increasingly interwoven into the texture of our daily life and embedded in our ecosystem. These new networks seamlessly pervade our lifeworld, demand our attention, make us increasingly dependent on what they have to offer, are incorporated by us, merge with us and, some authors say, eventually become us (Turkle 2004; de Preester 2011). This way, these technologies are becoming ubiquitous, that is, everywhere and nowhere. Disappearing from view, not only are they increasingly escaping our control but we are also not aware of their influence – making it only more effective. By influencing our biology, chemistry and psychology much more directly, their influence appears to come from inside, indistinguishable from inner drives (Zizek 2009, 2012). Techno-optimists highlight the advantages of this interconnectedness and seamlessness of devices, whereas critics consider it an additional and even more urgent danger.

Understanding Human-Technology Relations

Different theories and approaches have been developed to understand the relation between technology and the world as well as the increasing influence of technologies on humans and their lifeworld. Focusing on how these theoretical approaches frame the relation between inside and outside allows roughly grouping them into three accounts: an instrumentalist, a determinist and an interactionist. Approaches that take a movement from inside to outside as a starting point often adopt a so-called instrumentalist perspective of human-technology relations, which corresponds to common-sense approaches. The instrumentalist perspective understands technology merely as a neutral means developed by human beings to achieve certain goals. It is human beings who have intentions and form interpretations of the world; technologies are always secondary to those intentions and interpretations. They function as instruments and means to control, manipulate and change the outside world. A spear is seen as a tool that can be used to fight, defend against or hunt animals in order to survive; a house is seen as means that protect humans from harsh natural elements, preventing them being forced to live as nomads; a university is seen as an institute that enables the increase of knowledge and skills in order to prosper and live a good life.

Approaches that take a movement from outside to inside as a starting point often adopt a so-called determinist perspective of human-technology relations. They observe an increasing influence of technologies on humans. This perspective does not see technology as a mere means for achieving purposes set by human beings in order to control, manipulate and change the outside world, but as an autonomous force that cannot be controlled or regulated (anymore) by human beings. Jacques Ellul (1977) has argued that the impact of technology has become so great that it is transforming society into a technical system. And according to Günther Anders (1956/2002), the products that we make immediately escape our control and incite practices that we cannot predict. We are always behind the technical products that we make. He goes so far as to say that at the end of the 20th century, it is not the human being but technology that is the actual subject of history, people have become the product of the products they have produced. This determinist perspective originates in pessimistic views of the effects of the industrial revolution; the transition to new manufacturing processes, including iron production processes, new chemical manufacturing and the large-scale use of steam power, invoked the fear that humans have become cogs in giant machines. But the same perspective also manifests in other forms of determinism, like the view that it is our environment that shapes our behavior and personality (situationism) or that it is our biology and in particular our brain that is in control of our actions and decisions (brain-determinism).

These deterministic, outside-to-inside approaches currently have the above-mentioned new and emerging technologies in their sights, including things like nanoengineering, genetic modification and Artificial Intelligence. The direction has reversed suddenly, and technologies are making an inward turn; although these new technologies are often presented as aiding devices or as means of communication and interaction, these theoretical approaches attempt to reveal that humans are being controlled and modified by the technologies. Instead of technologies being means to realize goals that humans have set, humans are portrayed as means that are used to satisfy their goals, as vehicles that sustain and multiply these technologies (Zizek 2010, 132).

The instrumentalist and determinist perspectives should not be seen as consecutive, although in particular periods one or the other could be more dominant. Their proponents often defend them synchronically dependent on how they understand and value the effects of particular technological developments in particular contexts. These perspectives are attempts to make sense of the role that technologies play in our world, which can be evaluated as more or less sound. From a third perspective, which I heuristically call an interactionist perspective, the problem is that both poles are too one-sided. From this perspective, the instrumentalist approach in neutralizing the role of technology fails to see how technologies themselves help to shape human goals. The clock is not merely a neutral instrument that enables realizing a goal independently set by humans such as getting to work on time: it has synchronized the lives of masses of people, optimized the division of labor and ultimately created the place where we have to get to "on time" to start with. Without the clock, the whole Taylorian production process would be unimaginable (McLuhan and Gordon 2003). But from an interactionist approach, the determinist perspective is not a satisfactory position either, since it ignores the fact that technologies are always embedded in use contexts, and in order to function, they need to be interpreted and appropriated by human beings. Mobile phones and algorithms greatly shape our behavior and interactions, but can only do that if they are properly and often (but not always) deliberately integrated in our society and entrenched in our cognitive processes.

In order to sufficiently appreciate the proposed interactionist perspective, it is important to recognize that it aims to overcome the *categorical* separation of human beings and technology assumed in both instrumentalist and determinist approaches. The starting point is neither a movement from inside to outside nor one from outside to inside, but rather humans and technology are seen as reciprocally and mutually shaping one another. Different approaches could be related or even brought under the heading "interactionism": Technological Mediation Theory (TMT) analyses how technologies help to shape relations between human beings and the world, including how they influence

moral actions and decisions (Ihde 1990; Verbeek 2005, 2011). Extended mind thesis (EMT) examines how technological objects in the external environment could be utilized by the mind in such a way that they become extensions of the mind itself (Clark and Chalmers 1998). Material Engagement Theory (MET) studies how human constitution has always relied on material culture, particularly focusing on how material things transform and rearrange the structure of cognitive functions (Knappett 2005; Malafouris 2004, 2013). These are certainly not the only views that could be characterized as "interactionist," but I highlight them because the approach that I will develop in this book strongly concurs with or challenges certain elements contained in them. It must be clear that many other approaches that focus on human-technology relations could be regarded "interactionist," including Callon and Latour's Actor-Network Theory (Callon 1986; Latour 1992, 1993), Feenberg's critical theory of technology (Feenberg 2002), Haraway's cyborg theory (Haraway 1991), Sloterdijk's immunology thesis (Sloterdijk 1998) and Stiegler's philosophy of originary technicity (Stiegler 1998).

The approach that will be developed in this book falls within the interactionist perspective and will also directly or indirectly critically use and incorporate insights derived from the above-mentioned and other theories. However, it significantly distinguishes itself from those approaches by its method and focus. Although it will very seriously take the "principle of charity" (Davidson 1984, Chapter 13) as leading in criticizing opposing views and developing an alternative, it will also show their shortcomings. Existing "interactionist" views in philosophy of technology have provided sophisticated accounts of the influence of technology on humans from different angles, but have insufficiently examined the essentialist and dualist views that they attempt to overcome. The lack of thorough insight in these views prevents recognizing the stubbornness of the essentialist and dualist view on the relation between human beings and the technological world. As a result, current interactionist views have proved insufficiently able to uncover relevant presuppositions and blind spots in their own positions and end up rehabilitating the very framework that they are aiming to defeat. Overcoming the categorical separation of human beings and technology requires a comprehensive understanding of how that separation came about in the first place.

Successfully seeing beyond the categorical distinction between technology and humans requires, first of all, grasping on which grounds humans can be distinguished from their (material and technological) environment. In a constructive critique of Platonic-Aristotelian essentialism, that is, the idea that the real is ultimately invariable, independent and one, modern philosophy, guided by Descartes, has appropriated that essentialism and developed an inside-outside dualism, a dualism of an inside realm of ideas and an outside world of material things, including technologies. In order to understand how technologies could radically intrude in the human, it

is pivotal to adequately grasp what that essentialism and dualism entail as well as how they could be dismantled. Inspired by Friedrich Nietzsche, the interactionist view that will be developed will debunk the concept that attributes rigid *a priori*, essential qualities to the human and acknowledge that all seemingly steady coordinates for understanding the human can be "historicized," "contextualized" or "deconstructed." This essentialist conception is not only historically situated in the philosophy of Aristotle and Descartes but also revealed as a tendency found in contemporary approaches to the human, including approaches in philosophy of technology. From the proposed perspective, the human cannot be understood as an independent and isolated entity. When the social and material environment of the human changes, the human changes. Taking the interactionist perspective all the way, the human is characterized as a relational entity: there is no human without and prior to its relations; a human comes to be what she is and only exists by virtue of her relations and interactions.

Conceptualizing the view that relations and interactions are not additional but constitutive – that is, you do not have first relata (human and world) that then relate to one another but that rather the relation makes the relata – is quite a challenge. This challenge cannot be accomplished if at the same time the inside-outside dualism is not sufficiently grasped and dismantled. Inspired by the American pragmatist Charles Sanders Peirce, the interactionist view that will be developed is very radical and will reject the idea that thoughts can only be attributed to humans and their brains while matter can only be ascribed to the outside world. Instead of distinguishing inside and outside on the basis of mind and matter, an account will be provided in which mind and matter are structural dimensions of reality as such. Again, the distinction between an inside mind and outside matter is not only historically situated in the philosophy of Descartes but seen as a tendency found also in contemporary approaches to the human, including approaches in philosophy of technology.

Dismantling the view that the human contains an *a priori* given essence or core that fixes her nature and enables detaching herself from her material and technological environment as well as revealing how this assumption is covertly preserved in many contemporary views will enable a more profound understanding of what it means that technologies radically intrude in the human. However, initially this insight hardly provides a solution to a problem. On the contrary, it only highlights the urgency of the problem faced. If the human is shaped in interaction with various mutable and increasingly technological environments, is it at all possible to influence and control those interactions, and if so, what form should we impose on them?

Extimate Technology and Self-Formation

Overcoming the inside-outside dualism allows, first of all, acknowledging that new and emerging technologies do not unleash a situation without

precedent, but rather reveal that the human has never had a closed inner realm, detached from the outside world. They cogently disclose that the human has always found herself in what Lacan (2005, 2014) calls an "extimate" relation with her symbolic and material environment and can only discover and develop herself in and through interaction with that environment. This does not mean that, with the advent of the latest technologies, there is nothing new under the sun. New and emerging technologies are enormously expanding the possibilities to "intrude" in the human. Although the human has always been "directed outwards," these technologies are immediately intervening in her own bodily and mental constitution. Besides tradition, education and culture, now technology is becoming a dominant force in self-formation processes.

New and emerging technologies increasingly shape who we are and how we experience ourselves. They can no longer be identified as external technologies that we independently relate to. These technologies, as "another within," cannot be simply externalized, but are increasingly becoming something that is and is not us at the same time. They are increasingly becoming "extimate"; they are both intimate and external, both familiar and strange, both disruptive and lifesaving, both common and artificial (see also Zwart 2015, 2017). Their ambiguity also explains why they can be experienced as uncanny. Because they can be so deeply embodied, they can become an issue of intense and intimate concern in their own right; the pacemaker, iPhone, health bracelet and nanopill are not just functional tools but increasingly articulate who we are or want to be. "Extimate technologies" are actualizing the potential "extimacy" that has always been a dimension of the human to such a degree that they are becoming a permanent part of her fabric. Since extimate technologies increasingly *make* us who we are and, hence, cannot be simply controlled by us, the question of whether we can and how we ought to form ourselves becomes evermore urgent. Answering this question requires investigating if and how the human can *critically relate* to the worldly, increasingly technological influences that are profoundly shaping her actions, decisions and personality.

From a radical interactionist perspective, the human, despite her extimate character, does not necessarily have to completely coincide with and be reduced to particular antecedents that shape her conduct. Completely reducing the human to her particular antecedents and rendering her conduct entirely determined by them would reintroduce the inside-outside dualism again, making the inside completely dependent on the outside. Inspired by various philosophers, and especially Nietzsche, Peirce and Lacan, I will demonstrate that the values, goals and ideals that humans encounter in their environments not only shape their identities, but can also enable them to *critically relate* to their present state. Not completely coinciding with her present state allows the human, at least potentially, to make a distinction between who she is and who she

wants to be, that is, a distinction between, on the one hand, biological inclinations and short-term decisions and actions, and, on the other hand, longer-term goals and ideals.

However, the idea that the human does not completely coincide with her present state and is able to form herself by virtue of longer-term goals and ideals should not be conceived as a simple "escape route." Finding in societal values, goals and ideals a foothold in order to critically relate to our present state and the factors that have shaped it raises many different and additional issues and problems. These difficulties have not only been subject to philosophical inquiry but have also been a source of inspiration, reflection and entertainment in film and literature. Different levels of difficulty can be recognized in the way these issues are depicted.

In mainstream popular Western media, these problems are often still portrayed as relatively innocent and bridgeable. For example, coming-of-age films such as *400 Blows, Rebel Without A Cause* and *The Breakfast Club* show in different ways the difficulties and resistances that young people encounter and the resolutions that they find in their search for their true selves. In *The Breakfast Club*, for example, five high school students, all different stereotypes, report for Saturday morning detention and are assigned to write an essay about who they think they are. At the beginning of the day, the "nerd" diligently attempts to get something on paper but does not go beyond "who am I? who am I…?" During the day, their interactions with each other enable them to discover common values, goals and ideals, which help them to come to terms with who, what and how they are and want to be.

In more transgressive fiction, writers like Vladimir Nabokov, William S. Burroughs and Bret Easton Ellis do not depict generation gaps and peer pressure in the quest for a suitable identity, but show in different ways the inability or reluctance of certain characters to identify with the basic values and norms of society. The idiosyncratic mindsets and behaviors of the portrayed protagonists outrage and often violate existing morals and sensibilities. In contrast to "coming of age" stories, societal confrontations and struggles are not valued as necessary conditions for discovering and developing a more mature self, but rather are seen as suffocating restrictions and obstacles that prevent the emergence of or eradicate the existence of unique, divergent selves. Here not so much the importance of developing a socially valued and accomplished identity but rather enabling a *singular* identity is highlighted.

Literature that focuses on the implications of postmodern societies that have undergone a collapse of so-called "grand narratives," including Christian morals, shared fundamental values and homogenous cultural traditions, poses an additional problem: novelists like Ivan Turgenev, Samuel Beckett and Franz Kafka show in different ways how society seems to be no longer capable of offering sufficient orientations or "affordances" that could guide or support individuals in their efforts to

establish a coherent, stable and accomplished identity. Here the inability to hook into a society, culture or tradition that has paved the road (with or without hurdles) for identity formation does not only apply to individuals who have deviant and idiosyncratic preferences and values, but also to those who would like to blend in and to form a "mainstream" identity, precisely because what is "mainstream" or "a socially valued and accomplished life" has become ever less clear and obvious.

While the novels and movies mentioned so far mainly focus on how (deviant) *ideas* and *ideals* shape or fail to shape certain identities, there is also an ever-growing body of popular media and literature that deals with the influence of *things*, and especially emerging technologies on what we consider our selfhood, which is of course the main theme of this book. From Mary Shelley's *Frankenstein* to Philip K. Dick's *Do Androids Dream of Electric Sheep?* and William Gibson's *Neuromancer* the question of what makes and goes beyond a human has been explored. Movies such as *Gattaca, Avatar* and *Her* explore how biomedical technologies and ICT's could disrupt and displace established biological, social and technological configurations and create completely new life forms, such as genetically enhanced or digital selves without material bodies. These prospects as envisioned in popular media are echoed, as I have indicated earlier, in achievements, developments and predictions in science and technology. I have mentioned a number of new and emerging technologies as well as the advent of NBIC technological disciplines and approaches, which are so intrusive that they are increasingly shaping the very fabric of our selfhood.

The question that has to be raised now with regard to these intrusive technologies and the appeal to goals and ideals is: if new and emerging technologies are increasingly shaping the very fabric of our selfhood, will they not also affect and challenge our values, goals and ideals? Put differently, if new and emerging technologies are disrupting established biological, social and technological configurations, will they not also affect and disrupt the very standards, values and ideals that could be used to evaluate whether these new identities and life forms are socially admirable and individually desirable? We see here that an appeal to longer-term values, goals and ideals does not immediately provide a "way out"; even if it would enable the possibility to impose a particular form to our interactions with our environment, it does not yet make clear how this form could be considered "good." In the wake of univocal standards being challenged and disrupted by new ideas and ideologies as well as by new and emerging intrusive technologies, the question of how to form ourselves becomes evermore acute. How is it still possible against the displayed background to sustain an ideal of "good self-formation?" What would "good self-formation" mean, if anything?

The appeal to longer-term goals and ideals does not have to entail reclaiming accessibility to actual *a priori* and univocal standards by virtue

of which we can establish what "good self-formation" is, as I will attempt to show in this book. In fact, this would be yet another (concealed) attempt to rehabilitate the inside-outside distinction and situate *a priori* standards in an independent inner realm, shielded from external influences and disruptions (which lie at the heart of Cartesian philosophy). The most important merit of this study is, as I have indicated in different ways, to show how technologies are becoming "extimate" and to maximally recognize that the human, including her value systems, is compromised and unsettled: technologies are increasingly becoming both the necessary condition for forming a stable social self and an obstacle that prevents her from reaching the singularity that she seeks. The longer-term goals and ideals that will be put forward in this study aim to recognize the extimacy of the human and still offer direction. Important in this respect is that these longer-term goals and ideals are virtual and always situated in the future, which, on the one hand, excludes the possibility of reinstating invariable, uniform and independent criteria to establish what is "good self-formation," and on the other hand, resists accepting that the self is no more than a plaything of contingent influences and forces.

The approach to longer-term goals and ideals developed in this book will be characterized not so much as an ethical but rather as an *esthetic* framework, which will be elaborated and ultimately articulated in terms of self-formation as "sublimation." Ethical perspectives that encourage understanding technological self-formation as technological enhancement often assume univocal standards for establishing what are "normal" and "enhanced" capacities. In contrast, the esthetic framework that views self-formation in terms of sublimation is motivated by the aim to indicate what "good self-formation" is without endorsing essentialist and dualist views of the self that consider it an independent and invariable unity, disconnected from its dynamic and sometimes even disruptive technological environment. At the end of the book, the developed framework, which will be coined Technological Sublimation Theory (TST) will be applied to different technological fields.

Research Question

We are witnessing the rise of new and emerging technologies that are deeply intrusive, increasingly shaping the very fabric of our selfhood, and challenging and disrupting established standards that had hitherto guided self-formation processes. Against the backdrop of these unsettling developments a fundamental question can be raised, which forms the main research question of this book: *how can and should we form ourselves in a world that is being saturated with technologies that are profoundly intruding in the very fabric of our selfhood?* Instead of "human" I will use "self" as a concept that is as neutral as possible and can be the starting point of an examination of what it entails.

This research question can be broken apart into three sub-questions that also set the outline of this book: first of all: *how should the "self" in "self-formation" be understood?* "Self-formation" seems to require a type of self-reference, which in its turn seems to presuppose a self that does not completely coincide with itself; I can only form myself if in one way or another I am able to refer to myself, which seems to presuppose a self that can rise outside itself. But how to conceptualize and understand an "I" that forms itself? What kind of entity is this "I" that not only can attribute qualities and properties to itself but also can make adjustments and changes in its make-up? And how to understand this self as a dynamic relational being without rehabilitating essentialist and dualist frameworks (in which, for example, a soul or a spiritual distinct entity is forming a body)?

Second sub-question: *are we able to form ourselves?* If the self that I propose is to a great extend shaped by its surroundings, what then secures that this self is not the product of its environment? Put differently, is the proposed view of the self not deterministic? Moreover, if our environment is evermore a technological environment and if technologies are becoming ubiquitous, is the self then ultimately not a repository of those technological influences? And if the self lacks sufficient control or autonomy, does that then not render the very notion of "self-formation" unintelligible?

Third sub-question: assuming that self-formation is possible, *in which direction should we form ourselves?* If the self has no fixed essence that in one way or another contains or can validate univocal principles or guidelines, how do we know what is "good" self-formation? If the self is shaped within various and mutable environments that, in addition, offer ever less orientation, how can we know that the form that we impose on ourselves is a "good" form? Which meaning can "good" in "good self-formation" have, if any? And if new and emerging technologies are becoming deeply intrusive and "extimate," shaping our very selfhood, as well as challenging and disrupting established standards that hitherto guided self-formation processes, and even introducing new norms and values, has not establishing what is "good" self-formation become impossible?

Outline of the Book

This book contains three parts that attempt to answer the three formulated sub-questions and, subsequently, the main research question: in the first part of the book I will show how from an essentialist and dualist perspective, on the basis of Aristotle and Descartes respectively, the self is conceived as an invariable, unified and independent entity. Then, I will explain why in this framework, radical novelty and interaction cannot be given a proper place, which ultimately makes it untenable. On the

basis of Nietzsche and Peirce, I will provide a radical interactionist perspective beyond the inside-outside dualism, which recognizes that the self discovers and develops itself in and through dynamic interactions with its surroundings. This latter approach offers a non-essentialist and non-dualist account of self-reference, which will also map the contours of self-formation.

In the second part of the book, I will begin by discussing two opposing answers to the question of whether self-formation is possible: Kant's view of the self as an autonomous being will be confronted with two contemporary representations of a determinist view of the self, namely situationism and brain-determinism. Then, I will argue that both opposing positions assume a false categorical distinction between an "inside" and an "outside" realm, and so fall back in or are not able to sufficiently overcome an essentialist and dualist view of the self. Overcoming this distinction will not only make it possible to do justice to the dynamic and relational character of the self, but also to recognize what I will call the "artifactual character" of mind and self, which is a view that will be elaborated in a constructive critique of EMT: technology inscribes, as I will argue, increasingly its "grammar" in the structure of our mind and self. Recognizing this mediating role of technology will reframe "free will" debates: the self is conceived as something that is neither completely determined by neurological processes nor external influences, but rather as a relational being that with the help of technologies such as fMRI can attempt to critically regulate its behavior by virtue of longer-term goals and ideals and form a more attractive self, which is, I believe, a better index of freedom. At the end of this part, I will also reflect on the question of "how critical" the proposed notion of "critical self-identification and self-formation" really is and can be.

The third part of the book focuses on the question of how we ought to form ourselves. I will show that transhumanists and other techno-optimists have a seemingly clear answer to this question: if we are able to form ourselves, then why not also boundlessly enhance ourselves? Technological self-formation is, according to them, captured best as technological "self-enhancement." But I will argue that this conception of self-formation as self-enhancement is problematic for two reasons. First, transhumanists are confused about their own conception of the posthuman: they anticipate radical transformation of the human through technology and at the same time assume that the criteria to determine what is "normal" and what is "enhanced" are univocal, both in our present time and in the future. Inspired by Nietzsche's notion of the Overhuman, I will show that the slightest historical and phenomenological sense discloses copious variations of criteria, both diachronic and synchronic, of what can be considered "normal" and "enhanced." Second, I will argue that in our attempts to form ourselves – through education or technologically – we come across "something within" that

resists complete appropriation, something "uncanny" that we cannot simply mold as we please. I will elaborate this notion of the "uncanny within" with the help of especially Lacan and show that it renders an ethics of enhancement unintelligible. Complementing Lacan's notion of a structurally compromised "extimate" self with Nancy's view of the self as being "closed open" for intrusive technologies will enable recognizing how new and emerging technologies are becoming increasingly "extimate" and are shaping the very fabric of the self. I will propose that, instead of understanding technological self-formation in terms of making ourselves "better," "stronger" and "smarter," which, in fact, results in rehabilitating an essentialist framework, we should conceive it in terms of "technological sublimation," which could also redefine the very notion of "human enhancement." An esthetic conception of self-formation as sublimation will be developed based on insights derived from Peirce, Nietzsche, Freud and Lacan. This elaboration will result in an alternative theoretical framework for understanding what "good self-formation" in a technological world entails, namely TST. Finally, this theory will be applied to three technological fields: smart environments, brain imaging technologies and smart drugs.

Part I

What Is the Self?

Various attempts across a number of disciplines, including philosophy, psychology and neuroscience, have been made to disentangle the multitude of perspectives of the "self." Sometimes different, more or less well-defined terms are reserved to designate different dimensions or aspects of the self, such as "subject," "agent" and "person," or the "material self," the "social self," the "minimal self" and the "narrative self." Some philosophers (including Plato, Descartes and, more recently, Swinburne 1984, 21) have argued that we are partless immaterial substances or souls, or compound things made up of an immaterial soul and a material body. From a so-called animalist perspective, others have stressed that we are all the way biological organisms (Snowdon 1990, 2014; van Inwagen 1990; Olson 1997, 2003) or material things constituted by organisms (Baker 2000; Johnston 2007; Shoemaker 2011). Yet others have proposed that we are spatial parts of animals, in particular brains or parts of brains (Hudson 2001; Campbell and McMahan 2010; Parfit 2012). There are also philosophers who have argued that we are collections of mental states or events or bundles of perceptions (Hume 1978, 252; see also Quinton 1962 and Campbell 2006). Others still have tried to illustrate how the self or certain aspects of the self are constituted in narratives or actions (Velleman 1992; Schechtman 1996; Menary 2008). And still others have attempted to show that we don't really exist at all (Wittgenstein 1922, §5.631; Unger 1979; Russell 1985, 50; Metzinger 2003). Positions that attempt to establish "what is the self" are particularly, though not exclusively, exchanged in debates concerning "personal identity." This matter often deals both with the question concerning the features and traits that characterize a self and the question of how this self, despite all the changes that it experiences through time, remains the same person. Contributors to that debate have indicated that in order to identify the essential traits of a person, we need to know which changes she will and will not survive (Shoemaker 1984, 70; Garret 1998, 12).

Often reviewing from a critical perspective the more established positions – some of them touched upon above – fractured and fragmented ideas about the formation of self in social interactions have been developed. From this perspective we are not so much already "something" but

our decisions and actions, in particular social contexts, are continuously shaping who and what we are as well as what we could be and ought to be. Instead of considering the self as a pre-established entity, the self comes to know itself and develop itself only through its interactions with physical and social environments.

In different ways, Nietzsche has attempted to overcome the essentialist view that the self has a pre-established essence or nature situated in particular faculties or parts of its constitution. Instead, the self in his view is characterized as a challenge and a mission that needs to be continuously developed and regained in an intensive struggle with its social, material and especially moral environment. The pragmatist accounts of Peirce, Dewey and James explicitly support an externalist conception of the self: instead of identifying it with the mind, soul or brain, or being something "in the head" or "in the brain," the self is distributed through embodied practices into the social and material environment (see Gallagher 2011, 2013; Gilihan and Farah 2005).

Within pragmatism, George Herbert Mead is probably the philosopher who comes closest to the Peircean perspective that, in conjunction with Nietzsche, I will use to develop a radical interactionist view. The Peircean expressions, like "[w]e become aware of ourself in becoming aware of the not-self" (CP 1.324) or "the inner world, apparently derived from the outer" (CP 5.493), could have also been uttered by Mead. His symbolic-interactionism has had a great influence on many interactionist and constructivist theories regarding the self that were developed within the confines of philosophy, sociology and social psychology. We can find especially in the discussion regarding the so-called "dialogical self," which has been to a great extent influenced by Mead, interesting relations to Peirce's approach to the self (see also Wiley 2006).

Another influential thinker that could be mentioned in this context is the Russian developmental psychologist and founder of cultural-historical psychology Lev Vygotsky. Hegel could be an explanation for remarkable similarities between Mead and Vygotsky, who both were strongly influenced by him (Glock 1986; van der Veer 1987). Many other positions that have defended a relationalist or interactionist view of self and world could be mentioned, including Simondon's transindividualism (1995), Kaipayil's ontological relationalism (2009) and Barad's intra-actionism (2007).

It must be clear that the aim of this part of the book is not to provide an overview of different accounts of the self or a typology of its most important aspects. There are several encyclopedia and handbooks that fulfill that need (e.g., Noonan 1989; Gallagher 2011). What is needed here is providing the basic building blocks of an interactionist view that are required to develop a convincing alternative for an essentialist and dualist account of the self. This is a challenge that should not be underestimated because an interactionist approach raises many questions and

difficulties of its own. There is a lack of consensus about the analytic and normative value of relationality, and difficulties regarding the question of how to achieve consensus. Foremost the key idea of "relation" remains unclear and is difficult to conceptualize. Not seldomly, in conceptualizing a "relationalist" or "interactionist" view, essentialist elements are recaptured through the backdoor. It is not easy to conceptualize a "relation" without assuming first "relata" that enter into relations with one another, as, for example, is confirmed in Kaipayil's view:

> Relationism is not anti-substantivism. On an anti-substantivist view, things are not objects in their own right, but only events dependent on other events for their existence. Even if we grant the argument that relations are ontologically more fundamental than entities themselves, the question is, if there are no entities with some enduring substantivity, how do relations themselves exist? Relation is 'holding' between two or more things. If entities disappear, relations will also disappear (Kaipayil 2009, 8)

Unlike this view, I will attempt to show, on the basis of Nietzsche and Peirce, how from a more *radical* interactionist view, relations and interactions are ontologically primary and entities ontologically derivative and, at the same time, durability and regularity can be accounted for (see also Wildman 2010). Providing a credible interactionist view requires accurately elaborating how the essentialist and dualist view is reasoned for and justified, which I will do particularly on the basis of its most important representatives and founders, namely Aristotle and Descartes. An accurate and sufficient understanding of this account is pivotal, also because it is so persistent and stubborn, and is being rehabilitated in different views – also in interactionist views – that assume they have overcome it. In part two and three of this book, I will especially show that in positions that try to understand and evaluate the impact of technologies on our self, essentialist and dualist accounts of the self are far from obsolete or overcome.

   This part of the book contains two chapters. The first chapter will investigate essence ontology and show how it was appropriated in dualist conceptions of the self that have strongly shaped Western anthropology. After revealing for our thesis relevant problems and flaws in the essentialist and dualist approach, I will propose in the second chapter an alternative, more adequate grounding of the notion of the self, which I will call "event ontology." On the basis of this ontology, I will develop an interactionist view of the self that will display that the "self" must be conceived as "self-formation." In part two of this book, this notion of the "self" as "self-formation" will help us to provide an account of how technological self-formation is possible without having to rehabilitate an essentialist view of the self.

1 The Essentialist and Dualist Self and Why It Cannot Be Sustained

In this chapter, a refined and detailed account of the essentialist and dualist approach of the self will be provided. After having touched upon Plato's concept of *ousia*, I will discuss essence ontology on the basis of Aristotle's notion of *ousia* because he is the first philosopher who systematically explains in what sense "essence" is the first and most fundamental category and how it relates to other less fundamental modes of being – *ousia* is the Greek word for "essence"; "substantia" or "subiectum" is the Latin translation or employment of *ousia*. Then I will show how Descartes adopts the features of Aristotle's notion of "essence," but attributes them primarily to the self. He also highlights "independence" as the most important feature of "essence," which will have, as we will see, far-reaching implications for how the self is captured in Western thought. After having recaptured the most important features of the essentialist and dualist self and pointed out why it so alluring, I will show why this view cannot be sustained and indicate what is needed to develop an alternative, interactionist view.

1.1 Aristotle's Grounding of Essentialism

Plato's Concept of Ousia: Situating the Essence of Things in an Imperceptible World

Aristotle develops his notion of "essence" in a critical adaptation of Plato's view of "essence." To have a more sharply distinct conception of Aristotle's view, I will briefly discuss Plato's view of this concept. Plato is the first philosopher who attributes to *ousia* a philosophically significant meaning that will take root in our Western culture. His *Euthyphro* clearly shows what is at stake for him. Socrates addresses in this book the question of what makes a pious act pious. He believes that this cannot be individual acts. These are so diverse that they cannot offer, but rather require explanation and judgment. What makes a pious act "pious" must be, according to Socrates, identical in every pious act. Plato calls the identity-dimension of pious acts the "*eidos* as such." Because this "idea as such" is identical and uniform in every pious act, it can

function as a norm, standard or paradigm for what is a pious act (ER 11a; 6e).

Plato is very aware of possible objections against his view. In the *Cratylus*, he gives a voice to Hermogenes who defends the thesis that the correctness of words is based on agreement and convention. But at the same time, Hermogenes admits that one man is wiser than the other. Socrates argues that this would be impossible if "what appears to each would be true for each" (CR 386c–d; 384d). Things, therefore, must have a fixed *ousia*, and cannot depend on how they relate to us but must rely on their own nature (CR 386d–e). For Protagoras, another opponent of Socrates (and another view that Plato attempts to debunk), not only words but the things themselves are subject to our discretion and imagination: the human is the measure of things. In a similar fashion, Plato attempts to show how "true and certain" knowledge resists and endures different contingent opinions, which, Plato assumes, indicates that the essence of things is not dependent on how humans relate to it but relies on its own independent nature (TM 37; PB 59c).

Plato repeatedly emphasizes that *ousia* ensures stability; change and decay are responsible for confusion and impurity (CR 411c; SY 184b; TM 29b; PD 84b; PB 59c; see also Natorp 1921, 125, 147–150). An example that he gives in *Cratylus* explains why this stability secured by *ousia* not only functions as a standard but also as a standard for perfection. Plato attempts to show that a carpenter who builds a new shuffle will not take the broken shuffle as an example or model. He will "look up" to the *eidos* that also modeled the old shuffle. This invisible *eidos* is considered as the most real and perfect shuffle, which remains identical irrespective of the very different but never completely perfect shuffles that we see with our senses (CR 389b; see also ED 285–288). These kinds of examples clearly illustrate that, according to Plato, perceptible things are always less perfect than imperceptible Ideas. Perceptible things are, in contrast to Ideas, never what they ought to be (PD 74d–e). Plato goes even further: perceptible things are to such a degree not identical with themselves that they barely deserve to be considered as "real." They have "being," in so far as they partake of imperceptible, perfect Ideas that contain their essence (PD 101c).

Aristotle's Concept of Ousia: Situating the Essence of Things in the Things Themselves

Aristotle rejects Plato's degradation of the perceptible world to almost nonbeing as well as his doctrine of Ideas. His main argument for this rejection is that Plato's Ideas and his notion of "partaking" neither explain the state of affairs in the perceptible world, nor increase our knowledge of it (MT 991a19–26; 1003b16–19). In order to understand "how things are," Aristotle claims, we do not need to appeal to entities outside them,

but should rather make them and the way we talk about them the subject of research (MT 1030a11–14). At first sight, Aristotle seems to radically depart from Plato's ontology, but on reflection, as I will show, he endorses the basic features that Plato attributes to "essence."

Essence as Unchanging Ground

In the *Analytica Posteriora*, Aristotle distinguishes two types of judgments that he illustrates with the following examples: "The white is a piece of wood" and "This piece of wood is white." The first proposition he considers "improper," because this type of judgment does not adequately express what is "underlying." The second example, according to him, does not have this deficiency and is considered "proper": it is clear that the piece of wood, and not the white, is the "underlying." In addition, Aristotle discovers that "being" is an "analog concept" and that not all modes of being have the same meaning and value (see Seidl 1989, XXXIVf). Denotations of the "underlying" refer to "what" something is, which can also be attributed other characteristics, such as "how big," "where" or "when" it (the "what") is. However, it does not work the other way around: the "what" cannot be attributed to those other characteristics (AP 83a24–30; 83b22–23). Within proper propositions, we therefore need to make a distinction between essential denotations that refer to the *ousia* and less essential predicates that require an *ousia*. Aristotle, hence, understands *ousia* as "independent ground (*hupokeimenon*)" of less essential qualities or properties (MT 1003a33ff). In his words: "The things not said of a subject I call things in their own right [*kath hauta*], and that are said of a subject I call coincidentals [*sumbebekota*]" (AP 73b5–8; PY 185a31–32; MT 1028b36–37).

The *ousia* has primacy over other modes of being (MT 1004b9; see also 1051b22–28; DA 430b26–29): it has primacy in an ontological sense because it can exist independently; it has primacy in a conceptual sense because defining coincidental or accidental predicates is impossible without including what something is; it has primacy in an epistemological sense because we know something better if we know what it is more than if we know how big or how heavy it is (MT 1028a30–38; V11, 1019a2–4). If there is no *ousia*, then there is nothing (MT 1028a18–28; 1050b19; 1059b30; 1071b5–6).

The *ousia* as an independent ground is also what remains the same and is therefore resistant to the change of coincidentals (see Stegmaier 1977, 43). Rejecting Parmenides' qualification of becoming as not-being (I will explain this later), Aristotle defends a view of being from the perspective of his notion of *ousia*. He uses the following example to illustrate his view. We can say '"the human becomes musical," but also "the non-musical becomes musical." These propositions indicate that in becoming, we must make a distinction between what changes and what

remains the same: the human that becomes musical remains human but the nonmusical does not persist. From this, Aristotle concludes that in every type of becoming, there must be something underlying that is persistent and undergoes change (a human), whereas that what does not persist (being nonmusical) is being replaced by its opposite (being musical) (PY 190a35b1). In the transformation from black to white, to give another example, there must be something that remains the same and that can transform itself from black to white because black as such cannot become white. Change always presumes "something" that changes. The *ousia* is the invariable ground that makes change possible (MT 1069b3–7, 1042a32–34; see also König 1967, 16).

Essence as Immanent and General Ground

The *ousia* as ground must explain something about the perceptible things, which Aristotle thinks Plato has not seen. Since nothing can emerge from nothing, as Parmenides has shown (Diels and Kranz 1956, 231–240), all things must, according to Aristotle, have causes. In addition, Aristotle believes that we really understand something if we know its causes. The causes of things are not things themselves, but are responsible for the emergence and development of things and make it possible to have knowledge of things (MT 981b30–33; 1013a17; 1070b26; AP 94a20–23). Aristotle understands *ousia* as a cause (see MT 1041a6–10; 1013a16f).

In his search for what kind of cause the *ousia* is, Aristotle identifies it initially as matter (*hyle*). Matter appears to be something underlying (*hupokeimenon*) and a subject or substrate of predicates (see MT1029a2; 1042b9; 1070a11; 1033a9f; 1033a24–34; 1016a17–28; 1055a30). On reflection, matter, according to Aristotle, does not suffice because it is completely undetermined in itself and can therefore not lay down boundaries, provide unity and separate a thing from other things (MT1029a23–27; 1035a8 and 1036a8f.; 1037b26f.; see also Ulmer 1953, 74). As a result, matter cannot provide knowledge about a thing and is as such unintelligible (Stegmaier 1977, 47). Matter relates to *ousia* as bronze relates to a statue. Matter *is* and *is not*; it is as potentiality, but not as reality. Matter is potential *ousia*, not actual *ousia* (MT1069b15f).

Aristotle further explains the insufficiency of matter by showing how it addresses our sense organs. Our sense organs, he attempts to illustrate, cannot capture unity because they register parts and pieces of things without being able to detect in them a common structure. To connect a piece or part to another piece or part, sense organs must "remember" pieces and parts and connect them to other pieces and parts. Sense organs, Aristotle believes, do not have this ability. When our sense organs record one part, they have no knowledge anymore of another part (MT 1052a24–26, a29; 1016a5f. and 18f; see also Stegmaier 1977,

53f). Although matter is a ground or cause, it is only a ground or cause of what is particular and individual; it is *principium individuationis* (see MT 1016b31ff; 1034a7f.; 1037a1f).

Sense organs cannot capture unity and distinctness (*choriston*). This ability, Aristotle believes, should be attributed to thinking. Thinking can capture a common structure by abstracting from a variable multitude of individual parts and pieces. Therefore, thinking is directed at the general and not at the particular (MT 981a15f. and 981b10f.; DA 417b22–23). Irrespective of something becoming, for example, smaller or bigger, thinking can capture its unity and identify it as the same thing and as different from other things throughout time and space (MT 1016a35ff).

Aristotle's distinction between predicates and the subject is analogous to the distinction between matter and the general: because sense organs are directed at matter, they can only capture coincidentals. What structures and unites those coincidentals and, hence, gives them definiteness, can only be grasped by the mind (MT 1017b35–1018a2; 1029b19f). In contrast to Plato, however, the general does not exist (spatial) outside things, but is immanent to things. It provides them definiteness by distinguishing the durable from the variable (MT 1029a27–30).

Essence as Form and Goal

Matter as undetermined potentiality requires something that determines it and ensures its unity and distinctness. Aristotle calls this principle of determination "form" (*eidos*).[1] However, the form should not be seen as a separate entity localized outside the actual thing. It is not separate from the thing, but rather *is* its determinedness. The same applies to the relation between matter and form: matter is not spatially distinct from form, but can only reveal itself in it (MT 983a26–32, 1013a24–35; 1032b1f; PY 194b23–35; DA 412b6–8, 413b13–16). Bronze can only reveal itself in, for example, a bronze disc.

Form cannot be spatially separated from a particular thing. However, it also cannot be completely immersed by it. As something general, Aristotle believes, it exists independent from a particular thing. The form of a disc, for example, cannot exist without any materialization (e.g., in a bronze, wood or iron disc), but is not dependent on a *particular* materialization. According to Aristotle, it can also be conceived independent of a particular materialization. Although matter is not superfluous, it is clear that for Aristotle, form is more essential than matter. It is *ousia* in the strongest sense of the word because it ensures the identity of a thing, irrespective of variable materializations (MT 999a17–19; 1036a26–1037a5.; see also Stegmaier 1977, 70).

In addition, Aristotle states that matter is responsible for the "distance" between a thing and its essence, whereas its form coincides with its essence. Because there are infinite different ways of being insufficient

or deficient, there is an infinite multitude of particular things. Matter does not generate these variations positively, but the individual and the variable are the result of flaws that are caused by matter. Differences and changes are deviations. Matter is (negatively) responsible for variation and change, whereas form is (positively) responsible for unity and identity (MT 1033a20–22; 1037a33–b7 and DC 278a18).

We have seen earlier that matter is conceived as potential being, not as actual being. Insofar as matter *is not* (actual), it is the gap between a thing and its essence; insofar as it *is* (potential), it is directed at bridging that gap. The form (*eidos*) is, therefore, the aim of matter. Movement and change are attempts to realize the form of a thing. The form is the *causa finalis* of a thing. If we want to understand a thing, we therefore need to understand its purpose (MT 1050a7–10; PY 199a30–31; see also Berger 1968, 100–101). This has become known as Aristotle's teleological view of the world.

The epistemological implications of Aristotle's view on the relation between matter and form are far-reaching. Because matter is responsible for differentiation, it is in itself incognizable, which means that individual things cannot be defined: "If someone would, for example, wants to define you, he would say something like: skimpy or pale living being; however, that description might also apply to someone else" (MT 1040a6–7 and 12–14; see also 1036a8–9). The form makes it possible to define a thing, and makes it therefore cognizable, but only as something general.

Aristotle takes as a starting point for his investigation things as we know them in everyday life, but considers them insofar as they can be defined. Since the individual, according to him, cannot be defined, it is considered incomprehensible. It relates to the general form as a privation that must be overcome.

That the individual is subordinate to the general is also reflected in Aristotle's understanding of the Divine. God is the final ground of unity, determinateness and order. Unlike the form of material things, the divine no longer exists in conjunction with the sensory-individual and the changeable. The divine is free from all matter. It is without material parts and has no extension, so that it is indivisible. And because it is indivisible, it is imperishable; only compounded things can fall apart (MT 1003b17; 1028b27–32; 1071b20f). Aristotle's view of God as something that is completely (spatial) separate from the sensory-material brings him very close to Plato's doctrine of Ideas, which he rejected earlier (Happ 1971, 765 and 946).

The divine is the final source of all movement. If there was no initial "unmoved mover," there would be no change at all, and things would not strive for realizing their essence. Although Aristotle assumes that all that moves is moved by something else, he believes that God is not moving other things mechanically, otherwise it would be part of the world

and itself be moved. He argues that God has to move other things by being an object of desire; his divine stirs and brings other things in motion because it is longed for and loved, and is thought or imagined as desirable (MT 1071b14–17; 1072a25–26; 1005a35; PY IIX; DA 433b11–12). Just like God moves all other things by being a final goal, the form is the goal of matter (see Berger 1968, 103).

Essence as Invariable, Uniform and Independent

"Essence" is interpreted by Plato as a paradigm or example from which we derive the standard for what something is. Plato situates this example toward how our thinking is and should be oriented in an imperceptible reality. However, Aristotle situates this *ousia* of things in the things themselves. Crucial for him is the ontological distinction between *ousia* and coincidentals. He is interested in the difference between what is essential and what is incidental.

Although the differences between Plato and Aristotle are significant, there are also striking similarities (PY 192a20–29). Both characterize what is essential to a thing as invariable, uniform and independent, and what is inessential as variable, multiform and dependent. In addition, for both philosophers the essential is general, which can be captured by the mind, whereas the inessential is particular and individual, which is directed at sense organs. Moreover, they both believe that the individual is dependent on and can only be explained by the general. For Aristotle, the material is deficient and always strives for the ideal general form – something he believes Plato has not seen (PY 192a20–29). However, Plato in a quite similar fashion states that the perceivable is always measured by comparing it to the Idea, which is considered as a paradigmatic example (PD 74e; see also Berger 1968, 93–101).

Like Plato, Aristotle believes that the essential and ideal are invariable, uniform, independent and general. In contrast to Plato, he believes that the essential and Ideal (except God) should not be sought (spatially) outside the things, but rather in the things themselves.

1.2 Descartes' Elevation of the Self to an Independent Substance

Aristotle attempts to understand and explain how things are by studying how they appear to us and how we talk about them. In this investigation, he believes, eventually the final principles or causes of all things could be discovered. Hence, the first principles will be found last. Descartes' account has a different starting point and background. Unlike Aristotle, for Descartes it is no longer obvious that examining things as they appear to us will enable finding the first principles. At the end of the Middle Ages, different religious, social and scientific developments

had led to skepticism about our most basic theories, beliefs and intuitions. Descartes' philosophy represents a radical attempt to, once and for all, make an end to past, current and potential future doubts that are generated by the unstable "outside world." Inspired by the achievements of the natural sciences, he seeks a method that will enable him to find non-doubtable first principles. Everything else could then be derived from those axioms. Therefore, metaphysics should not come after "physics," as Aristotle believed, but it must precede all other sciences. Descartes' aim is not to find the first principles or causes of reality. He rather seeks a reliable foundation for the sciences (*Reg.* II:AT X, 362; IV:AT X, 377–378; XIV; see also *Disc.* II: AT VI, 362).

Descartes attempts to find a first absolutely certain ground on which all knowledge can be founded in a more reliable mental "inside world." Instead of worldly observations and investigations, Descartes relies on introspection. This change of perspective in the beginning of the modern era causes a fundamental shift in thinking about "essence." The "thinking I" discovers itself as first essence, and radically separates itself from the material world. Although Aristotle makes a distinction between different forms and levels of being, the idea that the human self is something independent from or outside of the physical world is foreign to him: all reality, including the human being, is explained by the principles of form and matter; these principles cannot be separated from each other, but intrinsically rely on each other. In contrast, Descartes defines the self as a thinking substance that can, as I will briefly elaborate, exist completely independent of the material world.

The Subjective Turn: Discovering the Self as First Essence

Science, according to Descartes, requires a certain and unshakable foundation which could only be discovered if reason is pure and attentive, that is, free from bias and hastiness. We can free ourselves from bias by challenging the prejudices that we have inherited in our education and upbringing. Hastiness can be prevented by abstaining from judgment as long as there is any doubt, and biases are generated by hastiness. Abstaining from judgment as long as there is any doubt in a structural manner proves to be the method that Descartes has been searching for. No other procedure or regulation besides this methodical doubt is allowed. This radical doubt will, Descartes expects, eventually overcome itself when it encounters a proposition that resists every possible attempt to challenge it (*Med.* I and II:AT VII, 17–34; *Princ.* I, 1–7: AT VIII–1, 5–7; *Disc.* IV:AT VI, 31–32).

At the peak of his experimental doubt, Descartes comes across something that by no means can be doubted: the reality of doubt itself. He can doubt everything, but cannot doubt the fact that he is doubting, and therefore, he needs to infer that he is thinking. And as far as he is

thinking, he, as the subject of that thinking, must exist. I think, therefore I am (*cogito ergo, sum*) is Descartes' first indubitable principle and unshakable foundation of all further scientific knowledge (*Disc*. IV:AT VI, 32; *Med*. II:AT VII, 25; *Princ*. I, 7:AT VIII–1, 7). Although the relation between, for example, "I walk" and "I am" or "I blow my nose" and "I am" is not less inferable, the premises "I walk" or "I blow my nose" are not indubitable facts. Those experiences could be easily challenged: they could be products of a dream or hallucination. However, it is impossible to contest that I am thinking, even if I would be dreaming or hallucinating. Hence, Descartes' *cogito ergo sum* does not rely on one but on two pillars, namely on the (indubitable) *experience* that he actually thinks and the (indubitable) *insight* that precisely because and insofar as he thinks it is necessary that he exists (*Resp*. V:AT VII, 53; *Princ*. I, 9:AT VIII–1, 7).

Descartes justifies his first certainty by claiming that it is based on nothing else than the ideas of clarity (*claire*) and distinctness (*distinct*). Then, he makes clarity and distinctness general criteria of truth: if we can see or think x clear and distinct, then x is true (*Disc*. IV: AT VI, 33; *Med*. III:AT VII, 35). Ideas (mental contents, contents of consciousness, images etc.) are "clear" if they are present and accessible to an attentive mind, if they are as it were in the center of our attention. An idea is "distinct" if it is precisely determined and sharply separated from other ideas, if it does not contain any elements that do not belong to it. Descartes calls unclear ideas obscure (*obscures*); ideas that are not distinct he calls confused (*confuses*) (*Princ*. I, 45–46: AT VIII–1, 21–22).

Clarity and distinctness can also apply to relations between ideas. That whatever thinks, must exist, is an example of a clear and distinct necessary relation between the idea of thought and the idea of existence. The validity of such necessary relations or eternal truths (*veritates aeternae*) does not depend on the existence of anything and must be therefore innate. Moreover, it is, says Descartes, obvious that any attempt to clarify these eternal truths would lead to embezzlement (see Vennix 1996, 159).

Since the ideas of clarity and distinctness are, according to Descartes, innate, it is possible for the human subject to close his eyes and turn inward to his own clear and distinct ideas, which enables him to reconstruct reality entirely independently of experience. This is the so-called "subjective turn" of Descartes' philosophy. Although we can recognize Aristotle's principle of unity/distinctness (*choriston*) in Descartes' principles of clarity and distinctness, there is also a great difference: unlike Aristotle, Descartes situates the principle of unity/clarity/distinctness not in things but in the mind. The criteria for truth are not derived from investigating objects in the world, but are rather found in the mind of the human subject. This "subjective turn" distinguishes Descartes's philosophy from that of Aristotle and all Greek philosophy and has brought about what has grown to be called modern Western philosophy (Löwith 1967, 25).

A Perfect God as a Warranty for True Knowledge

Descartes recognizes that his first certainty and his general criteria for truth are not sufficient to ground science. Further certainty is needed in order to ensure that the truth that we find in our mind is really true. Descartes needs a truthful God that can provide him that certainty. If we would not know that our first, fundamental certainty comes from a perfect being, then we would not have any guarantee that the ideas that we have found in our mind, regardless of how clear and distinct they are, are really true (*Med*. III: AT VII, 36). Although God is not known earlier than the *cogito ergo sum*, the insight in this first certainty only comes to full completion in the knowledge that God exists and that he is a truthful being (see Vennix 1999, 14).

Because it is the only certainty he has so far, Descartes' proof that God exists can have no other basis than the *cogito*, that is, the mind and its contents or ideas. That explains why he further examines the origin of those ideas and asks whether they all have the same status. Is there an idea in our mind whose origin cannot lie in the human self? After all, if there is an idea that could not have been originated or invented by the human subject, then that idea must have been induced by something else and then something else must exist besides the human self. According to Descartes, it is the idea of a perfect being (God) which cannot be produced by us, because something imperfect cannot create something perfect. That I doubt is already a mark of my imperfection. From this it follows that there must be something that has produced the idea of perfection in me. And because the idea of perfection has not been produced by me and cannot come from nothing, a being must exist that is as perfect as that idea, a perfect being that has produced that perfect idea in me (*Med*. III: AT VII, 40f; *Princ*. I, 17–18: AT VIII–1, 11–12; *Resp*. II: AT VII, 167). In addition, Descartes provides a so-called ontological proof of God, which is based on the idea that a perfect being would not be perfect if it did not exist; perfection already involves existence (*Med*. V: AT VII, 65f; *Princ*. I, 14–15: AT VIII–1, 10; *Resp*. 11: AT VII, 166–167; *Disc*. IV: AT VI, 36).

According to Descartes, the cause of our errors should not be sought in our capabilities, which are created by God, but in the way we use them. Errors arise when we fail to keep our will within the limits imposed by our mind. Errors are prevented if we refrain from passing judgment on ideas from which we have no clear and distinct knowledge. If we do have that knowledge, then we can be certain that the truth of those judgments is warranted by God's truthfulness (see *Med*. IV: AT VII, 59–60; *Princ*. I, 29–44. See also Löwith 1967, 28–29).

The Soul as a Thinking Substance

Descartes' first certainty is *that* he is (existence). The next question that he investigates is *what* he is (essence). Descartes states that he is

a substance whose whole being is nothing but thinking. To exist, this thinking substance does not need place, nor does it depend on something material. A substance is for Descartes something that for its existence depends on nothing else than itself (*Disc*. IV: AT VI, 32–33; *Princ*. I, 8: AT VIII–1, 7). Strictly speaking, only God is a substance, because only God, who is an uncreated substance, is completely independent. All other substances, which are created by God, are for their existence entirely dependent on God. Descartes, however, generally uses the term "substance" in a less strict sense, referring to something that for its existence depends on nothing besides God. Descartes relates "substance" primarily to independence. Unlike substances, properties always exist in something else; they are attributes of something else. "Thinking" is for Descartes a property or attribute that cannot exist on its own, but that necessarily presupposes something that thinks: a thinking substance (*res cogitans*) (*Princ*. I, 51–52: AT VIII–1, 24–25 and 11:AT VIII–1, 8). Descartes appropriates the Aristotelian argument that a property is always dependent on a substance.

However, Descartes does not say that he is a substance, which can be attributed to, among other properties, thinking. He considers himself a substance whose whole being is thinking. As a pure thinking substance (a soul) that can exist on its own, he does not contain anything material. The soul can be clearly and distinctly thought without, thereby necessarily including something material or corporeal. Therefore, the soul can exist completely separate and independent from the body. A thinking substance or soul is totally self-sufficient and autonomous (*Med*. VI: AT VII, 78; *Princ*. I, 60: AT VIII–1, 28–29; see also Vennix 1999, 16).

However, "thinking" cannot be thought independent of the idea of a substance. Of "thinking," no distinct idea can be formed independently of the idea of a substance. Therefore, thinking cannot exist independently of a substance. And this is true for all attributes (*Resp*. I: AT VII, 121, III: AT VII, 175–176 and IV: AT VII, 223). Moreover, ideas such as doubt, imagination, volition, sense and feeling cannot be thought independently from the idea of thinking. Therefore, they cannot exist without that idea. We cannot form an idea of doubting-without-thinking, while we can form an idea of thinking-without-doubting. Such attributes can only exist as modes of a thinking substance (*Princ*. 1, 53: AT VIII–1, 25).

According to Descartes, every substance has a main attribute that forms the core of that substance, and which all other attributes as modes depend upon. In this case, the soul is the substance of which thinking is the main attribute that is involved in all other attributes. Therefore, Descartes can say: "I doubt" (mode), therefore "I think" (main attribute), therefore "I am" (substance) (*Princ*. 1, 53: AT VIII–1, 25).

The Body as an Extended Substance

Besides the soul, Descartes distinguishes the body, which is part of the material world. His famous "ball of wax experiment" makes clear what is the essence of the body. The essence of the piece of wax cannot consist of its varying properties, such as shape, size, color, smell, taste and density. When the wax is warm, its shape, size, color and so on change. Nevertheless, one cannot deny that, says Descartes, it is the same wax. The essence of the wax must be something that is maintained through all these changes. What is variable and mutable is for Descartes incidental and therefore not essential. Here, Descartes fully endorses the Platonic-Aristotelian idea of being as self-identical and invariable. By putting all that is variable as nonessential between brackets, Descartes finally comes to the conclusion that the essence of matter can be nothing else than extension in three dimensions, namely length, breadth and depth. Extension is the main attribute or essence of the body, which Descartes calls "extended substance." Size, shape, motion, position and so forth are modes of extension (*Med*. II: AT VII, 30; *Princ*. II, 11:AT VIII–1, 46).

According to Descartes, this knowledge of the essence of the piece of wax is not provided by our senses because our senses can only register the varying qualities of the wax and not what is preserved in the wax. Our knowledge of the essence of the piece of beeswax as well as the material world in general, rests, according to Descartes, on a clear and distinct idea. Because extensiveness can be clearly and distinctly thought, it can exist completely independent, also independent of the mind. And because God is not a deceiver, it is assured that the external world exists also outside our mind (*Med*. VI:AT 7:79–80).

Descartes reduces physical reality to pure geometrical extension, to what can be measured and quantified. Sensory qualities as well as qualitative changes in the physical world have no objective reality, but must be reduced to the clear and distinct ideas of extension and movement. All physical changes are mechanical in nature, that is to say, they are, in reality, changes of place, local movements of particles extension caused by push and pull and can be explained in mathematical terms. Sensory experiences and sensations of pain, hunger and thirst, which exist only in the mind, merely have a practical function: they are useful for life, because they teach us what to do and not to do in order to live healthily (*Princ*. IV, 189: AT VIII–1, 315–316; *Le Monde* I: AT XI, 4–5).

It must be clear that the idea that in order to grasp the essence of things, thinking is more suitable than our senses was strongly influenced by Plato and Aristotle. However, here we find also a significant difference. Sensory qualities such as smell, color, taste, heat, hardness and softness exist for Descartes, unlike Aristotle, not in physical reality outside the mind, but only in our thinking. These properties are nothing

but the purely subjective responses of our mind to the movements, resting states or configurations of particles extension (*Princ*. IV, 189: AT VIII–1, 315–316; *Le Monde* I: AT XI, 4–5). These qualitative properties cannot be separated from our thinking because they cannot be thought independently of it. We see repeatedly how Descartes uses the idea of "independence" as the main criterion in his reconstruction and recalibration of scientific thinking.

Cartesian Dualism and the Concept of "Freedom"

For Descartes, the human being is an "assembly" of two essentially different substances, which could exist entirely independent of each other. The bodily functions of the human being could be explained entirely in mechanistic terms, which for Descartes also mean that the human body is without any form of consciousness (*Disc*. V: AT VI, 45–46; *Traité de I'Homme*: AT X1, 201–202). However, the functions of our rational life cannot be understood mechanistically and are not bound by natural laws; in Descartes' words: they are mode of a free soul that "renders us in a certain way like God by making us masters of ourselves," (*Passions of the soul*, art. 152). The human is a soul that is connected to a machine. Because both substances are self-sufficient, their interaction is mysterious (*Resp*. IV: AT VII, 227–228; *Disc*. V: AT VI, 59; *Med*. VI: AT VII, 81; *Med*. VI:AT VII, 81; *Lettre à Elisabeth*: AT III, 693, 695).

Although the human consists of soul and body, Descartes stresses, at the same time, we seem to identify ourselves primarily with our minds. The idea of the self as a conscious agent is very much dependent on his principle of introspection. "Introspection" refers to the capacity to access the contents of our thoughts in an authoritative (others cannot challenge our belief of being in a particular mental state), privileged (we know the contents of our own minds always better than we know the contents of the minds of other people) and immediate (knowledge of our mental contents is non-inferential and non-evidence based) manner, which radically differs from how we encounter the external world of material things (Macdonald 2007); because I have this privileged access to my mind, but not to my body and the world, I identify myself primarily with my conscious thought.

The discovery through introspection that I am a completely independent "thinking substance" assures that I have a free will. The correct use of free will is dependent on rational judgment and identified as the critical factor in the attainment of knowledge. Descartes explains this, proving being the father of rationalism:

> If (…) I simply refrain from making a judgment in cases where I do not perceive the truth with sufficient clarity and distinctness, then it is clear that I am behaving correctly and avoiding error. But if in

such cases, I either affirm or deny, then am I not using my free will correctly. If I go for the alternative which is false, then obviously I shall be in error; if I take the other side, then it is by pure chance that I arrive at the truth, and I shall be at fault since it is clear by the natural light that the perception of the intellect should always precede the determination of the will (*Med*. IV (AT VII 59–60))

The human possesses an innate, internal faculty by virtue of which it is an independent substance and can direct its will and decisions. I am free insofar as I have the ability to refrain from affirming confusion and falsehood: "the more I incline in one direction [towards truth and goodness] the freer is my choice" (*Med*. IV (AT VII 57–8)). For Descartes, freedom is self-control guided by pure reason; in cases of clear and distinct perception, the will is free, in that it is not confused and constrained by influences external to the mind but spontaneously affirms that which the inner mind puts forward.

1.3 Why the Essentialist Approach Is Alluring but Untenable

The Essentialist Self as One, Invariable and Independent

Aristotle and Descartes' views have been pointedly elaborated and presented as foundational representations of an essentialist and dualist approach of world and self, respectively. If we compare how these philosophers depict "essence," we find significant differences: in a criticism of Plato, Aristotle situates the essence of things not outside but in the things themselves; Descartes detaches the self as a thinking substance from the world and highlights "independence" as the most important mark of the real. Obviously, many other philosophers, including Leibniz, Spinoza, Kant and more recently Wiggins (1980, 2001), have contributed to the development, reception and critique of the notion of "essence" in different directions and forms. Although the divergent views, characterizations and valuations of this concept should be recognized, throughout its history recurring features are attributed to it: the "essence" of things is conceived of as "one," "invariable" and "independent" (see Aydin 2003 for a more extensive elaboration of the history of the concept of "essence").

   The essentialist approach is not only an expression of probably one of the most deeply rooted convictions in the history of Western thought, but it also seems to mirror a common-sense intuition. From this perspective, an essence secures that a thing can possess a basic, invariable identity, despite being subject to continuous change; it makes it possible that a thing can form a unity, despite having different parts and properties; and it enables that a thing can be separated from other things, despite being involved in various interactions. By virtue of these features,

essences can secure truth, stability and order in the world. They are also seen as the ground, substrate or cause of the phenomena that we perceive. Essences are the "furniture of the universe:" *if there are no essences, then there is nothing.*

This essentialist view has profoundly shaped how in Western thought the self is understood: although I change continuously, have different faculties and properties and am deeply intertwined with other people and things, I still assume that there is something in me that secures that what "I" think, do and experience can still be attributed to "me." This essentialist view of the self seems to further strengthen the common-sense intuition of how we relate to ourselves and others. Descartes' appropriation of essentialism through highlighting "independence" as the most important element also matches the liberalist and individualist view esteemed so highly in Western culture. In the modern Cartesian era, the "self," the "real self" or "what makes the self real" is conceived of as one, invariable and independent; strongly valuing the element of independence enabled radically detaching the self from its instable world and securing its rationality and autonomy. The self is primarily identified with an inner realm that is independent of the outside world. This inner realm contains the criteria for truth and secures the possibility to acquire reliable knowledge. The radical distinction between "inside" and "outside" has also far-reaching moral implications: it safeguards the existence of freedom and responsibility in a world that is determined by laws of nature. All these things considered the essentialist and dualist view of the self looks appealing.

The Essentialist Account of "Multitude", "Change" and "Relationality"

If essences are the bedrock of the world, and "essence" is characterized by "unity," "invariability" and "independence," how then does one value the opposite qualities "multitude," "change" and "relationality"? From an essentialist perspective, these qualities are considered to be less or not real because they cannot exist on their own: "multitude" is incognizable and associated with chaos and disorder; "variability" has only reality insofar there is something (invariable) that is subject to change; "relationality" presupposes things that can enter into relations. Nietzsche has stated that all philosophers from Plato until Schopenhauer believe: "What is, does not *become*, what becomes, *is* not…" (TI, 167).

The most extreme nullification of becoming can be found in a philosopher that has also left his traces in the views of Plato, Aristotle and Descartes, namely Parmenides. For Parmenides, change was a logical impossibility. He defines change as "becoming something else," that is to say, something that exists cease to exist; something else that did not exist comes into existence; and so forth. But something, a being, cannot become something else, another being, because in order to become

something else, it first has not to be. There must be a transition of something that is in a certain way to the not-being of it. But being cannot become not-being, because that is not. And from not-being nothing can emerge. There is no middle between being and not-being, something is or it is not. And change is not. Parmenides sums up his philosophy with the thesis "what is, is." He argues that all change is an illusion and that in reality nothing changes. Our senses make us believe that reality changes. But if we think logically, we must conclude that change is not a real entity, and that our senses deceive us in believing that there is change (Diels and Kranz 1956, 231–240).

In his response to Parmenides (and Plato), Aristotle grandly attempts to do justice to change and our experience of change. Without an explanation for movement and change, which we encounter in all facets of reality, nature, according to him, could not be understood. Also, being a biologist, Aristotle argues against Parmenides that change is not a transition from not-being to being, which Parmenides took for a logical impossibility, but a transition from possible being (*dynamis*) to actual being (*energeia*) (PY 200b12ff). Aristotle's introduction of the concept of "possibility" is a stroke of genius.

But then his notion of change is greatly relativized and weakened. Although Aristotle attempts to secure the variability in the world by the introduction of the notion of "possibility," he firmly holds on to the principle that every possibility is based in an actual reality. Change and variability are considered in Aristotle's view of reality, but these are always relative and additional. As coincidentals, they need something else by virtue of which they can exist and be known, they need an invariable, unified and independent substance, as we have seen earlier. Aristotle does not completely deny the existence of "multitude," "variability" and "relationality," but he also does not consider them "essential." Because they cannot be positive, persisting constituents of what a thing is, they are not a genuine part of its identity. Applied to the human self, Aristotle recognizes that we in one way or another change, interact and engage in different roles. However, these facets of the self are not constitutive for what the self "really" is and should become. What it "really" (potentially) is and should become ("a rational being") is already inscribed in its essence, which is part of the species that it belongs to. The only (negative) reality that is ascribed to these elements is a transitory one: they express that a thing is not yet what it is in essence, which is also what it eventually should become. Every type of becoming derives its meaning and value from a stable destination that it wants to reach, which is a view that lies at the heart of Aristotle's teleology.

Every object and every being that emerges in the material world is, according to the teleological framework of Aristotle, the realization of a predisposition that was present in advance. Without that predisposition, the realization of that object or being would not be possible.

A predisposition or proclivity can only be brought to completion by something that is already complete; a predisposition cannot do this by itself (AP 73b5–8; PY 185a31–32; MT 1028b36–37). Aristotle finds this completed-being in the *form* of an object, which is also the common attribute, the *eidos*, of all the members of a particular species (MT 983a26–32; PY 194b26–27). The essence of a thing is, at the same time, its goal: a thing strives to become what it essentially is (DA 434a31–32). Change is, therefore, always a sign of indigence, of the unfinished, of life that has not yet matured into a particular shape (MT 1048b28–29; DA 417a16–17). Becoming is not an autonomous process, but rather a mere transition that can only find completion in something invariable and static. It has being as far as it is a transition to invariable being (DC 279b1–2; 7, 274b13–15). Ultimately, Aristotle's Unmoved Mover is, as indicated earlier, the prototypical account of this point of view.

In short, becoming has by no means the status of the radically novel or newly created; nothing emerges which was not already there beforehand. Becoming is a movement toward completion of a form that was given to it beforehand (MT 999b4–17; Heimsoeth 1958, 182ff). This is also why the idea that species are the result of a particular evolution is for Aristotle ungraspable. Essence philosophy represents a worldview in which being and becoming are oppositions, and becoming is explained and valued from fixed, invariable principles, forms or formulas. The Aristotelian teleological approach of the self builds on the idea that changing qualities presuppose an invariable substrate by virtue of which the self can change and still remain the same. The self is capable of changing and developing itself into a profoundly rational and virtuous being. However, the pedagogical development of the self consists of liberating itself from deceiving forces so that it can see what was already inscribed in its nature and gradually acquire the right attitude; change does not entail a radical transformation into a completely different self. This framework seems pervasive but raises two fundamental issues.

Two Fundamental Problems: Real Novelty and Causation

The essentialist approach has two fundamental difficulties: first, the assumption that all phenomena can be explained from one or more pre-given essences, which are not themselves a product of an evolutionary or empirical-historical process, precludes the possibility of real novelty. From an essentialist perspective becoming can never be radical: something cannot evolve into something of a completely different nature. Deducing becoming from being, that is, presupposing pre-given instances that explain and determine all becoming in the world, precludes the possibility of real novelty.

From the second half of the 18th century on, this presupposition has become ever more problematic. Observations of factual increase in diversity and variety in the world as well as experiences of the difficulty of

finding anything that will stand the test of time, indicate that this view is unsatisfactory. Darwin's *On the Origin of Species* (1859) has ferociously shaken the world awake: the idea of the possibility of the transmutation of species undermined the belief that species are unchanging parts of a designed hierarchy. As a consequence, the last bedrock that could secure the view that, despite the abundant (accidental) changes that we can observe in the world, there were invariable essences that ascertain a fixed order became untenable. Theories that could not account for the possibility of real novelty were rapidly consigned to the wastepaper basket of nonscientific beliefs.

The conception that even species are not pre-given but have rather emerged radically undermined essentialist views (Hausman 1993, 168–193), which also have implications for views of the self. Although our identity might be fixed to a certain extent, it is at the same time profoundly undetermined. By virtue of this indeterminacy or potentiality, we are able to realize possibilities "in ourselves" that we in no way are able to anticipate and that, on the basis of our contemporary criteria, might be even considered "unnatural" or "inhuman."

A second, not less fundamental difficulty that is inherent in the essentialist view is that it fails to account for the dynamic connectedness of things. If everything that happens is understood in terms of individual substances, all of which have their own properties, how then is it possible to understand that these substances act on other things, let alone that they are the cause of something else? If there is nothing that connects the effect to the cause, there can be no causal relation. Aristotle sometimes seems to want to explain causality on the basis of his concept of efficient cause and identifies different efficient causes (the father as the cause of the child, hard work as the cause of appropriateness, invasion as the cause of war) (MT 1032a12–b22): efficient causality refers to the transfer of the form of the efficient cause to the effect. This is done by a movement that incites the efficient cause (GA 715a4–7; 21–22, 729b–730b31; AP 94a20–22; MT 983a30 and 1032a12–b22). However, how the form is transferred from the efficient cause to the effect is one of the greatest puzzles in Aristotle's theory of causality. Because of his essence ontology, relations in the world, of whatever nature, ultimately only exist by virtue of the interaction between substances. How that is possible, however, is difficult, if not impossible, to understand from the essentialist perspective, since substances are essentially characterized by independence, unity and durability. Aristotle presupposes that objects and people engage in relations with one other, but these relationships exist "between" them, in an enigmatic manner belonging to none of them as such. Relationality and interaction have in Aristotle's conception no actual place in the world and consequently no actual significance in human identity (Debrock 2003a, b).[2] This is difficult to reconcile with our experience that we are strongly influenced and shaped by other things and people.

Overcoming Essentialism: Toward an Interactionist View of the Self

From the 19th century on, the view that dynamic relations and interactions are not additional but rather constitutive has gained ever more ground. This shift has not only been caused by developments in the sciences, but also various philosophers have proposed a nonessentialist account of the self, which gives import to the possibility of radical novelty and the constitutive character of interactions. The view that there are no individual selves prior to their interactions can already be found in Hegel, who has influenced many philosophers who advocate a dynamic, evolutionist worldview. Also, philosophers before Hegel have, more or less, explicitly developed conceptions that could be considered "relationalist," "interactionist" or "processual." We could even go back as far as Heraclitus, who proposes the idea that everything continuously flows and that things are what they are by conflict of opposites.

The alternative ontology that I will propose and use to criticize essentialism can be called "event ontology." From this perspective, not essences but events are the "basic constituents" of the world: *unless something happens, there is nothing at all*. This not only means that events are ontologically prior to what is, but also that being is derived from events rather than the other way around. Whatever is, is an abstraction derived from what happens. This also implies that nothing happens in isolation and unless it involves interaction. On a more particular level, I understand events in terms of interactions. From this perspective, it is crucial to recognize that interaction does not occur between already existing things, but that interacting things come to be what they are by virtue of interactions: "chess players do not play chess, but *playing chess* makes 'chess players'"; "lovers do not make love but *love-making* makes 'lovers'"; "murderers do not murder but *murdering* makes 'murderers'"; and so on. These examples indicate how the identities that we ascribe to both things and humans have only reality and meaning in the light of certain events that are generated by particular interactions.

An interactionist approach has been elaborated, as mentioned earlier, in many different directions. In the last decades, poststructuralist like Michel Foucault, Jacques Derrida and Judith Butler have conceived the self as a result or effect of power structures and/or (textual or institutionalized) discourses. Paul Ricoeur and Charles Taylor, both in their own way, have proposed the notion of "narrative" to understand the self and have illustrated how selfhood implies otherness (other humans, tradition, values, etc.) to such an extent that selfhood and otherness cannot be separated; the self implies a relation between the same and the other (Ricoeur 1990; Taylor 1992; see also Giddens 1991 and Rasmussen 1995).

Remarkable in the broad and complex discussion on personal identity is the slight recognition of the constitutive character of relations

and interactions. Contemporary philosophers like B. Williams, S. Shoe-maker and D. Parfit formulated in different ways the necessary and sufficient conditions for persevering the same identity. However, in spite of all the great differences between the contestants in this debate, the starting point is often that ultimately identity is constituted by something (psychological or physical) that, in one way or another, is prior to the interactions with the social and material environment. The essentialist inclination to disregard interaction and radical novelty, without which an adequate account of the self is impossible, seems to be also present (if concealed) in current discussions about personal identity. However, a critical analysis of these approaches will not be provided in this book because that would make the run-up to developing an interactionist view on human-technology relations too long.

This essentialist inclination certainly does not only apply to accounts of personal identity. It seems to be very difficult to overcome the essentialist approach; explicit and bold attempts to overcome it are often diagnosed as "still essentialist" by later, self-proclaimed "more advanced" critiques of essentialism. Essentialism seems to be engraved in our thinking and grammar; thinking in terms of subjects and predicates already mirrors, as we have seen in Aristotle's philosophy, the substance-properties structure. The idea that relations precede relata seems to be very difficult to conceptualize, as indicated earlier; we seem to necessarily presuppose that there must be first "things" that then relate to one another. However, in our practical encounters we experience change and exchange everywhere. We find ourselves in different ever-changing webs of relations and networks, both physically and virtually. We seem to be torn between two ideas: in theory, it seems very difficult, almost impossible, to overcome an essentialist approach to the self, while in practice, we are always engaged in relations and interactions. Western thinking seems to be schizophrenic with regard to the question of the self.

It is not the first time that we come across the discrepancy between theory and practice. This discrepancy, one could say, has been responsible for the birth of phenomenology in the 19th and 20th centuries. In order to develop an account of the self that does justice to the elements of novelty and relationality, it is, I believe, for that reason important to concur with the phenomenological approach. The view that I will develop is aimed to prove itself fruitful for my developing a sound view of self-formation and the question of how technologies affect self-formation. An important distinctive feature of the view that I will propose in this respect is that it accentuates more than Technological Mediation Theory (TMT), which has been also strongly influenced by a phenomenological perspective, a dynamic and pragmatist view of the self: technologies do not only influence and mediate our self-*conception* and self-*experience*, but they also actively instigate in a dynamic interaction the *formation* of a particular self. It is probably not a coincidence that TMT often

goes back to the more hermeneutical phenomenology of Husserl and Heidegger, whereas my interactionist perspective is, as I will elaborate, more inspired by the anti-essentialism of Nietzsche and the pragmatist phenomenology of Peirce (see Aydin 2007b for a comparison between Husserl and Peirce). Although both Nietzsche and Peirce are often not considered part of the phenomenological canon, there are, as we will see, strong affiliations between their thought and phenomenology. In a broader sense of the word, I do consider them both phenomenologists (Peirce explicitly named himself a phenomenologist), insofar as they attempt to take as much as possible "pre-theoretical" everyday life experience as a leading method in their analyses.

A question that might arise here is why Foucault's view, and especially his notion of "technologies of the self" that he introduced shortly before his death in 1984, is not used to guide the development of the intended framework. Foucault discusses different procedures and practices, available in all civilizations in different forms, through which individuals constitute, develop and maintain an identity (Foucault, 1984, 369; 1989, 134). This mechanism in the constitution of the self – also characterized as "subjectivation" – is what he referred to as "technologies of the self." For Foucault (1988, 18), technology does not refer to material matters but rather to "a matrix of practical reason," a kind of applied knowledge that makes self-transformation and subjectivation possible. He devoted significant attention to so-called ascetic practices in Greek and Greco-Roman philosophy and early Christian spirituality.

I have two reasons not to go for Foucault as a lead guide. The first reason is that his philosophy does not explicitly and directly (I do not exclude that it does implicitly) offer an externalist ontology of relations and interactions. The second reason, which I must elaborate more extensively, is that his focus on the individual and individualism does not fit well with my intended externalist view of the self. This particular focus also seems to manifest itself in Foucault's tendency to see an antagonism between the individual and society, which explains his particular interest in domination and repression and on how societal institutions exert power over individuals. In this view, I recognize a Nietzschean legacy that I do not subscribe either, as I will explain later. Nietzsche's influence on Foucault's view of power and power technologies does, however, make Foucault resonate indirectly in this study. I know that there is discussion about the extent to which the social and the community plays a role in Foucault's processes of self-cultivation and care for the self (see, e.g., Myers 2008 and Wong 2013). Nor do I argue that Foucault does not pay attention to the individual's social surroundings. The point is that I see a prioritization of the individual over the social, which does not correspond well with the framework I want to develop. As Foucault points out himself: the "care for others should not be put before care of oneself. The care of the self is ethically prior in that the relationship with oneself is ontologically prior"

(Foucault 1997, 287). However, this is not to say that there are no elements in Foucault's philosophy that can complement, enrich and nuance the framework that I will develop. I think there are.

Another candidate that needs to be mentioned explicitly is Alfred. N. Whitehead; why is not he chosen as a guide for the development of the proposed ontological perspective? It is clear that Whitehead has greatly contributed to the establishment of process philosophy. He, nevertheless, does not provide the most suitable framework for developing an interactionist account of the self, which becomes clear when his views on continuity, discreteness and time are considered. For Whitehead, the ultimately real things of the universe, from which all other realities are derived, are actual entities or actual occasions. Although the moment of becoming of these actual entities is also their moment of perishing, as a result of which they become a fully fixed, unchanging past, they strangely enough do not change while maintaining an identity. They are, as Whitehead puts it, "incurably atomic" (Whitehead 1929/1978, 195). The coming to be of time requires for Whitehead a succession of ontologically discrete units. These discrete units are not temporal, because their atomic succession gives rise to time. Put differently, because they are the building blocks of time, they cannot be "in time" (Whitehead 1925, esp. 10, 126; Whitehead 1929/1978, 112, 363). Whitehead's focus on atomicity and succession of discrete units seems to be in discrepancy with the view that interaction and continuity are primary.

From a Peircean perspective, "time itself is", as Sandra B. Rosenthal puts it, "a continuous spreading out in which quasi-discretes, which are themselves continuous processes, emerge in the passing present through the interaction of dynamic tendencies constitutive of the ongoing temporal/ontological advance" (Rosenthal 1996, 561; see also 542f, 547f). Charles Hartshorne believes that the emphasis on radical continuity is precisely where Peirce and other "process pragmatists" fail. According to him, Whitehead gives us, with respect to this point, a more adequate conception (Hartshorne 1964, esp. 460f., 467, 473; see also Hartshorne 1980, esp. 283). I believe that Whitehead and Hartshorne's view that continuity presupposes atomic units that are subject to change in the end embraces an Aristotelian ontology that cannot account for the constitutive character of interactions and the possibility of real novelty.

I will try to show how the anti-essentialism of Nietzsche and the pragmatist phenomenology of Peirce offer us a framework that allows doing justice to the elements of interaction and novelty without disregarding the importance of durability and regularity. Insights of these philosophers will be used to provide the basic building blocks of an interactionist view of the self, which is needed to overcome an essentialist and dualist approach of the self. However, throughout this book, this basic framework will be complemented by ideas taken from various other philosophers and approaches.

Notes

1 It must be clear that Aristotle does not conceive the form of a thing in modernist terms as a category of reason à la Kant, but rather as a dimension of the thing itself; the form is not a property of reason, though it can be captured by reason (see *De anima* II 4, 415a14–22 and 415a14–22; III 4, 429b21–22, 430a3–5 and 423b26f; *Categoriae* 7, 6a36f and 6b33–36).
2 For a more extensive discussion of the problem of "the missing interface" in the substantialist view of the world, see Aydin 2003, Section one and two; see also Hulswit 2002, Chapter 6.

2 From "Self" to "Self-Formation"

Unless something happens, there is nothing at all; everything that "is," emerges from and is constituted through interactions. To illustrate this, I will lay out the ideas of Nietzsche and Peirce in this chapter. First, Nietzsche's notion of "will to power," which represents an ontology of dynamic interrelatedness, will be interpreted as a *radical critique* of the essentialist approach. This Nietzschean ontology provides fundamental building blocks for developing a radical anti-essentialist view of the human self, which will also be employed later in the book to analyze and evaluate what is considered an "enhanced human." Then, Peirce's "list of categories" will be discussed, which enables developing a more systematic and more detailed interactionist account of self and self-formation. Nietzsche and Peirce's fundamental views will complement and strengthen one another,[1] laying the groundwork for developing a perspective that recognizes that the self is formed through dynamic interactions with its changing social and material environment without rendering it a mere product of contingent environmental influences.

2.1 Nietzsche's Will to Power as an Anti-Essentialist Ontology

For Nietzsche deducing becoming from being entails negating the primordial reality of becoming (TI, 176). From an essentialist perspective being and becoming are opposites, and becoming is explained from fixed, invariable principles. Against this view, Nietzsche postulates that reality is a variable multiplicity of wills to power: the world does not consist of a collection of durable and independent unities, but of a multiplicity of variable and relational wills to power.

Nietzsche's intention is not, as will be argued, to explain reality from a new principle that determines its essence (the "will to power"), but rather to show from a "new principle" what happens if something is declared as the essence of the world or the self, when something is determined as real or unreal, or as true false. Nietzsche repudiates the Aristotelian idea that one cannot ask what substance is, because substance is the underlying instance of everything that *is*, that is, because it *is* in the strongest

meaning of the word. If reality does not consist of pre-established dura-
ble and independent unities, then it has to be clarified what "precedes"
them, that is, what generates essences and how that generating should
be understood. The notion of the will to power is Nietzsche's attempt
to provide this clarification, which should be also seen as an attempt to
do justice to the dynamic character of reality without denying that real-
ity is also to a certain extent organized. In the last part of this section,
Nietzsche's framework, which I will call the Organization-Struggle The-
ory (OST), will be applied to the problem of decadence (also known as
the problem of nihilism). This application will display the first contours
of the proposed interactionist notion of self-formation.

The Ontological Status of the Will to Power

From the beginning of the second half of the 1880s, Nietzsche proclaims
explicitly that all reality is will to power: "[t]he world viewed from in-
side, [...] it would be simply 'will to power' and nothing else" (BG 36; see
also SZ II, On Self-Overcoming; BG 13). Reality has only one intrinsic
quality: the will to power. At the same time, the will to power is the only
principle of interpretation (*Deutungsprinzip*) for reality.

This claim raises many questions: is the will to power a teleological
principle or a substantial cause? Does it involve a homogenization of
reality and, therefore, a negation of diversity and richness? In addition,
Nietzsche calls his notion of the will to power repeatedly a hypothesis.
How should we understand all this?

Reflection on Nietzsche's notion of "power" already provides some
clarity. "Power" in "will to power" is a peculiar concept. It is character-
ized, and this is a crucial point, by intrinsic relationality: power is only
power in relation to another power. Nietzsche says: "A power quan-
tum is characterized by its effect and its resistant" (KSA 13:14[79]; see
also KSA 12:9[151] and KSA 12:2[159]). The concept "power" would be
meaningless if a power would be detached from an opposite power. That
power is inherently relational implies further that it is characterized by
a relation without relata that precede it or that can exist independent of
it. Nietzsche's principle of the will to power implies that relation is not
an additional element of things, but something that constitutes in a fun-
damental way what a thing is. In other words, there are no first things,
which then have relations with each other; rather, things are what they
are by virtue of their relations; what constitute things are for Nietzsche
ultimately power relations.

Furthermore, Nietzsche's concept of power implies that reality is
dynamic in the strongest meaning of the word. Power, in Nietzsche's
view, entails a directedness or causation without there being something
(durable), a fixed cause, which can be separated from that directedness
or causation; power is in its essence "something" that does not coincide

with itself. It is an always-being-on-the-way. Additionally, this structure implies that power must be understood as a necessary striving for more power (KSA 13:14[82]). Power is a necessary striving to expand itself. Power is only power insofar as it can maintain itself against other powers, and strives to predominate over them.

There is, in Nietzsche's worldview, nothing that has existence and meaning outside the game of power relations. Because of this, withdrawal from this "game" is impossible. Even rejecting the claim that reality is will to power is an expression of will to power. Making a statement about the cause or (pre-given) goal of a thing is nothing else than the formulation of a will to power, which always can be challenged by other wills to power. This power game also constitutes what we consider our "inner world" (KSA 11:25[185]). Willing is an outcome or an expression of this game of forces: somebody who wants something commands something outside or in himself that obeys, or that he thinks it obeys. Every view, action or decision is understood as a power seizure, or as the effect of it. Although the necessary striving for more power can be called teleological, it is not teleological in the traditional Aristotelian sense. What we have here is, in a certain sense, a teleology without *telos*. The crucial point is that the "teleological" character of the will to power not only has no pre-given, fixed end, but it precisely *precludes* such an end.

Such a pre-given end that is precluded, and that Nietzsche frequently attacks, is self-preservation. Nietzsche characterizes the notion of self-preservation as one of those "*superfluous* teleological principles" (BG 13). At the same time, this conception is exposed as an attempt to negate the reality of becoming. The statement that all life strives for self-preservation presupposes that there is a substantial self that wants to preserve itself. Nietzsche repudiates that there is such a self (KSA 12:9[98]; see also GS 349).

The notion of the will to power can be conceived as a kind of hypothesis. It is, however, not the kind of hypothesis that can be proved to be a true and valid thesis through sufficient verification or lack of falsification. Such is not possible because all conditions that have to be fulfilled and all the standards that determine when a hypothesis is verified or falsified are themselves expressions of will to power (van Tongeren 1989, 174–177). In the game of power relations, every power tries to impose its own conditions and standards on the rest. A proposition never loses its conditional character, because it is continuously being questioned by other powers. There is no proposition that can ever be determined definitively as being true. The hypothetical character of the notion of the will to power expresses this provisional status, this "always-being-at-risk," of every proposition.

For Nietzsche, every actualization, dependent on a particular constellation of power relations, is the realization of only one possibility. There is what he sometimes calls a permanent chaos at work, which is

a condition for discovering evermore and alternative worlds. The chaos is, therefore, not a mere burden that we have to overcome to survive or make our life easier; that is only one aspect of it. It also plays a very positive role. It is the basis for all creation and creativity. Without it, nothing *novel* could emerge. The more that chaos breaks into our ordered world, the more our creative power is stimulated (KSA 8:5[188] and KSA 10:5[1]). In *Thus Spoke Zarathustra*, Nietzsche puts it in the following poetic formulation: "I say to you: one must still have chaos in oneself in order to give birth to a dancing star. I say to you: you still have chaos in you" (SZ I, 9; see also KSA 9:11[121] and KSA 10:24[5]). Nietzsche believes that this element is operative in every aspect of reality. Not only is there no final ground or divine order to which ultimately everything can be reduced, but there can be also no phase in which there is no chaos anymore. For Nietzsche, "underlying" every order there is an element of chaos that is operative: we never can reach a final ground or ultimate end, but are always confronted with a multiplicity of possibilities (Aydin 2004, V1).

The principle of the will to power proves to be a special kind of "principle," one that deprives every principle that serves as the basis of our interpretation of reality of its unconditional character. The homogenizing of reality in this way does not lead to the negation of the vitality, diversity and richness of the world. On the contrary, due to it, every determination of reality, every interpretation, can be continuously questioned by opposing powers; because of this, other interpretations always remain *possible*.

For Nietzsche, all reality is the result of a continuously changing hierarchical order of smaller and bigger coups (KSA 13:14[81]; KSA 13:14[98]). Change can be the symptom of both the establishment of a new hierarchical order ("substantializing") as well as the collapse of an old order ("de-substantializing"). Nietzsche's ontology aims to clarify how the processes of individuation proceed, that is, how a variable and relational multiplicity arranges itself, decays and rearranges itself in different directions and in multifarious ways, how different functions and phenomena form and decay, which is also displayed in Nietzsche's concept of "organization."

Organization and Form

Nietzsche's interest in the concept of organization is not surprising: if multiplicity, variability and relationality are essential constitutive aspects of reality, and if, as a result of that, there are no pre-given forms, then a seemingly independent and durable unity, that is, every perceivable form of reality, can only exist as a variable and relational multiplicity that is held together in some way. In Nietzsche's words: "[a]ll unity is *only as organization and interplay* a unity" (KSA 12:2[87]). A variable

and relational multiplicity that is kept together is an organization – that which *keeps* it together is, according to Nietzsche, will to power.

Any instance of will to power as such cannot be a durable and independent unity – it is always a variable and relational multiplicity held together, and those wills to power exist only as a multiplicity of wills to power and so on *ad infinitum*. There are no last unities that one ultimately bumps up against. Speaking about "a will to power" is, therefore, always misleading. "All reality is will to power" can be determined more accurately as: "all reality is 'will to power'-organizations."[2] And because "interaction" is dynamic in the strongest sense of the word, the term "organization" should be understood not as a noun but as a verb; organization is an activity. The variable and relational multiplicity has to be organized continuously. An important implication of the ontological status of the will to power is that reality is always necessarily organized to some degree. An absolute *dis*organized reality is a *contradictio in terminis*. Organization is inherent to life (KSA 10:7[174]).

The will to power is the ground of existence of an organization, that is, which constitutes and drives an organization. But how does that work? Bound up with the concept of "will to power" itself are the facts that: (1) "will to power"-organizations act on each other and (2) this acting is directed at gathering more power. "Will to power"-organizations assess each other, judge each other with the intention to subdue each other. A "will to power"-organization is subdued when it is converted into a function or functionary of another "will to power"-organization (KSA 12:2[76]). This converting into a function or functionary of the suppressed means that certain rules of behavior are imposed on it – that it is disciplined. The result of a successful conquest is that the conqueror becomes stronger – that she grows.

The suppressing and converting into a function of a "will to power"-organization does not happen without a struggle. The "will to power"-organization that is being subdued resists, because every "will to power"-organization is inherently directed at subduing. Subduing a "will to power"-organization requires making equal what is unequal. The "will to power"-organization that has to be subdued must be arranged and transformed in such a way that it fits in with the organization of the ruler (KSA 11:40[7] and 13:14[186]). Elements that can in no way be converted into the new unity are removed or at least made harmless. However, the ruler must transform its organization in such a way that it is able to accommodate the suppressed. How many concessions have to be made on each side depends on the strength of the different "will to power"-organizations. If the ruling "will to power"-organization is not able to organize what it has subdued, decline sets in (KSA 12:2[76] and 12:9[151]).

Nietzsche's view on organization also sheds light on how regularity, stability and "truth" can exist. Stability is, says Nietzsche, the result

of ordering "will to power"-organizations in different ways according to a certain "form and rhythm" (KSA 10:24[14]). If a "will to power"-organization is able to maintain a certain hierarchical order for a long time, the "illusions" of durability, unity and independence emerge; one then believes that a certain form is substantial. If this belief becomes so strong that it is no longer questioned, a ground-form is established, forming the basis for what we hold to be reality and truth. Ground-forms are prejudgments (*Vor-urtheile*) of the mind. Knowing is nothing other than organizing a multiplicity of new impulses according to a certain incorporated ground-form. Ground-forms are a certain type of filter that simplifies and equalize the multiplicity of impressions in such a way that we think each time that we perceive the same thing (KSA 11:26[94] and 11:41[11]).[3] Nietzsche emphasizes repeatedly that the oldest incorporated forms are the strongest. The reproduction of ground-forms is an activity that engenders structure in the multiplicity that organizes it according to the longest incorporated ground-forms (KSA 11:26[94]).

Insofar as ground-forms fulfill a pragmatic-naturalistic function and stimulate the constitution and growth of new and stronger "will to power"-organizations, Nietzsche does not value them negatively. On the contrary, they are of vital importance. An organization maintains itself only insofar as it fixates itself continuously in a certain way, according to a certain form and rhythm. Life is possible only by virtue of substantializing, which always means falsifying, that is, making equal what is not equal. Truths or incorporated forms are life-ensuring "fictions" (KSA 11:43[1]). They secure stability, which is needed for self-reservation and growth. The problem, however, is that ground-forms remain operative, even after they cease to fulfill that function. In a physiological context, Nietzsche states: "every drive is bred as a temporary *condition for existence*. It is inherited, even long after it has ceased to be that" (KSA 11:26[72]). The danger Nietzsche is warning against is the preservation of ground-forms that no longer stimulate growth and novelty. Ground-forms can also obstruct the constitution and growth of new and stronger "will to power"-organizations. How exactly they do this and how those obstructions can be overcome will be discussed in the next section, which deals with the notion of struggle and its relation to organization and form.

Organization, Form and Struggle

The will to power is, as we have seen, essentially directed at subduing as many other wills to power as possible; the opposing wills to power, however, are also directed at that (KSA 13:14[186], 11:36[22], 11:40[55], and 11:26[276]). Hence, the interaction between wills to power is characterized by "struggle." That the will to power only exists by virtue of its actions and its resistances, as we saw earlier, means that it only *is* by

virtue of struggle. And since everything that happens is will to power, Nietzsche claims: "All happening [*Geschehen*] is *struggle…*" (KSA 12:1[92]; see also 2:9[91]). This is *not* to say that all reality is *based on* struggle or all reality is *determined by* struggle. Such interpretations already assume that struggle is an additional quality of something that distinguishes itself from it. Struggle, however, is a constitutive relation, not an additional or secondary element.

I repeatedly stressed that Nietzsche conceives organization as inherent in all reality. Reality is such by virtue of both struggle and organization. How should we understand the relation between these two concepts? Clarifying this relation will also shed light on the salient difference that Nietzsche makes between strong or healthy and weak or sick "will to power"-organizations.

A "will to power"-organization overpowers another "will to power"-organization by the force that is released through the "discharge" of its internal tension (KSA 12:9[92]).[4] In addition to the force that is needed to manage its own organization, there has to be enough force to incorporate the "will to power"-organization that has to be subdued (KSA 10:7[95]). Internal tension is generated by building up the internal struggle in an organization. At the same time, this tension can only be built up if the opposing parties are related to each other in a certain way; if, in other words, the struggle is organized. Moreover, overpowering is only possible if the internal struggle is organized in such a way that the tension can be discharged *at the same time and in the same direction*, which requires a hierarchical ordering. Thus, on the one hand, the internal struggle in an organization has to be increased and intensified to such a degree that enough internal force is generated. On the other hand, this internal struggle has to be organized in such a way that the organization does not fall apart and that the discharge has a certain direction (KSA 12:9[92]).

We can now also take up again the concept of "ground-forms." Ground-forms are manifestations of "will to power"-organizations that are hierarchically ordered in such a way that subduing other organizations is possible. These are incorporated because they have proven themselves to be "means" that can preserve a certain kind of life, on the basis of which a "will to power"-organization can become stronger. Ground-forms can, however, as indicated earlier, also obstruct the constitution and growth of new and stronger "will to power"-organizations. That happens when they are given an unconditional status. One then "forgets" that they are for the use of the becoming-stronger of "will to power"-organizations. The effect of declaring a certain ground-form unconditional is that the struggle between and within "will to power"-organizations is weakened. Other ground-forms, which are manifestations of other "will to power"-organizations, do not get a chance to develop themselves. Without struggle, a "will to power"-organization cannot become stronger.

That is why Nietzsche campaigns against every type of unconditionality: "the world of the unconditional, if it existed, would be the *unproductive*. But one must finally understand that existent and unconditional are contradictory predicates" (KSA 11:26[203]).

In order to vitalize internal and external struggle, Nietzsche summons taking up the fight against dominant ground-forms. Only continuation and intensification of the struggle between a multiplicity of "will to power"-organizations can guard against disintegration. Therefore, he states: "[t]here has to be struggle for the sake of struggle" (KSA 11:26[276]); "[t]he struggle [...] becomes a *regulating principle*" (KSA 10:7[190]).

However, it is important to recognize here why Nietzsche often emphasizes the importance of struggle: he believes that in his (our?) culture, Platonic metaphysics and Christian morality have organized life in a uniform way to such a degree that every possible counteraction is destroyed. That does not mean that the element of organization is less important. In some of his thoughts on the possibility of a future Europe, for example, which is an issue that is relevant today more than ever, he emphasizes the element of organization, because in that case he finds tremendous struggle and division, but no organizing force (BG 208 and 256). Moreover, life forms that have declared themselves absolute can only be effectively challenged if the opponent is well organized. Although struggle is indispensable for generating continuous growth, it is able to do that only if the different struggling parties remain, at the same time, related to each other. Only then are "will to power"-organizations able to subdue other "will to power"-organizations, to become stronger and to guard themselves against disintegration.

OST shows that Nietzsche's view of "strong" and "weak" is nuanced. Only the combination of strong organization and intense struggle is a trait of strength and health. If a high degree of organization is achieved by excluding all struggles, then that is a sign of weakness. Additionally, intense struggle without great organizational force is a sign of weakness. A strong or healthy "will to power"-organization is characterized by considerable divergence and struggle that is forced into a unity in a structured manner. A kind of discrepancy or instability or chaos (Aydin 2004), as Nietzsche often calls it, is therefore inherent in the strong type. That is why the strongest organisms can be the most vulnerable: the more internal struggle an organization contains, the more difficult it is to maintain a unity, the bigger the chance that it will fall apart. Weak or sick "will to power"-organizations, in contrast, which contain little divergence and struggle, can be more easily organized and have less chance of falling apart. They are more stable than a strong organization, which also explains why Nietzsche can say: "the strongest and most happy are weak when they have the organized herd instincts, the fear of the weak, the majority, against them" (KSA 13:14[123]; see also GM 3, 18 and KSA 11:27[40]). That, however, does not alter the fact that the

weak type is unfruitful and, in the long run, liable to a process of dis-integration. This refinement of what Nietzsche conceives as a strong or healthy type obviously makes quite untenable the view that the notion of the will to power simply implies brutal, bodily force.

Decadence and Nihilism

Nietzsche has examined weak or sick types most prominently in the context of his reflections on "decadence" and "nihilism." In his influential work on Nietzsche, Müller-Lauter defines the decadent, very much in line with the above account, as a type of will to power that is unable to organize the conflicting variance within itself. He further explains it as an "[i]ncapacity for organic formation," as a physiological deficiency of organizing force, which manifests itself as a becoming independent of the part with respect to the whole (Müller Lauter 1971, 1999a, 4; Aydin 2002). Degenerating will to power has, according to Müller-Lauter, still enough unity to strive for disintegration in striving for nothingness. Avenging its impotence to organize itself, the weak negates itself – and with that, reality as will to power – with the help of a fabricated, "true" reality. A difficult question that arises with regard to nihilism in relation to Nietzsche's concept of the will to power is how nihilism is possible at all. How can the will to power negate itself – and not acknowledge itself? How can life, of which the will to power is the principle, turn against itself? How can a degenerating will to power will to fall apart? How can a part detach itself from the whole? If the will to power is the ultimate quality and the only principle of interpretation, then nihilism too must be explicable in its terms.

OST could help understand how that is possible. On this view, the possibility of nihilism can be explained from two perspectives, which are (necessarily) connected. The first explanation is that nihilism is bound up with the structure of organization. A "will to power"-organization can become stronger only if it is able to organize the multiplicity of wills to power in itself according to a certain form. It must impose a certain identity on itself. The imposition of a form on itself is necessarily accompanied by a weakening or restraining of the internal struggle. That internal struggle is however, at the same time, necessary for growth. Organization is necessary for gathering more power, but it contains at the same time the danger of decadence.

This danger can be elaborated by clarifying how decadence is bound up with the structure of a ground-form insofar as it functions as life-ensuring and life-enhancing. An important function of a ground-form is to regulate the chaos of impressions, that is, the impact of other "will to power"-organizations; a ground-form enables a "will to power"-organization to defend itself against those influences so that it does not fall apart. Regulation occurs by "the arranging of the new material on

the old patterns (= Procrustes bed), a *making* equal of the new" (KSA 11:41[11]). This type of inertia can be life-ensuring or even life-enhancing to a certain degree. However, the power constellations can at a certain moment have changed so much that maintaining a certain ground-form becomes counterproductive. The inertia inherent in ground-forms can inhibit the "perception" of this (in time to do something about it). Thus that a "will to power"-organization is not able to reorganize itself on time, that is, to impose on itself a different form, is caused by the inertia, the life-ensuring equalizer and negationist of what is new, which is inherent in ground-forms.

Decadence can also be the result of the success of a certain ground-form. A certain ground-form can be so successful in its submission of opposing "will to power"-organizations that it destroys all internal and external struggle. Since a "will to power"-organization only exists and grows by virtue of struggle, excessive success has *dis*integrating consequences in the long run. The structure of form contains both growth and decay.

The second explanation is that decadence is internal to the dynamic character of struggle. Internal struggle is a condition for a "will to power"-organization to become strong. When, however, the internal struggle cannot be organized in such a way that a tension is generated that can be discharged in a certain direction, decline occurs. Struggle is necessary for gathering more power, but it contains at the same time the danger of decadence. The greater the struggle, the more difficult it is to organize it, the greater the chance of decay.

This explanation of decadence will prove itself useful in shedding light on the problem of self-formation. In the last part of this book, this explanation will be used to clarify how a distinction could be made between a "good" or "healthy" type and a "bad" or "unhealthy" type of self-formation. In that respect, commanding will be understood as commanding oneself, as self-overcoming (Heidegger 1961v2, 37, 651, 265), which requires recognizing the internal divergence in the structure of the will to power; "self-commanding" and "self-overcoming" mean being able to organize the multiplicity of wills to power, such that the struggle is not weakened or destroyed, but intensified. A "willing person" does not command herself, but rather commands something in herself (Müller-Lauter 2000, 87). The "self" is only a "unity" as organization. A decadent or nihilistic type of self-formation that lacks the capacity of self-commanding and self-overcoming will be explained as the antonym of what I will later in this book will call "sublimation."

Toward a More Comprehensive Notion of Self-Formation

OST shows that, in contrast to Aristotle's teleological account, the direction of the development of reality is not *a priori* established, thereby

guaranteeing the possibility of fundamental novelty and radical change; the world is, in a sense, continuously pregnant with a measureless variety and multiplicity of possibilities that are still unknown to us. This theory also does justice to the dynamic connectedness of the world. The notion of struggle indicates that a thing exists only by virtue of its actions and its resistances, which implies that interactions do not occur between pre-existing things, but that interacting "things" come to be what they are by virtue of interactions. Moreover, that Nietzsche's worldview is dynamic in the strongest sense of the word does not mean that there is no regularity operative in the world: if a "will to power"-organization is able to maintain a certain hierarchical order for a long time, the "illusions" of durability, unity and independence emerge. A certain ground-form then is established, which forms the basis for what we hold to be "real" and "true."

The Nietzschean OST gives some grip on how we should understand a dynamic and relational worldview. However, a more encompassing account is needed of how regularity and form can include multiplicity and variability, of how an entity that is regulated in a certain way still is able to transform itself in something significantly novel. This requires a better understanding of how something like radical change is possible, how exactly something *novel* can occur. Adopting a certain kind of potentiality seems unavoidable. Although Nietzsche gives us notions such as "tensile force," "chaos" and "chance," which sheds some light on the phenomenon of potentiality, a more systematic analysis of how a non-Aristotelian notion of potentiality should be grasped is needed. Only then we would also be able to better understand the relation between potentiality and actuality, that is, how something that is not present can nevertheless actualize itself and emerge in "the world." Clarification of that relation will help us to understand how exactly a self imparts a form to itself without completely fixing itself and how it can continuously change without falling apart. In short, the relations between the different dimensions of an entity (potentiality, interaction and regularity), eventually applied to the self, have to be explained and elaborated more systematically.

The second point that needs to be recognized concerns the possibility of a normative account from an interactionist worldview. Nietzsche seems to derive a kind of norm or ethical standard of evaluation from the notion of a strong or healthy "will to power"-organization, itself conceived in terms of the only ontological quality that he recognizes: striving for more power. The criterion for what makes for a strong or healthy organization cannot be external, because besides the will to power, there is no other quality or principle. A "will to power"-organization is strong or healthy insofar as it does justice to its nature or essence, which is the directedness at more power. From the degree of internal and external struggle and the way it is organized, one can make out to what extent

that has been met. So *if* the characterizations "strong" and "weak" can be understood as value judgments (strong is good; weak is bad), Nietzsche seems to have found ethics on ontology (see KSA 10:13[10]): something is good, insofar as it coincides with its "essence," which is the will to (over)power. In addition, the tenor throughout Nietzsche's texts is that the strong type is an ideal that should be pursued: you must become what you (really) are.

This view raises many questions and difficulties. First, the idea of the strong as an ideal that *can* be pursued sometimes seems to be incompatible with Nietzsche's notion of reality as will to power. An implication of Nietzsche's "homogenization" of reality as will to power seems to be that we cannot ask *why* a certain "will to power"-organization is strong or weak, what causes this. There is, besides the will to power, no other quality or (external) cause that could serve as an explanation. Nietzsche writes in a note from the late *Nachlass*: "that something is as it is, so strong, so weak, that is not the result of obeying, or of a rule, or of a compulsion…" (KSA 13:14[79]; see also KSA 12:2[139]). Something is, it appears, simply what it is. But if something is just (necessarily) what it is, then every form of self-control, and, hence, self-formation, is in vain (see also Moles 1990, 185–222). Nevertheless, throughout Nietzsche's work there are appeals to pursue certain goals and ideals, which assumes the possibility of self-control and self-formation.

Second, if the strong type could be considered as an ideal that should and could be pursued, then still realizing it seems to be a great challenge. How can we strive for, on the one hand, maximum struggle and dividedness, and, on the other hand and *at the same time*, for maximum relatedness and unity? (Müller-Lauter 1971, Chapters 7–8; Van Tongeren 1989, 213ff) In the next section, I will turn to Peirce's list of the categories in order to further develop the groundwork for the notion of self-formation. I will argue that this difficult and paradoxical appeal to organize the multitude without weakening the struggle, which we can find not only in Nietzsche's work but also, as we will see, in Peirce's, opens up a very promising answer to the question of how to form ourselves.

2.2 Peirce's Phenomenological and Pragmatist Categories as Basis for an Interactionist Account of Self-Formation

Peirce argues that in our encounters with the world, we always and necessarily must adopt or presuppose three phenomenological categories, which he simply calls the categories of Firstness, Secondness and Thirdness. Whatever we describe or experience, regardless of it being a thing, a fact, an event, a fantasy or a hallucination, it always necessarily contains or presupposes these three categories. Phenomenology is in his classification of the sciences the "highest," that is, most basic discipline

within philosophy, being in its turn – after mathematics – the "highest", that is, most general discipline of all sciences. Within philosophy, Peirce arranges, after "phenomenology," a cluster of three disciplines, which he calls the normative sciences, namely esthetics, ethics and logic. Logic finds its justification in ethics, which in its turn is dependent on esthetics. Phenomenology defines the framework in which these three disciplines and the other "lower" special sciences can develop their analyses, explanations and judgments in a proper manner. I will show that self-formation from a Peircean perspective can only be properly understood on the basis of these basic categories, but ultimately must be conceived as an esthetic process.

Firstness: Potentiality

Peirce defines Firstness as "the mode of being of that which is such as it is, positively and without reference to anything else" (EP II, 328). In whatever we experience, we must recognize a "state" that cannot be related to anything else. This "mode" or "state" is often regarded as a kind of quality (CP 1.418). The following example slightly illustrates the peculiar character of what Peirce considers as quality. He writes:

> [i]f a man is blind to the *red* and the *violet* elements of light and only sees the *green* element, then all things appear of one colour to him […]. Yet since all things look alike in this respect, it never attracts his attention in the least […]. For the very reason that it is his own kind of sensation, he will only be the more entirely oblivious of its *quale*. Yet for all that, that *is* the way things look to him" (CP 6.222)

Although this example may mislead us to think that Firstness refers to a subjective psychological state, we need to keep in mind that it, being part of phenomenology, precedes qualifications in terms of physical or psychological, subjective or objective and even real or non-real. In contrast, contemporary conceptions of qualia (Lewis 1929, 124f; see also Block 1978), Firstness must not be misunderstood as a subjective experience that could not be shared with others, but rather as "a sort of consciousness […] with no self," (CP 2.85; see also EP II, 146f. and 258f). It represents "nothing but sensation minus the attribution of it to any particular subject" (CP 1.332).

 Peirce's understanding of quality must also not be understood in an Aristotelian sense. Within Aristotle's framework, what is given in perception is an instantiation of a distinct universal concept (e.g., *Met*.1039a3–7 and *DA* 427a2). Peirce, however, claims that what is given always contains something that we cannot classify or describe; its absolutely unrelated and irreducible singular character makes identification by comparison or contrast impossible (CP 1.303; see also Gallie

1966, 187). What Peirce calls the *quale* of the color-blind man's visual experience is something he does *not* and *cannot* understand as a discernable quality or kind in an Aristotelian sense. Nevertheless, it is difficult to deny that he experiences a kind of "green," but this "green" is not the green that we could recognize consciously as green. The example at least makes plausible that a "state" must be assumed prior to conscious distinctions, which Peirce sometimes denotes as "*presentness*" (CP 5.44): a "passive consciousness of quality, without recognition or analysis" (CP 1.377; see also 5.44). What is immediately present is not the product of reflection, which means we cannot relate it to something that we have in our head and, consequently, cannot derive it from something that we already have in our head. Peirce denotes it hyperbolically as: "[w]hat the world was to Adam on the day he opened his eyes to it, before he had drawn any distinctions, or had become conscious of his own experience" (CP 1.357; see also Gallie 1966, 185).

Peirce emphasizes that quality, as he understands it, has a simple and singular character. At the same time, however, he claims that qualities show not only a variety, but even "unlimited and uncontrolled variety and multiplicity" (CP 1.302; see also 5.44). From his view on quality, both indications do not exclude but rather include one another: the absolutely unrelated and irreducible singularity of Firstness implies that each phenomenon contains an element characterized by undefinable originality, uniqueness and spontaneity. By virtue of this fundamental openness, there *can* be an unlimited variety and multiplicity of things, which refers to the *impossibility* of completely ordering, identifying, confining and appropriating what presents itself. This boundless variety and multiplicity cannot be understood as a plurality of distinct, particular and repeatable entities, but only as something that is indistinguishable, vague and indefinite. This element can be characterized as "possibility" or "potentiality."

From an Aristotelian perspective, attributing a quality to a subject always means attributing something secondary to something primary; expressed differently, first there must be something actual, an apple, before it can have a quality, for example, the quality "red." Potentiality presupposes for Aristotle, as we have seen earlier, actuality. In contrast, quality from Peirce's perspective is something that is not dependent on any subject: it is neither dependent on an actual entity that contains it nor is it dependent on being perceived or thought by a mind. Firstness refers to a phase of experience that "precedes" the recognition of distinctive sensory qualities (CP 1.422; 1.377) and "precedes" thinking and language (Rosenthal 2001). Potential qualities "precede" actual qualities, which secures the possibility of radical spontaneity and novelty. It must be clear that distinguishable qualities are not quantitatively distinguished from the unlimited variety and multiplicity of possible qualities, but a manifestation of that unlimited diversity of possible qualities

in a certain direction. Actual qualities are crystallized possibilities (CP 1.302; Alborn 1989, 6).

The following example might further clarify Peirce's difficult category of Firstness. Imagine a clean blackboard as a kind of diagram of the *original* vague potentiality. Suddenly I draw a white chalk line, with the result that something particular or more particular appears on the board (CP 6.203). First of all, I have to assume two *qualities*: the "white" of the chalk line and the "black" of the blackboard, which denote Firstness. However, the crucial point is that this chalk strip is only an actualization of one possibility. The school board occupies an unlimited variety of possibilities: the continuous vague indefiniteness of the school board is inexhaustible, irrespective of how many types and combinations of lines I draw on it. This also means that the continuity of the school board has *no a priori tendency* of any kind to crystallize in one particular direction, to crystallize into one particular entity, namely *this particular* line. The school board houses an infinite number of variations of lines that can never be predicted completely. Firstness, therefore, does not provide any positive basis for predictions in any direction, but precisely precludes such a positive basis.

Of course, the comparison of the blackboard with the *original* vague potentiality is not entirely adequate, as Peirce points out: a "blackboard is a continuum of two dimensions, while that which it stands for is a continuum of some indefinite multitude of dimensions," and a "blackboard is a continuum of possible points; while that is a continuum of possible dimensions of quality, or is a continuum of possible dimensions of a continuum of possible dimensions of quality, or something of that sort" (CP 6.203). The possibilities of actualization of the blackboard are not unlimited: the blackboard cannot actualize itself into a hamburger. The original potentiality, however, has no limits. Nevertheless, the example of the blackboard illustrates to a certain extent that it is not predetermined in which concrete forms the potential wealth of the world will be actualized: "[t]here are no points on this blackboard. There are no dimensions in that [original] continuum" (CP 6.203). Peirce stresses that "there are doubtless manifold varieties utterly unknown to us," which might never become actualized (CP 1.418).

That does not mean we cannot predict anything. It just means there will always be things that we cannot predict. The absolutely unrelated singularity of *Firsts* refers to the *impossibility* of so ordering the world that we can conceive of a limit to it. Put differently, Firstness refers to the realm of unactualized possibilities that always "transcend" the domain of our world of actual objects and facts. The world contains a dimension of potential novelty, which cannot be depleted by existing concepts and conceptualizations. This potential novelty implies that identities are not *a priori* established. Identities are the result of a particular *actualization process* and always contain "material" in order to realize other forms.

The category of Firstness strongly expresses the radical evolutionary character of Peirce's philosophy: there is an element of absolute spontaneity in all reality. The cosmos as such is subject to an element of absolute coincidence, so that the direction in which it develops can never be predicted completely. The idea that the world of possibilities is independent, that is, unlimited, ensures that an identity can never be captured or identified definitely, nor reduced to an essence. Firstness expresses the element of indefiniteness that is inherent in any form of identity, and thus also in the human self (see Hookway 1985:166; Hausman 1990; Rosenthal 1994, 51–62).

Secondness: Interaction

Although the element of Firstness is a condition for every actualization – every actualization is a realization of something that was at least partially not yet there – this category cannot sufficiently explain this process of actualization: the category of Firstness cannot explain how something undetermined is transformed in something more determined. Firstness cannot explain how actual identities manifest themselves. To account for that transformation, Peirce introduces a second category that he calls *Secondness*.

Peirce defines Secondness as "the conception of being relative to [...] something else" (CP 6.32; see also 5.66). This element of relationality "brings the subjects together, and in doing so imparts a character to each of them" (CP 1.326; see also 5.66). The following example Peirce gives provides a first indication of his understanding of relation in this regard: "God said, Let there be light, and there was light" (CP 1.327). We can make a distinction between two subjects, "God" and "light," which are causally related: "God creates light." However, the two aspects "God who commands the existence of light" and "the coming-into-being of light" are not a sequence of two distinct facts. They are two aspects of the same fact, or formulated differently: two perspectives on the same event. The relation, that is, "the *act* of creation," or seen from the other perspective, "the *coming*-into-being of light," determines the identity of the relata, "God's power of creation" and the "existence of light." The event not only precedes in a certain sense the relata, but the relata also are derived from the event in a certain way rather than vice versa. The relation does not exist between things that already have a certain identity, but that identity is generated by virtue of the relation. This example also indicates that Secondness should not be understood as a static relation but as an active or dynamic one, as a force that grants things a particular existence.

Contrary to the idea of Firstness, which is "is so tender that you cannot touch it without spoiling it," the idea of Secondness is "eminently hard and tangible" and "forced upon us daily" (CP 1.358). We experience

this element very clearly every time we are confronted with a stubborn thing or a hard fact, that is, with "something which is there, and which I cannot think away, but am forced to acknowledge as an object or second beside myself" (CP 1.358; see also 4.3), though it is also present in the "softest" experience. Secondness refers to a pre-conscious experience of resistance or confrontation (CP 1.24; 1.323).

Characteristic for actual factuality is for Peirce its spatiotemporality. Secondness "consists in its happening *then* and *there*" (CP 1.24; see also 1.532, 1.405, 3.460 and 3.434) and expresses the particular "thisness" of an individual thing. This "thisness" of a thing, or its *hæcceities*, as Peirce calls it in line with Duns Scotus, is not something that we recognize from the similarities between its qualities, nor is it something that we can conceptually determine or grasp (CP 1.405, 3.460 and 3.434; 7.266). Initially, an object does not distinguish itself from other objects by uniquely recognizable features, but by the compulsion or pressure it exerts on our attention. It is by virtue of this compulsion, this antagonistic perseverance (CP 6.318, 1.325, 7.551, 1.320 and 1.432) that it has existence and individual reality.

According to Peirce, existence, actuality and individuality are fundamentally characterized by interaction, that is, by action and counteraction (CP 1.336, 1.320 and 1.322). However, interaction in this regard must not be understood as the impact of one object and the response of another. "Interaction" in this respect is irreducible and undeductible. A further clarifying example:

> [s]uppose I try to exercise my strength in lifting a huge dumb-bell. If I strive to lift it, I feel that it is drawing my arm down. If I suffer no consciousness of having my arm pulled down, I can have no consciousness of exerting force in lifting the dumb-bell. To be conscious of exerting force and to be conscious [of] having force used upon me are the same consciousness (CP 7.543)

Of course, we can distinguish between a weight lifter that pushes a halter upward and a dumbbell that resists being pushed up. But we cannot say that *first* there is a "weight lifter" and a "halter" that *then* interact with each other. Rather, the reverse is the case: weight lifters do not lift weights, but weightlifting makes someone a weight lifter. "Halter" and "weight lifter" are what they are in and by virtue of the weight-lifting activity.

Expressed more generally, there is not a self, on the one hand, and a world, on the other, which then interact. It is precisely this essentialist view that is challenged here: "the idea of a reaction is not the idea of two *plus* forcefulness" (CP 7.266; Lachs 1980, 12). We do not first have actual things that are independent, and then they are connected to one another by some external power. Prior to interaction there are only

possibilities. Individual objects can only manifest themselves in and by virtue of interactions: "as long as things do not act upon one another there is no sense or meaning in saying that they have any" (CP 1.25).

The earlier used example of the blackboard can further elucidate this view. The clean blackboard represents, as we have seen, the original vague potentiality. I draw a line on the blackboard. What is responsible for this line? Although we can say that the school board is a continuum of potential points that could form a line, there was nothing in or on the blackboard that already possessed these dimensions in an actual way. The blackboard was clean. The category of Firstness can in no way explain that modification. The chalk line is, in fact, a plane; the only line that there is, is the *boundary* between the white plane and the black plane. That boundary between the white plane and the black plane is neither an aspect of the white nor of the black, nor of both, and neither can it be reduced to the white, nor the black, nor to any other quality. At the same time, however, white and black are recognized as white and black precisely *by virtue of* that limit. By virtue of the interaction between white and black, two different entities are generated and we can recognize the chalk line as a chalk line. Without their contrast and interaction the "white" and the "black" are mere abstractions, mere possibilities. Individual entities are the result of an activity that generates discontinuity in potential qualities: Secondness is "one of those brute acts by which alone the original vagueness could have made a step towards definiteness" (CP 6.203).

However, the individual entity that is brought about by a brute act out of the continuity of the blackboard is never completely fixed, but always displays a certain degree of indeterminateness. I can draw a second line diagonally over the first, which results in a cross. Then I can connect the end of the lines by drawing four other lines, which brings about a kind of square, and so forth. Each actualization is material for further actualization. The element of Firstness secures this inexhaustible potential richness, which is not only contained to the blackboard; each individual entity, including the human self, has potential for further determination.

Thirdness: Regularity

The idea that things and people are to a certain extent indeterminate and derive their individuality from brute interactions is insufficient to understand how they can have an identity which is more or less fixed, and which we can know and anticipate. Firstness and Secondness cannot explain how the diverse and varying reactions that a particular phenomenon prompts can be recognized as aspects of a whole. To be able to understand this, we must adopt a third element, which Peirce calls Thirdness.

Peirce defines Thirdness as "the mode of being of that which is such as it is, in bringing a second and third into relation to each other" (CP 8.328; see also CP 5.66). Thirdness is characterized as something that connects, mediates or intervenes. It is the "thread of life" (CP 1.337; see also 1.328 and 5.104), the glue that connects time. However, Thirdness should not been seen as something that connects two "parts" of time, past and future, which have a distinct reality, but it rather challenges the idea that past and future must be distinguished in such a way. The world does not consist of discrete events that are entirely self-contained, but of events that always anticipate future events (CP 1.675, 1.325, 6.142 and 6.141). Without this element, there would be no order and continuity in the things we encounter. Thirdness is a kind of "synthesizing law" (CP 1.351, 1.343, 1.536, 1.675, 1.325, 6.142 and 6.141), a kind of law that dictates how an endless future must be.

When we approach Thirdness not so much from the "outside," but more from the "inside," we can, according to Peirce, also characterize it as "thinking." A thought has a completely different character than a fact in the sense of Secondness: a fact is particular while a thought is general, not only because it can be communicated, but also because it not only applies to things that exist but also to all possible things (CP 1.420, 1.477 and 1.478). A thought should also not be confused with Firstness: the generality of Firstness refers to an unlimited variety of possibilities that do not indicate any direction. However, the generality of Thirdness refers to an idea that suggests a tendency toward realization in a *certain direction*, but can never be fully absorbed in a particular realization. Thirdness, in a sense, not only does not coincide with itself, but also constantly refers to a certain future (CP 1.342).

One of Peirce's most important but also most controversial views is that the tendency of future events to conform to a general rule does not have a mere conceptual status. Against what he calls nominalists who believe that a "general rule is nothing but a mere word or couple of words" (CP 1.26; see also 1.27), Peirce states that the indisputable fact that the general is of the nature of thought does not imply that it is also a product of our thinking; nominalists confuse "thinking" with "thought." Realists – and Peirce considers himself, especially in his later work, "a scholastic realist of a somewhat extreme stripe" (CP 1.27; see also Friedman 1997) – do not make that mistake and understand that our thinking only apprehends and does not create thought, and that thought may and does as much govern outward things as it does our thinking. Ideas can have an objective reality, which manifests itself in their predictive power.

It is indeed always possible that we encounter future events that falsify certain expectations, but that does not mean that all expectations have no basis whatsoever. Ascribing the fact that, for example, stones have fallen to the ground without verifiable exception for ages to a strange

coincidence, and therefore claiming that those experiences of uniformity provide not the slightest guarantee that the next stone that shall be let go will fall, is, according to Peirce, an absurd hypothesis. The only reasonable hypothesis is acknowledging that these kinds of uniformities are generated by some kind of orientation or predisposition, and that they are governed by some kind of active general principle or idea (CP 5.97, 5.98, 5.100 and 5.101).

The capacity of interconnecting requires a kind of intelligence or "synthetic consciousness" (CP 1.377; see also 1.381) that can be neither reduced to potential qualities nor to oppositional forces. The following analogy we find in Peirce sheds some light: by drawing a diagram, the geometrician is able to demonstrate relations between certain things that did not seem to be there before (CP 1.383). The (ideal) relations that are revealed by the diagram are in no way reducible to something else and must be recognized as real.

This element of intelligence can also be illustrated by our example of the blackboard: possible quality and contrast are insufficient to make, so to say, the white chalk line appear or manifest itself as a white chalk line; we also need to *compare* or relate the "white" to the "black" within a certain respect, in this case under the aspect of, among others, "color," which denotes Thirdness. Let's say we draw a line, and then another, and another and so on. After each line we draw, our eye triggers us to find a new line, and to recognize that this line is connected in a certain way to an earlier line, and that we can expect another line being connected in a certain way to a previous line; this goes on until all lines are connected in such a way that we recognize the final product: a square. Although the different individual lines that constitute a square have a distinct character by virtue of Secondness, Secondness is not sufficient to evoke the concept of a square in our mind. All the different lines have to be analyzed and synthesized in such a way that a continuous and unified whole, a square, is recognized. To relate the different chalk lines to one another a certain *orientation* or general *idea* is required. Ideas ascribe meaning to an object by relating its different parts to one another within a certain respect. The distinct lines gradually lose their independent individuality in this process and become part of a continuous whole (CP 6.204, 6.206, 1.492, 1.494 and 1.487). This process has no beginning or end; rather, each perception begins and ends with this process.

This view does not fit into the essentialist and dualist scheme. There is not first an independent object, which then reacts to a subject in a regulated manner and invokes an image in its mind, and there is also no *a priori* established subject, which then identifies an object by means of a special power. Interaction and regulation are not additional, but constitutive aspects of a thing. From Peirce's perspective, the identity of a thing is indistinguishable from its responses. His famous pragmatic maxim is based on this idea. In its most cited form: "[c]onsider what effects, that

might conceivably have practical bearings, we consider the object of our conception to have, then, our conception of those effects is the whole of our conception of the object" (CP 5.402; see also 1.343). What a thing is coincides with the way it reacts to us and tends to shape our behavior according to a pattern according to which it is itself formed. To determine the identity of an object, we must ask what it does and what can be done with it (see Rosenthal 1994, 31f, 40f).

However, not anything goes: Peirce's pragmatism and his realism go hand in hand. In our interaction with the things we encounter, eventually certain patterns will impose themselves on us. Such patterns or laws we then summarize under certain general terms such as "square," "tea cup" and "hard," which function as hypotheses that enable us to understand and anticipate what we encounter. General concepts are collections of laws that tell us how a certain thing will react under certain conditions (CP 5.403; Hookway 1985, 171f). This does not change the fact that – by virtue of Firstness – the identity of a thing is never completely fixed. When we encounter findings that do not confirm the general schemes that guide our reason, that is, when certain hypotheses that we apply produce a persistent incomprehensible sequence of interactions, we are forced to adjust them or further specify the conditions under which they are still valid. Such decisions result in the formation of other laws or habits that guide and regulate our future behavior and determine the identity of things. As the identity of an object coincides with its variable interactions and habits, not only our knowledge of the object, but the object as such must be understood as a dynamic process; and this also applies to our own identity (Gallie 1966, 203; Hookway 1985, 171f).

Discovering the "Self" as "Self-Formation" on the Basis of Peirce's Categories

Possibility, interaction and regularity are not only the general constitutive elements of material objects, but also of living beings, including the human self – which does not imply that these categories are equally dominant in, for example, stones, spiders and human beings (see Muoio 1984; Colapietro 1989; Andacht and Michel 2005). The category of Firstness indicates that the human self is not an invariable substance, but is fundamentally characterized by indeterminateness. By virtue of this indeterminateness, we are not necessitated by pre-given structures or instincts in the degree that animals are, and this allows for the adjustment and readjustment of our conduct as well as for real novelty.

Peirce's category of Secondness implies that the subject has only an individual identity by virtue of its natural and social interactions: we do not have a self before interacting, but can only manifest our-selves by virtue of our interactions. One could say that a person is nothing else than his or her interactions.

This seems to result in contrasting views: on the one hand, we are characterized by indeterminateness by virtue of Firstness; on the other hand, we seem to be necessitated by virtue of Secondness. We are, according to Peirce, indeed both undetermined and necessitated.

But there is more: it is not a coincidence that Peirce defines his third category as "the Idea of that which is such as it is as being a Third, or Medium, between a Second and its First" (CP 5.66). Thirdness mediates between absolute indeterminateness and sheer necessity. By bringing both these aspects together, it imparts, on the one hand, form and regularity to the indeterminate multiplicity of interactions that we are constituted by, and, on the other hand, it soaks off the rigid constraints that the world imparts on us. What we call our self is a never completely fixed multiplicity of interactions that is governed and regulated by certain habits, laws or ideas. By giving our various scattered interactions a durable form, which we can adjust continuously, we possess a certain degree of freedom.

Understanding the self on the basis of Peirce's three categories displays the self as divided and multifocal. In an article named *What Pragmatism is* (1905), Peirce defines, in the line of Plato, the thinking of a person as an inner dialogue (see TH 189e-190a and SO 263e.).[5] He writes:

> a person is not absolutely an individual. His thoughts are what he is "saying to himself," that is, is saying to that other self that is just coming into life in the flow of time. When one reasons, it is that critical self that one is trying to persuade; and all thought whatsoever is a sign […]. The second thing to remember is that the man's circle of society (however widely or narrowly this phrase may be understood), is a sort of loosely compacted person, in some respects of higher rank than the person of an individual organism (CP 5.421)

We can first learn from this passage that a person does not completely coincide with her actual individuality, with her *thisness* (*hæcceities*) at this time and in this place, in one word, with her Secondness. By virtue of Thirdness, she can transcend her actual state and possess a certain degree of self-control; she can govern and regulate her interactions with the world, and form and modify her conduct. This capacity enables her to anticipate a future critical self that she wants to convince – maybe we could call this future critical self a person's conscience.

Second, Peirce emphasizes in this passage that a man's social circle is a kind of person that is in a certain sense of a higher order. This indicates that the habits or laws that a person must incorporate in order to regulate the multiplicity of her interactions cannot be determined by sheer individual preferences, but should be discovered in a communal quest for, what Peirce sometimes calls, the *summum bonum*. This *summum bonum* is a kind of orientation or goal that is situated

in the future and enables deliberate and critical self-reflection. The future critical self is the critical voice of community (CP 1.588, 5.433 and 5.566).

Peirce not only claims that we can govern and regulate our interactions with the world, but he also presupposes that one can make a distinction between good ways and bad ways to give ourselves a certain form. The fact that (self-)criticism is possible already implies for Peirce this distinction. He presupposes that we can make a distinction between good reasoning and bad reasoning, and between good conduct and bad conduct. In short, he presupposes that there are norms, criteria and goals that indicate what is good self-formation.

Peirce believes that in our practical life, these criteria are not expressed by individual reasoning but by sentiments. The sentiments that are expressed by moral rules, however, are not individual feelings, but in a certain sense exactly the opposite; for Peirce, a moral code is ideally an "instinctive or sentimental induction summarizing the experience of all our race" (EP II.32). Individual reasoning is much more susceptible for mistakes than social instinct, because it is often influenced by various particular, contingent circumstances that narrow its scope. Social sentiment, however, is the result of a long process of interaction between correction and modification of different perspectives (de Waal 2006). That is why Peirce can say: "It is the instincts, the sentiments, that make the substance of the soul. Cognition is only its surface, its locus of contact with what is external to it" (CP 1.628). In short, the criterion for what is good self-formation should not be situated in individual reasoning, but in social sentiment.

Because instincts are the result of beliefs that were incorporated after having proved themselves for ages, they deserve to be taken very seriously. Ethical theories (both utilitarian and deontological) that are too one-sidedly rational do not take enough into account that the worth of rational judgments that are established on the basis of rational calculations always depend on what is conceived as and has preserved itself as admirable in a certain culture. This also explains, as we will see later on, why for Peirce the task of ethics is not to establish directly whether a decision is morally right or wrong, but rather to investigate under which conditions an adequate (sentimental) disposition can be developed and a self could form itself in the right way.

Self-formation and Ideals

Self-formation presupposes certain norms and goals. These norms and goals however are, as we saw above, not situated in individual reasoning, but can only be discovered and embodied through a long process of social interaction. Because thinking is not an isolated, theoretical process but fulfills a certain function in our interaction with the world, there

is a direct relation between (correct) thinking and (reasonable) action. Peirce writes:

> Thinking is a kind of action, and reasoning is a kind of deliberate action; and to call an argument illogical, or a proposition false, is a special kind of moral judgment (CP 8.191)

Correct thinking is a form of dealing successfully with the resistances that we encounter in the world. Within this context, a proposition can in principle be understood as an expression of a belief and a belief can be understood as an intellectual habit. And because a habit is a willingness to act, a proposition can be understood as a formulation of a willingness to act. In addition, a willingness to act anticipates on a certain expectation, on a goal that wants to be realized. The meaning of a proposition refers, therefore, in the end to a goal that somebody is attracted to (see Debrock 1992, 11f).

The normative sciences, which are arranged under phenomenology in Peirce's classification of the sciences, study the norms that should be respected in critical thinking and action. Logic studies and formulates the rules that should be followed in reasoning. Rules, however, must refer to a goal that justifies why those rules should be followed (CP 2.198; see also Potter 1997, 34). Ethics studies the goals upon which thinking and action should be directed. Richard Bernstein claims justly that Peirce gives here his "own version of the primacy of practical reason" (Bernstein 1972, 193).

This also means that, for Peirce, the object of ethics is not action, nor right action. He writes:

> The fundamental problem of ethics is not, […] What is right, but, What am I prepared deliberately to accept as the statement of what I want to do, what am I to aim at, what am I after? To what is the force of my will to be directed? (CP 2.198)

The task of ethics is to discover the goals that can generate certain habits by virtue of their attractive power, habits that could increase the chance of making morally right decisions, that is, increase the chance to explain and overcome the resistances that we encounter in the world. Ethics is the quest for goals and ideals that enable desirable self-formation.

We see here that self-formation not only presupposes that we do not completely coincide with our actual state, but also that we have a capacity to guess and predict that can be further developed. This capacity confirms for Peirce that there must be a basic affinity between the human mind and the world. Peirce writes: "retroduction goes upon the hope that there is sufficient affinity between the reasoner's mind and nature's to render guessing not altogether hopeless, provided each guess is checked by comparison with observation" (CP 1.21; see also CP 6.417

and 6.477). Further cultivating this basic affinity will enable developing ever better ideas with a predictive power, anticipating our future conduct and increasingly possessing self-control. That is why Peirce says: "A rational person […] not merely has habits, but also can exert a measure of self-control over his future actions" (CP 5.418; see also Hookway 2000, 14). Self-control is, according to Peirce, a complex process, which roughly involves the following elements:

> first, in comparing one's past deeds with standards, second, in rational deliberation concerning how one will act in the future, in itself a highly complicated operation, third, in the formation of a resolve, fourth, in the creation, on the basis of the resolve, of a strong determination or modification of habit (CP 8.320)

We can learn from this passage that self-control is a process of self-criticism. Self-criticism in this respect means first of all that an actor reviews each of his (important) actions, comparing them with certain standards or ideals that he values highly and wishes to realize. Peirce writes: "His ideal is the kind of conduct which attracts him upon review" (EP II.377). Next, the actor judges how he will act in the future and whether he wants his future conduct to be in accord with the ideal in which he thoroughly believes. His decision to devote himself to the realization of that ideal will influence his disposition so as to modify what he is naturally inclined to do and will, in the long run, result in the modification or formation of a (new) habit.

Peirce emphasizes that an ideal of conduct should not be confounded with a motive to action. Every action has a motive. An ideal, however, refers to *deliberate* conduct. Although a habit may modify future action, it is not a moving cause of action. A habit is a kind of mental formula that predicts how one will act and wish to act under certain conditions. The distinction that Peirce makes between a motive of action and an ideal of conduct corresponds to the Aristotelian distinction between efficient cause and final cause. The great importance that he attaches to the development of a reasonable disposition brings Peirce even closer to Aristotle's virtue ethics.

There is a kind of reciprocal reinforcement between ideals and the formation of habits: the more we devote ourselves to a certain ideal, the more we will be able to form and cultivate certain habits; and the more we form and cultivate a certain habit, the more we will be sensitive to the attractive power of the ideal that we wish to incorporate. In reality, the incorporation of ideals is identical to the formation of habits; we can only disconnect the two processes from each other *in abstracto*.

This account of the process of self-control requires another addendum: because reasoning can be understood as a certain form of interacting with the world, "self"-criticism cannot be a sheer individual process

(Bernstein 1972, 190; Bernstein 1980, 82). Peirce's notion of Secondness and his claim that thinking is a form of internal dialog presuppose a community with certain norms and ideals. Reasoning is, as we saw earlier, an attempt to persuade a critical self: "that other self that is just coming into life in the flow of time" (CP 5.421). Because that critical self is principally situated in the future, the community of critical minds cannot be an existing, finite community, but must be a community "without definite limits, and capable of a definite increase of knowledge" (CP 5.311).

To sum up: (1) the ideals that we anticipate through critical reasoning can only be discovered by virtue of a real and virtual interaction with our social and natural world; (2) these ideals can never be completely realized, but remain in principle always subject to criticism; and (3) real criticism, however, must again presuppose certain ideals that cannot be completely dependent on subjective preferences, since those ideals must prove themselves in the interaction with our social and natural environment.

And Peirce makes another decisive step: not only does critical conduct necessarily presuppose certain ideals, but those ideals must presuppose an ultimate ideal that is admirable in itself. Ethics, Peirce says, shows how deliberate conduct is governed by certain goals and ideals and how these ideals modify our conduct. It, however, cannot tell us what is the state of things that is most admirable as such. Ethics is, according to Peirce, dependent on another discipline, which he calls "esthetics."

Self-formation as an Esthetic Ideal: Organized Heterogeneity

What is the ultimate ideal, that is, the "admirable without any reason for being admirable beyond its inherent character"? (CP 1.612; see also 1.191). Peirce's answer to this question is: reason! This answer is less traditional than it appears at first sight. Contrary to the view of many traditional and modern philosophers, Peirce does not consider reason (which is equivalent to Thirdness) to be a personal or subjective or even a specific human faculty. Although human beings to some extent embody reason as something manifesting itself in the mind, reason is for Peirce at the heart of nature or evolution itself. He writes:

> The creation of the universe, which did not take place during a certain busy week, in the year 4004 B.C., but is going on today and never will be done, is this very development of Reason. I do not see how one can have a more satisfying ideal of the admirable than the development of Reason so understood (CP 1.615)

From this evolutionary point of view, Peirce states that the only entity that in its essence always is directed at endlessly improving its results is reason itself, although it can never reach complete perfection. Because human reason and conduct are part of a larger process, the development

of an ever better disposition contributes to the ultimate ideal of making the world more reasonable. In Peirce's words:

> Under this conception, the ideal of conduct will be to execute our lit-tle function in the operation of the creation by giving a hand toward rendering the world more reasonable whenever, as the slang is, it is "up to us" to do so (CP 1.615)

In order to understand Peirce's view on the ultimate ideal within this context, we have to take as starting point his conception of the relation between thinking and action. The experience of an object coincides, as we saw in the above discussion of Peirce's phenomenological categories, with the interactions (Secondness) that it prompts. An object, however, can only be given a certain meaning by virtue of linking its individual parts to one another in a certain respect (Thirdness). But clearly, this function of Thirdness coincides with the function of ends: an orienta-tion toward an end relates different parts to each other, thus generating a unity, which is a necessary condition for the occurrence of meaning. Because it is always possible that we are confronted with new, unpre-dictable events (Firstness), which destabilize our conception of a certain object, the quest for the perfect orientation will never come to an end.

That quest for the perfect orientation consists concretely in the de-velopment of ever better habits. There is, as we saw earlier, a recipro-cal relation between the pursuit and embodiment of ideals, on the one hand, and the formation of habits, on other hand. We give our world meaning by virtue of the orientation toward certain goals. Those goals, however, can only generate meaning inasmuch as they are and can be embodied in our concrete conduct. Embodied goals and ideals become habits of action and thought. Those habits are subject to criticism, but that criticism can, as we saw earlier, only be authentic if it presupposes certain ideals that are pursued. Those ideals are also subject to criticism, but again that is only possible in the light of yet a higher or other ideal that we want to realize and so on and so forth.

Peirce believes that this accumulation of criticism can gradually bring about a "habit of feeling":

> the ideal must be a habit of feeling which has grown up under the influence of a course of self-criticisms and of heterocriticisms; and the theory of the deliberate formation of such habits of feeling is what ought to be meant by esthetics (CP 1.573f)

The different degrees of self-criticism correspond to different degrees of self-control:

> When a man trains himself, thus controlling control, he must have some moral rule in view, however special and irrational it may be.

> But next he may undertake to improve this rule; that is, to exercise a control over his control of control. To do this he must have in view something higher than an irrational rule. He must have some sort of moral principle. This, in turn, may be controlled by reference to an esthetic ideal of what is fine (CP 5.533)

The ultimate ideal consists of developing an ideal that justifies all other particular ideals and relates them to one another in a harmonious way. This explains the esthetical character of the ultimate ideal. Peirce writes:

> I should say that an object, to be esthetically good, must have a multitude of parts so related to one another as to impart a positive simple immediate quality to their totality; and whatever does this is, in so far, esthetically good, no matter what the particular quality of the total may be (EP II, 201)

It is by virtue of the development of "habits of feeling" that we are able to become ever better "in tune" with the world, prompting it to give up its secrets and to acquire evermore meaning. The development of a sense for the esthetic ideal entails the formation of a practical and emotional disposition that enables the orientation toward goals by virtue of which evermore meaningless, useless and inefficient aspects of the world are related to one another in such a way that they become meaningful, useful and efficient.

We see here that the basic affinity between the human intellect and the world, which is a necessary condition for the possibility of acquiring knowledge in the first place, is important not only at the relative beginning of the quest for knowledge, but that it is also a structural component of knowledge production. The formation of, respectively, "habits of thought" and "habits of action" leads to the further development of this basic feeling or instinct, that is, to the development of "habits of feeling," which prompt further developments and so on *ad infinitum* (see CP 1.648). The ultimate ideal appeals to a maximally developed disposition, which enables us to *feel* toward which ideals we have to orient ourselves with and which particular habits we have to develop with in order to discover ever better ideals. Consequently, this ultimate ideal shapes the course of our life and our self. That is why Peirce can say: "the good is the attractive, – not to everybody, but to the sufficiently matured agent; and the evil is the repulsive to the same" (CP 5.552).

This view requires, however, a further refinement: the regulation of the multitude of parts by a simple quality, form or idea must not be achieved at the expense of the multitude. Peirce warns that self-control and self-regulation should not be excessive. In one of his manuscripts, he writes: "See that self-government is exercised; but be careful not to do violence to any part of the anatomy" (M 675, 15–16; see also Colapietro

1989, 111f). Although excessive regulation can go at the expense of variability and multiplicity, diversity can only be preserved and intensified *by* submitting it to a form or idea. The development of concrete reasonableness involves the differentiation of the undifferentiated, the evolution toward what Peirce sometimes calls "'organized heterogeneity', or, (...) 'rationalized variety,'" (CP 6.101; see also 6.191). Our moral task is to contribute to this evolution.

Peirce understands this cosmic evolution as an expression of the principle of "evolutionary love" or "creative love," which signifies the principle of relating things to one another without destroying their independence. In doing so, he refers to the belief that "God is love" as set out in the gospel of St John, whom Peirce calls the "ontological evangelist" (CP 6.287). In an article entitled *Evolutionary Love*, Peirce writes:

> The movement of love is circular, at one and the same impulse *projecting creations into independency* and *drawing them into harmony*. This seems complicated when stated so; but it is fully summed up in the simple formula we call the Golden Rule. (...) It is not by dealing out cold justice to the circle of my ideas that I can make them grow, but by cherishing and tending them *as I would the flowers in my garden*. (...) Love, recognizing germs of loveliness in the hateful, gradually warms it into life, and makes it lovely [my italics] (CP 6.288f)

On the one hand, we must pursue and embrace as much multiplicity as possible; but on the other hand, we have to relate the different parts to one another so as to establish a harmonious unity. For Peirce, this is what genuine love does. Harmonious unity cannot be generated by exercising brute force (Secondness), but only by conjuring, by evoking, arousing, infecting (Thirdness). That is why Peirce sometimes says that though ideals do not have force, they do indeed have power. They stimulate the thing evoked to realize its potential to its maximum. Peirce somewhere compares this passive reinforcement with the way in which the sun prods a flower to realize its potential (see CP 2.274 and 5.520). Hard facts, brute interactions are transfigured into meaningful life by imparting on them a form by virtue of a certain ideal.

Consequently, synthesizing a multitude without destroying it is only possible by virtue of the orientation toward the incorporation of goals and ideals that favor the inclusion of as many parts of the (social) organism as possible. The realization of a good life and good self-formation consists in the continuous incorporation of and devotion to evermore inclusive ideals. For, even though ideals may sometimes do harm to certain aspects of life or even exclude other important ideals, it is our moral task to pursue the ultimate ideal, which can be defined in the words of Vincent Potter, as the ideal that "can be consistently pursued in any and

all circumstances" (Potter 1997, 49). The failure to pursue an ultimate ideal amounts to settling for ideals that do violence to certain aspects of life and fail to prompt their development. Peirce writes in this line: "The only moral evil is not to have an ultimate aim" (CP 5.133).

Although the ultimate ideal is the final ground of moral judgment, it does not represent an *a priori* principle that is beyond every form of critique, but it has, as it were, a *post*-critical status: it refers to a phase that could be reached after an endless process of "self-criticisms and of heterocriticisms." That the ultimate ideal can be "consistently pursued in any and all circumstances" means that it does not encounter any resistance within an endless community of critical minds. It means that it has generated a perfect emotional disposition that enables us to be perfectly in harmony with our surroundings. While the achievement of such an ultimate phase is based on "nothing more" than hope, it is a presupposition of crucial importance: on the one hand, real criticism is only possible in the light of such an ultimate ideal, and, on the other hand, the ultimate ideal prevents absolutist claims that bring further criticism and improvement to a halt. The ultimate ideal can be conceived as a regulative principle (CP 1.405, 1.173, 1.121, 6.610 and 4.61).

Self-formation Beyond the Individual Self

It is now clear that self-formation does not involve the development of an isolated, subjective, personal disposition, but rather of an orientation about which in the long run everybody would agree on. This explains Peirce's often almost hostile attitude toward individualism. In a manuscript written in 1906, he contends:

> Now you and I – what are we? Mere cells of the social organism. Our deepest sentiment pronounces the verdict of our own insignificance. Psychological analysis shows that there is nothing which distinguishes my personal identity except my faults and my limitations – or if you please, my blind will, which it is my highest endeavour to annihilate (CP 1.673)

If one reads this passage (and other passages) accurately, one discovers that Peirce's aim is not the destruction of the individual self, but rather the situation of the individual (the cell of the organism) within a broader whole (the social organism). According to Peirce, we should not concern ourselves too much with questions about what we as individual moral actors should do or should not do. We should rather consider our thinking and acting as part of a broader development with a common history and common future. What is important is not what I think, but what eventually should be thought. From that perspective, the very first command is that we ought to acknowledge that there is a higher business

than our own. For Peirce, this higher, common business is not something to be performed after one is finished with all one's individual tasks. For one's personal duty is part of a more general duty. Only such generalization of "duty" can complete one's personality by "melting it into the neighboring parts of the universal cosmos" (CP 1.673; see also CP 1.631 and 1.639).

So, conceived generalization is inherent to the pursuit of common ideals. Only by embodying ideals that in the long run everybody would agree on can we make the world more reasonable. As Peirce sees it, true reasoning, far from being synonymous to cold, logical thinking, amounts to being welded into the universal continuum. To generalize is to embody common ideals in the deepest emotional layers of our life (see CP 1.673).

There is some similarity between Peirce's notion of generalizing and Kant's categorical imperative, inasmuch as both notions involve the idea of overcoming an individual-subjective standpoint (see Colapietro 2006, 182). But there are also great differences between the two philosophers. For Kant, the categorical imperative appeals to and coincides with the rational part of the human being, which is completely detached from its empirical nature. In addition, the *a priori* status of Kant's categorical imperative is completely autonomous and, therefore, completely beyond our control. For Peirce, however, the ultimate aim that ought to govern our conduct does not coincide with our actual rational nature, but it is something to be attained in the *future* by an unlimited community (see Erny 2005, 134). For him the necessary condition for moral conduct is not *knowledge* of what our duty is, but the adoption of a self-reflexive and self-controlled *attitude*. By the "indefinite replication of self-control upon self-control" (CP 5.402 n.3), we can make our very *empirical* world more reasonable.[6] The basis of Peirce's notion of self-formation cannot lie in knowing what the good is, for the function of ethics consists precisely in the search for what the good *could* be. We can never know with certainty that the goal we pursue will ultimately turn out to be a good goal. The only requirement is that the goal we pursue *can* be continuously pursued as an ultimate aim.

Here again, Peirce's ethical perspective both resembles and differs from that of Kant. Peirce agrees with Kant that we can never be completely ascertain that our conduct is good. But he has a completely different reason for this claim. Kant connects this claim to the fact that it is always possible that our conduct is not exclusively motivated by the categorical imperative, but may very well be the result of our natural inclination. For Peirce the reason for never being completely certain that our conduct is good is not due to an alleged discrepancy between reason and nature, but to the fact that the good is not something that is already present; the good for Peirce is, as we have seen, something that is the object of a common quest and has to be developed through a process of embodiment.

Moreover, although there are certain similarities between Kant's Kingdom of Ends and Peirce's ethical ideals, their divergent notion of rationality creates an unbridgeable gap between them.

Peirce's concept of generalizing and its implications for his view on the individual and individuality may go against our modern taste. But one must take into account that from Peirce's point of view, we are not born as unique, independent and self-sufficient individuals in some kind of liberalist meaning of the word. We are what we are by virtue of our interactions. And we can only become persons through a process of giving a general form to our interactions. In other words, we can become persons only by virtue of the formation of social habits, that is, by virtue of the belief in and embodiment of common ideals, which are no less real than concrete, physical events.

Having said this, I don't want to circumvent Peirce's anti-individualism. Even though Peirce emphasizes that the process of generalization or habit formation and the incorporation of ideals can only be realized in the concrete conduct of individual entities, his concept of concrete reasonableness implies that the particular lives of people are in a certain sense "mere" means in the completion of a cosmic, evolutionary development. Peirce can be considered an anti- or posthumanist philosopher in the sense that he does not take the human being to be the final product of evolution. Indeed, what differentiates a human being from lower animals is his ability to contribute to something higher, to something more developed. For Peirce, becoming a self is possible only if we seek beyond our actual, individual lives and become part of a cosmic process that transcends us: only by overcoming our disconnected and fragmented human individuality can we become genuine persons. Maybe this view makes Peirce a humanist in a different and even a better sense.

Notes

1 Relating Nietzsche to pragmatism might seem odd at first sight, though it is less extraordinary as it looks. At the start of the 20th century, Nietzsche's thinking was frequently related to pragmatism (Berthelot 1908, 1909, 1912; Stein 1908; Warbeke 1909; Müller-Freienfels 1912; Kleinpeter 1912; Waibel 1915, esp. 28, 29, 42f., 48, 66, 71f; Scheler 1926/1977, esp. 44–47; Durkheim 1993, 10–168), but that relation was never worked out thoroughly. After a long silence, the last decades have seen contemporary scholars pointing out similarities between Nietzsche and pragmatism (Granier 1966, esp. 482–498; Abel 1984, 175f; Hallman 1985; Steilberg 1996, 228–234), but again without elaborating fundamental implications (the dense study by Hingst 1998 forms an exception). There is, however, virtually no research that relates Nietzsche's thinking to that of Peirce. If Nietzsche's thinking is associated with pragmatism, then the focus is often on similarities regarding the concept of truth, especially regarding James' notion of truth (Danto 1964, 384–401, 567f, 588; Danto 1965/1980, esp. 72, 79f., 98f., 191f; Donadio

1978, esp. 15f, 31f., 42f; Wilcox 1980; Rorty 1982, XVIII, XLII, 150–153, 155–158, 161–166, 205; Wilcox 1986; Rorty 1989, 10, 39, 46; Rorty 1991, 2f., 117f).

2 Nevertheless, I keep using, where it is possible, for frugality the term "will to power." "Will to power" however always means "will to power"-organizations.

3 See, for a discussion of epistemological issues related to this view, Stegmaier 1992, especially 345–351.

4 In 1867, R. Mayer published a collection with the title *Mechanik der Wärme*, which contained, among others, his groundbreaking book *Bemerkungen über die Kräfte der unbelebten Natur* from 1842 as well as a new article *Über Auslösung*. Nietzsche appropriates Mayer's notion of "discharge" in his view of struggle and organization with regard to the interaction between wills to power.

5 There are intriguing relations between Peirce's view on the self and the notion of the so-called "dialogical self" that has received growing interest in the last decades, especially from scholars who work on the border of psychology, sociology and philosophy. See, for an illuminating overview, Wiley 2006.

6 See, for a convincing interpretation of the "flesh and blood"-embodiment of habits with respect to the formation of our practical, historical identity, elaborated within a critique of Kant's purely formal identification of the self with an utterly abstract law, Colapietro 2006.

Conclusion Part I

In the first chapter of this part of the book, I discussed essence ontology on the basis of the views of Plato and mainly Aristotle. Then, I showed how appropriating this Aristotelian ontology, Descartes developed an essentialist and dualist notion of the self, which has greatly influenced the Western view of the self. The human self is conceived as an agent that is essentially independent of its social and material environment. In addition, the self is primarily identified with a mental, inner realm that is capable of detaching itself from the outside world. The Cartesian distinction between "inside" and "outside" also safeguards freedom and responsibility; although it is recognized that the world influences our behavior, the existence of an inner realm secures the possibility of determining our own actions and decisions and, therefore, also the possibility to ascribe those actions and decisions to ourselves.

This essence ontology cannot account, as we have seen, for the constitutive role of interactions and the possibility of real novelty, which are, from a phenomenological and practical everyday life point of view, necessary dimensions of selfhood. I have argued that Western thinking seems to be schizophrenic with regard to the question of the self. In theory, it seems very difficult, almost impossible, to overcome the essentialist and dualist approach to the self: the idea that first there are independent things that then relate and interact with one another seems to be engraved in the very grammar of our language and thinking, which makes conceptualizing the primacy of relationality a very difficult, almost impossible task. However, in our practical encounters, we experience the constitutive influence of relations and interactions everywhere; we find ourselves, both physically and virtually, in different, ever-changing webs of relations and networks, which greatly determine our actions, decisions and identity.

As I indicated earlier, I am certainly not the first who has challenged the essentialist and dualist view of the self. From a certain angle, all philosophy from the second half of the 18th century on, and more explicitly and strongly from the 19th (with Nietzsche and the pragmatists) and 20th centuries on (with the postmodernists), could be seen as an attempt to overcome essentialist and dualist views and to provide more suitable

alternatives. Many "multiplicity and dividedness" thinkers (often histor-ically influenced by Nietzsche!) focus primarily on the view that the self is *fragmented and divided* by variable and instable relations, which is often valued as a good thing. Although often these approaches do justice to the importance of multiplicity and change in self and self-formation, they are insufficiently able to account, within a coherent framework, for the elements of organization and regularity, without which self and self-formation would be incomprehensible. Diana Fuss has argued that poststructuralists and postmodernists are inclined to situate difference and dissimilarity *outside* the self, as a result of which the self falls apart into radical heterogeneous pieces (Fuss 1989, 103). In contrast, I believe the proposed "event ontology" on the basis of Nietzsche and Peirce's views situates variable interactions not outside, but rather *within* the self, in the sense that they structure and constitute it. An adequate non-essentialist alternative needs to be radical, but at the same time sufficient and balanced: not only does it have to account for change and multitude, but also for organization, unity and durability.

Another way of being radical but too one-sided is assuming that repu-diating the essentialist and dualist view of the self amounts to rejecting the existence of the self altogether. A helpful distinction that we can make in this regard is that between the anti-realist and the realist position with regard to the self (Zahavi 2015). A radical adaptation of the anti-realist position we can find in Metzinger's book, *Being No One* (2003). Draw-ing strongly on neuroscientific research, Metzinger denies the existence of the self altogether, as he claims: "no such things as selves exist in the world. Nobody ever *was* or *had* a self" (2003, 1). Very much in line with Zahavi (2015, 1), we can ask what exactly it is that Metzinger takes himself to be denying. It becomes clear that Metzinger understands the self as an ontologically invariable, independent and unified substance (2003, 626; 2011, 280). Denying the existence of such a substance, for him comes down to rendering the self an illusion. Although Metzinger admits that the no-self view comes in different flavors and strengths, he fails to see (see also Zahavi 2015, 1) that the same holds true for the "no no-self alternative," since only if the untenable account of the self under-stood as an unchanging, independent and unified substance or essence were the only possible realist account of the self would an anti-realist no-self account of the self be the only possible defendable alternative. On reflection, Metzinger's skeptical view seems to be not that contro-versial, because the essentialist notion that he rejects is already largely abandoned not only by most of the empirical researchers who these days investigate the development, structure and function of the self, but also by most 20th-century continental philosophy, as also Zahavi (2015, 2) points out.

Although I agree with Zahavi that most 20th-century (continental) philosophies have attempted to overcome the essentialist view, I believe

it is much more persistent and stubborn than Zahavi seems to assume. Many attempts to overcome it have resulted in rehabilitating and sustaining it by other means. This is especially the case within the field of philosophy of technology, as I will demonstrate in the rest of this book. In order to provide an alternative, more adequate notion of the self, it is pivotal to sufficiently understand the arguments that the essentialist and dualist approaches provide to justify its outcomes. Not taking those carefully into account results in developing approaches that seem radically different but are not that different, or are radically different but too one-sided. The real challenge is not only rejecting the essentialist account of the self, but also providing a balanced alternative.

After having given a detailed account of the essentialist and dualist view of the self and explained why that approach is alluring but untenable, I have discussed an alternative, and I believe a balanced approach, on the basis of Nietzsche's notion of the will to power and Peirce's categories and normative conceptions. Because this "event ontology" forms the basis for further developing my interactionist account of self-formation in a technological world, I will now indicate how the combination of Nietzsche and Peirce's views provide a firm foundation for the rest of this study (see also Aydin 2006).

First, both Nietzsche and Peirce try to do justice to the possibility of the emergence of real novelty; they believe that the world possesses the possibility to actualize itself in an unlimited multitude of forms that might be completely beyond our current comprehension. Although Nietzsche does not elaborate this element of radical novelty in a very explicit and systematic manner within his "ontology" of the will to power, his conception of reality as a multitude of contesting wills to power, both internally and externally, indicates that one can never completely predict what will be the outcome of a power struggle and, therefore, in what kind of forms reality will actualize itself. In his anthropology of the human and Overhuman (*Übermensch*), this dimension of indeterminateness is, as we will see in third part of this book, more extensively and explicitly elaborated: the human is defined as the "not yet determined animal" as well as a being that is capable of continuously overcoming her present state. In Peirce's philosophy, the possibility of radical novelty is systematically elaborated within his doctrine of the categories. His category of Firstness indicates that things always contain an element of potential novelty, which rules out the possibility to predict with certainty in which direction things will actualize and develop themselves. From Nietzsche and Peirce's perspective on radical novelty, in combination with the idea that interactions have a primordial, constitutive influence, a teleological view à la Aristotle is unsustainable; the idea that things contain an *a priori* given essence that determines both their identity and the course of their development is precluded.

This element of indeterminateness is sometimes also elaborated in a more epistemic context. For Nietzsche, every proposition is fundamentally hypothetical, in the sense that there is always the possibility that it will be undermined or corrected (in his language: subdued) by other or future interpretations. Every form of knowledge exists by virtue of a power struggle between competing interpretations, which never reaches a final end (GM II, 12). For Peirce, every type of knowledge, including the laws and mechanisms accepted by scientists, could at any given time turn out to be incorrect or insufficient after all. What is true and real is dependent on what the open-ended community will accept in the long term, but cannot be fixed by what any finite group of believers thinks at some given time (CP 5.172, 591; see also Haack 1992, 28f; Meyers 1999, 642).

Second, for both Nietzsche and Peirce, interactions are not additional but constitutive aspects of reality. Struggle is at the heart of Nietzsche's notion of the will to power: power is only power insofar as it can maintain itself against other powers and strives to subdue them. Peirce's notion of Secondness entails that individual entities can only manifest and actualize themselves in and by virtue of interactions.

The importance of interactions is also expressed in both Nietzsche and Peirce's conceptions that the meaning of a thing is completely exhausted by listing the practical consequences that it implies. For Nietzsche, this idea is, as we have seen, contained in the structure of power: "A power quantum is characterized by its effect and its resistant" (KSA 13:14[79]). In other texts, this view is expressed more explicitly and clearly. Nietzsche states that "a 'thing' is a sum of its effects, synthetically bound together by a concept" (KSA 14[98]13.275; see also 2[85] 12.104). Along similar line, Peirce states that "[f]or as long as things do not act upon one another there is no sense or meaning in saying that they have any being" (CP 1. 25). Peirce's pragmatism attempts to explain what this view entails, as formulated in his famous pragmatic maxim: "Consider what effects, that might conceivably have practical bearings, we consider the object of our conception to have, then, our conception of those effects is the whole of our conception of the object" (CP 5.402). In a similar way Nietzsche states:

> A thing = its qualities; but these equal everything which *matters* to *us* about that thing: a unity under which we collect the relations that *may be of some account* to us. (…) the "object' is the sum of the *obstacles* encountered that we have become *conscious* of (KSA 2.77)

For both Nietzsche and Peirce, things derive their existence as well as their meaning from their relations and interactions. It must be clear that for Peirce, as also expressed in his pragmatic maxim, interactions are a necessary but not a sufficient condition for generating meaning.

Third, organization and regularity are for both Nietzsche and Peirce structural elements of reality. At first sight, Nietzsche seems to be primarily concerned with the importance of struggle, which can be explained by the urgency that he promoted regarding liberating humans from (Christian and humanist) conceptions that try to fix their identity permanently and deprive them from the possibility of developing other and evermore diverse and rich life forms. This also explains why Nietzsche's philosophy has foremost an antagonistic character. As a consequence, he is suspicious of explanations that provide a more positive account of what does constitute the world and self. However, on closer examination the notion of organization is, as we have seen, of no less significance for him than that of struggle and change: if the ruling will to power is not able to organize what has been integrated, decline sets in. Although the world is characterized by struggle, variability and multiplicity, things have *being* insofar as struggle, variability and multiplicity are organized in a particular way.

In Peirce's account, organization and habit formation are strongly highlighted and structurally elaborated; they provide important directives for understanding how identities can have a durable form. His category of Thirdness expresses that a thing can only have a durable and unified reality and meaning if various, scattered interactions are given a particular form. The different parts of a thing have to be synthesized in such a way that a continuous and unified whole is recognized. To relate the different parts of a thing to one another a certain orientation or general idea is required. "Ideas" ascribe meaning to an object by relating its different parts to one another within a certain respect. The different parts gradually lose their independent individuality in this process and become part of a continuous whole (CP 6.204, 6.206, 1.492, 1.494 and 1.487). It should be stressed that for both Nietzsche and Peirce, organization is not a noun but a verb; it is an *activity* that does not exclude but includes multiplicity, change and novelty.

Despite their anti-essentialism and "fallibilism," Nietzsche and, to a stronger extent, Peirce do not endorse, and this could be considered as a fourth similarity, radical relativism. Although Nietzsche often repudiates every type of truth, especially in his radical criticism of metaphysical doctrines, his notion of the will to power indicates that within a (temporarily) fixed power configuration certain rules do apply, which enable (always provisional and conditional) certain truth claims (see Aydin 2003, Chapter 7). It must be noted here that the extent to which Nietzsche is or is not a relativist also depends on what exactly is meant by "relativism." In this respect, Peirce is more outspoken. Against what he calls "nominalism," he passionately defends his "scholastic realism," which is the doctrine that reality "though it has no being out of thought, yet it is as it is, whether you or I or any group of men think it to be so or not" (CP 8.153; see also 8.12). This conception distinguishes him from

pragmatists like James and Schiller, and from contemporary pragmatists like Rorty (see Rorty 1982, especially 3–18, and 1972, 649–665). Peirce's realism should not be confused with Platonism, which treats universals as abstract particulars (CP 8. 10). The pragmatic maxim explains general terms by means of conditionals expressing "habits," that is, general rules of interaction to which future events conform (CP 5.90, 107 and 105). This has also implications for the human self: although the self is never completely fixed, it is fixed to a certain extent by habit formations that are generated by certain deep-rooted and proved beliefs, goals and ideals.

Nietzsche and Peirce's event-ontology provides basic building blocks for a normative view of self-formation. A reflection on Nietzsche's notion of "decadence" in the context of his "ontology" of the will to power offered some parameters for understanding an ethics of self-formation. On the one hand, a strong "will to power"-organization is characterized by a strong internal dividedness; that is why Nietzsche says: "The greater the urge to unity, the more weakness we shall encounter; the greater the urge to variance, difference, internal decay, the more force we find there" (KSA 11:36[21]). A great internal dividedness requires that the parts are sufficiently "independent" to keep each other in a certain equilibrium. Put differently, a certain balance between parts prevents the multiplicity – and with that, the struggle and tension – from being destroyed: "If two wants to remain two, an […] equilibrium is needed" (KSA 12:5[82]). On the other hand, those "independent" parts must be related to one another in such a way that they form a whole. When that relatedness is missing, decay and decadence will set in. In Nietzsche's idiom: "[t]he whole no longer lives at all: it is composite, calculated, artificial, and artifact" (CW 7). Thus, a well-organized struggle requires that a "will to power"-organization is internally strongly divided, and at the same time that multiplicity is forced into a synthesis. Only then a strong enough tension can be generated that can be discharged on another "will to power"-organization. "Good self-formation" seems to entail the paradoxical challenge to allow as much dividedness and multitude as possible and, at the same time, to strongly organize that dividedness and multitude into a particular form.

Peirce's ethical-esthetical expansion on the basis of especially his category of Thirdness arrives at a similar but more systematically elaborated account of the ideal of self-formation. Peirce's realist notion of ideas, goals and ideals explicitly displays self-formation as an ethical-esthetical challenge: forming ourselves is only possible if we impart a form to ourselves by virtue of an orientation toward certain goals and ideals. Esthetic good self-formation is characterized by relating a multitude of parts to one another in such a way that their independence is recognized, and at the same time, a certain unity and harmony is imparted to their totality. This ideal does not exclude (self)criticism, but is

a "habit of feeling which has grown up under the influence of a course of self-criticisms and of heterocriticisms" (CP 1.573f). Good self-formation consists in the continuous incorporation of and devotion to ever more inclusive ideals, which requires cultivation of a disposition that is responsive to an infinite (internal and external) community of critical minds. As I will show in the rest of this study, this idea of self-formation as organizing a multitude by virtue of an esthetic ideal will prove itself fruitful with regard to the question of what is meant by "good self-formation" as well as with regard to the question of whether and how we can technologically form ourselves.

In order to make the contours of my approach sharper in the context of contemporary discussions with regard to the self, I will briefly take up again the conversation with Zahavi. Zahavi (2015) proposes an experiential minimalist notion of the self that approaches selfhood as a built-in, integral feature of our experiential life: my experiences do not only take place "in me," but they are also "for me," which amounts, he believes, to a primitive and minimal, first-personal and pre-reflective form of selfhood. He also argues that this experiential minimalism does not exclude social constructivism. For the social constructivist approach, being a self is an achievement rather than a given: to be a self is to stand in a particular interpretative and evaluative relation to oneself; the self is constituted by choices and actions (Korsgaard 2009) or by organizing and unifying our experiences and actions according to narrative structures (Taylor 1989). Not only do both experiential minimalism and constructivism reject the essentialist account of self espoused by anti-realists like Metzinger, Zahavi argues, but they also complement one another: who we are is both found and made, though Zahavi stresses that we can only make our-selves if there is a self that we can attribute to ourselves; normative commitments and narrative self-interpretations presuppose a pre-reflective, experiential ownership (Zahavi 2015, 4).

The notion of the self that I propose also attempts to overcome an essentialist and dualist view without embracing a no-self account. In addition, the fundamental self-referentiality (*ipseity*) that Zahavi (2015), in agreement with Sarte, endorses is inherent in my conception of self-formation: self-formation requires a self that can relate to itself. I am also sympathetic to Zahavi's view that the self is not only a social construction but presupposes a certain real selfhood, although I interpret and develop the relation between the two in a different way and direction. The interactionist view of the self that I propose endorses a realist account that attempts to do justice to both the existence of an actual, pre-reflective individual self as well as to the possibility of forming that self in a certain direction: there must be a "pre-reflective form of selfhood" (in Peirce's idiom: Secondness) that precedes deliberate attempts to form itself in a certain direction (Thirdness), made possible by a structural indeterminateness (Firstness); a person can transcend her actual,

individual state by orienting herself toward goals and ideals, and in this way forming a more preferred self.

However, I neither adopt nor reject the idea of a "primitive and minimal" self. This notion has raised questions of *how* primitive and minimal that self really is and has solicited accusations of endorsing a metaphysics that is still too thick (see the debates in Durt et al. 2017). Although this is an important issue, it cannot be dealt with in this study. The main aim of this book is not investigating to what extent the pre-reflective self is or is not interpersonal or a (historical) construction, but developing an interactionist framework that enables understanding whether and how self-formation is possible in an intrusive technological environment. After having developed that framework, one could evaluate whether the idea of a minimal self that justifies a distinction between consciousness and self-consciousness is inevitable or a symptom of a view that needs to be left behind. However, that question I leave for future research.

Part II

Is Self-Formation in a Technological World Possible?

In the first part of this book, an interactionist notion of the "self" has been developed in a critique of essentialist and dualist approaches to the self, which has culminated in a view of the self as "self-formation"; one could say that the "self" has been deconstructed, on the basis of Nietzsche and Peirce's views, as "self-formation." "Self-formation" has been explained in terms of organizing or regulating multiplicity and struggle by virtue of an orientation toward certain goals and ideals.

Now, this basic framework must prove itself fruitful in the context of a world that is saturated with technologies and a repository of ever increasing technological explanations, technological explanations not only for scientific but increasingly also for psychological and philosophical phenomena. The influence of technologies and technological explanations has in the modern Cartesian worldview always been restricted to the realm of the external world. Currently, technologies and technological explanations are increasingly extending their reach, intruding also in the realm of the human self; the self seems to be greatly influenced by its environment, which is increasingly a technological environment, and liable to technological explanations. But if the self is to a great extent shaped by its environment and if this environment is evermore a technological environment that is becoming ubiquitous, what then secures that this self is not the product of this technological environment? And if technologies are increasingly capable of revealing (necessary) laws and mechanisms that explain our actions and decisions, what then is left of the autonomy that in everyday life is ascribed to human agents?

The primary question in this part of the book is then how the proposed interactionist framework can explain *if* it is at all possible to form ourselves in an intrusive technological environment. In general, theories that highlight the dominant influence of external factors on the self tend to strongly constrain its possibilities to form itself, whereas views that attribute unrestricted autonomy to the self and a capacity to detach itself from natural and societal influences often also render self-formation possible. But, how to sufficiently acknowledge the influence of the environment, which is increasingly a technological environment, and, at the same time, account for the possibility of self-formation?

It must be clear that it is not my aim here to provide an overview of the sophisticated positions (hard determinism, compatibilism, hard incompatibilism, libertarianism, two stage models, non-/agent-/event-causal frameworks etc.) and fine-grained arguments that are exchanged in the debate regarding autonomy, free will and responsibility, which is impossible. I will rather briefly touch on a selection of positions pro and contra autonomy and attempt to show that, despite relevant differences, these views have a common blind spot: they assume a problematic categorical distinction between an "inside" and an "outside" realm, which is both a manifestation of an essentialist and dualist view of the self and an obstacle to overcome that view. I will, in both this and the next part of this study, "reveal" this blind spot in contemporary views in philosophy of technology and argue that it is greatly responsible for their failure to provide a framework that sufficiently accounts for the intrusive influence of technologies on the human self. In this part, I will focus on the extended mind theory and the view that brain imaging technologies could provide functional evidence in free will debates, using primarily Peirce's framework as a filleting knife; in the next part, I will engage with other contemporary views in philosophy of technology, using Nietzsche, Peirce and other philosophers as "tuning fork." The engagement with and critique of these approaches will allow me to further develop my interactionist notion of self-formation and to make clear in what sense self-formation in a technological world is possible (which is the topic of this part of the book) and how the self ought to form itself, that is, what "good self-formation" entails (which is the topic of the third part of this book).

In Chapter 3 of this part of the book, I will set the stage by presenting the freedom-determination dichotomy on the basis of Kant's influential deontological ethics, which states that, despite different influences, human agents possess an ineradicable autonomy. Kant's attempt to show that the human self contains an "inner agent" that can detach itself from the outside world and govern its corporeal behavior has prompted various responses. I will briefly engage with contemporary "Kantian" theories that revolve around the idea that the human can only be free and responsible if it possesses a *real self* that secures its *mine-ness*. Parallel to these debates in philosophy, developments in social psychology and neuroscience have flared up the question of whether the self is "master in its own house." Within this context, I will discuss the freedom-determination dichotomy in a conversation with situationists, brain-determinists and brain-non-determinists. In the last part of this chapter, I will argue that, despite many relevant differences, these approaches share that common blind spot: the assumption that "inside" and "outside" could be categorically detached from one another.

Chapter 4 will argue that radically overcoming the inside-outside distinction requires showing that the mind as such, which was for Descartes,

Kant and other essentialists and dualists the mark of the internal, is material and technological; conversely, it will turn out that the material world contains a mental dimension. Overcoming the inside-outside distinction will make it possible to do justice to the dynamic, relational and interdependent character of the self and hence to recognize its "artifactual character," which is a view I will elaborate in a constructive critique of the extended mind thesis.

In Chapter 5, I will attempt to go beyond the inside-outside distinction by challenging the view that the brain is the locus of the self and the organic "inner self" that controls the body and utilizes the material environment. Analyzing and criticizing how brain imaging technologies frame the relation between brain, body and world, and sustain and strengthen the view of the brain as an isolated agent, I will show how from an interactionist view of the self, "self-determination" could be reinterpreted and further specified in terms of "self-formation" and how brain imaging technologies, instead of proving whether free will exist or not exist, could contribute to forming a critical self. Brain imaging technologies allow, as I will propose, critically reflecting on whether our habits are consistent with the longer-term goals or ideals that we pursue: they potentially display to what extent our biological constitution is in par with or poses certain limits or constraints to our personal or social ambitions. This further illustrates how a seemingly internal, biological reality can merge with what is considered the external world.

In the last chapter of this part, I will further sharpen the proposed view of critical self-formation by reflecting on the question "how critical" critical self-formation is; the question is here to what extent both longer-term goals and ideals and cognitive and neurological means to form ourselves are not again determined by certain external factors, namely society and institutions. Self-formation will appear to be not only an individual but also a collective process.

3   The Autonomous Self and
        the Determined Self

3.1 Kantian Notions of Autonomy and Responsibility

Kant's deontological ethics had and still has a great influence on positions and arguments regarding autonomy, free will and responsibility. After reviewing his approach, I will briefly discuss three contemporary views that are influenced by but also struggle with it. These views often endorse a compatibilist take and focus on the question of what kind of self is required to secure autonomy and responsibility. This discussion will make clear that, while Kant's position might be and is very problematic, what is at stake in Kantian adaptations of autonomy, free will and responsibility is far from obsolete. A brief elaboration of relevant social psychologist and neuroscientific approaches to freedom and responsibility in the next section will make clear that a more empirically oriented approach as such is not sufficient to overcome the inside-outside distinction. Sufficiently fathoming of what is at stake here is crucial to make further progress.

Kant's Pure Autonomy

In his *Groundwork of the Metaphysics of Morals*, Kant sets the tone of his ethics with the following famous statement: "It is impossible to think of anything at all in the world, or indeed even beyond it, that could be considered good without limitation except a *good will*" (Kant 1997, 7). This assertion is an allusion to many ethical theorists who preceded Kant and tried to ground morality on, for example, the law of god or a sovereign monarch, a special feeling that we experience when we act morally good, or the perfection of natural qualities. Kant admits that qualities like wit, courage and perseverance can be good in different respects, but these "gifts of nature" may also turn out to be extremely bad if the will that makes use of them is not good (Kant 1997, 7). Only a good will is good without restrictions.

   What Kant has seen sharply is that grounding morality on an externally imposed law or foreign institute – and we will see later on that also our natural inclinations are external to our genuine self – jeopardizes the

autonomy of the will. Because we can only be held accountable for our autonomous actions, acting under orders of a will that is not our own discharges us from being accountable for our deeds. Put yet differently, we can only be held responsible for what is within our control. All that happens in the external, empirical world is, according to Kant, beyond our control (Kant 1997, 8, 28). The only thing that I can control is the decision that I make before I physically act, that is, my intention. That is why Kant can say that only a good will is intrinsically good. The equation of will with decision-making implies, as we will see, that the will for Kant is not something that is detached from reason, but that it rather is a particular modus of reason.

If qualifying an action as good or bad is only justified if the action is the result of an autonomous will, then the question that emerges is: when do we act autonomously? Distinctive for agents with an autonomous will is that they do not simply act, but are able to reflect and decide whether they would want to act in a certain way. In other words, they have the capacity to act deliberately. Acting under the influence of passions or impulses is, according to Kant, a feature of a heteronomous will (Kant 1997, 41, 47f). It is action that is not the result of a deliberate and self-legislated choice, but action that is governed by a foreign, external faculty. We act autonomously only if we act in accordance with a law that is dictated by our own reason.

Acting autonomously and therefore reasonably is thus not the same as doing whatever I please. Reason is for Kant not determined by subjective preferences or particular circumstances, but has rather its own objective laws, which are valid for every rational being. That we, moreover, experience these laws as imperatives that we ought to follow, as constraints, does not imply that we act under the influence of a foreign legislator, but rather indicates that we are only partly rational, that is, we also have a bodily nature, which is not necessarily in accordance with our reason (Kant 1997, 24, 25).

Kant believes that a principle can only be called reasonable if it is valid in all situations and circumstances, regardless of subjective inclinations or desires; reason is the same at all times and for all people. From this, it follows that only categorical, unconditional imperatives (in contrast to hypothetical imperatives) can be considered reasonable in a genuine sense (Kant 1997, 30f). Consequently, reason becomes not only the source for morality but also the criterion for establishing the moral value of an action.

Once we recognize that moral law is universal, we can determine the "content" of the categorical imperative. If we strip away all empirical considerations that are part of a hypothetical and heteronomous point of view, we discover that the content of the moral law is in a certain sense its form. It says that an action is morally good if it is not dependent on arbitrary impulses or preferences, but rather is motivated by a maxim

that we could will in all possible circumstances – "arbitrariness" is, we could say, the opposite of "rationality" (Kant 1997, 30f).

Kant's categorical imperative can be characterized more sharply by discussing shortly his two most known formulations of it. His first formulation is: "I ought never to act except in such a way that I could also will that my maxim should become a universal law" (Kant 1997, 15). At first sight, this seems to be no more than stating in a more weighty way a principle that is used commonly in everyday life, namely "Do not do to others what you do not want done to yourself" or "Treat others as you would like them to treat you" (St. Luke 6:31). However, Kant in a footnote warns us against confusing his categorical imperative with such popular sayings (Kant 1997, 38). Those everyday life maxims usually appeal to the wants and likings of particular persons and are therefore based on self-interest, whereas Kant's imperative wants to radically exclude self-interest from moral consideration and overcome particularity. His categorical imperative does not say that I, for example, may not steal because I would not like it if people would steal from me, but rather indicates that stealing is morally wrong because I cannot will it to be a universal law, irrespective of my personal desires or particular circumstances. Actions ought to be performed exclusively *for the sake* of duty (Kant 1997, 11f).

Kant's second formulation of the categorical imperative is: "So act that you use humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as a means" (Kant 1997, 38). I may not treat other rational beings as mere means to my purposes because in pursuing their goals, they, like myself, never consider themselves as mere means to other purposes. Put differently, rational beings cannot be treated as mere means because they are simultaneously the authors and the subjects of the laws and principles they execute through their will. An ideal world would be a kingdom in which persons never are treated merely as means but always also as ends. Kant's ethics has laid the foundation for the great importance that we ascribe to autonomy today.

The major problem that Kant's ethics has to contend with from the start is that morality for him is only possible *if* we can really determine our own law for ourselves, *if*, in other words, we are really free. In the last chapter of his *Grundlegung* (Kant 1997, 52–66), he shows that the reality of freedom can never be proven by empirical observation because the world that we perceive and understand is governed by the principle that every event is caused by another, prior event. Kant's way out of this determinism is through his introduction of the distinction between an intelligible world and a world of appearances. Kant not only argues that we are members of both worlds, but he also states that the first has a certain primacy over the latter: appearances are no more than the picture that reason develops in making sense of the world. Although we can

never prove the reality of freedom, he concludes in the *Grundlegung*, we necessarily presuppose it, especially when we engage in practical endeavors. If we do not attribute our judgments and actions to our conscious intentions, we cannot understand what we do and what happens around us (Kant 1997, 62f).

Real Self View, Autonomy View and Reason View

Kant's view that a decision or action is *mine* if and only if it can be ascribed to *my* will has been taken up by different contemporary authors and has been embedded in different ways in various contemporary positions. Pauen (2004, 2007), for example, takes as starting point that any action that is said to be free has to meet two minimal conditions: the "absence of compulsion" and "the absence of pure chance." He argues that these criteria positively entail that my action is free if (and only if) it can be attributed to *my* (and not to an enforced or a random) will. From this perspective, "autonomy" or "freedom" can be translated as "self-determination." Additionally, Pauen (2007, 5) points out that it "is completely unclear what it means to act in a self-determined way, as long as it remains to be spelled out what the self is."

Pauen's approach seems to fit within what sometimes is called the Real Self View of free will and moral responsibility, a label coined (and a position rejected) by Susan Wolf (1990). More recently, different versions of this approach, sometimes also calling it the Deep Self View, have been proposed and discussed by T.M. Scanlon (1998), Nomy Arpaly (2003), Angela Smith (2005, 2008), Sarah Buss (2012) and Chandra Sripada (2016), among others. The basic idea in most Real/Deep Self Views is that for an action to be my own, it must flow in some way from my real/deep self: only actions that originate in factors internal to me can be morally attributed to me. What is the correct criterion of internal sourcehood (rational judgments, higher-order desires, cares etc.) is a question that is subject to different and contesting conceptions. Often real/deep self views endorse a compatibilists approach, though what distinguishes them from more classical compatibilists views is that they can account for cases of internal compulsion (such as addiction), which are considered unfree, by making a distinction between two (or more) levels or motivational systems. The real or deep self is then located at the level or in the system that can reflect on and potentially govern the more superficial self (Faraci and Shoemaker 2010, 321). What this often boils down to is a distinction between a "superficial or unreal within," which is in a certain sense external, and a more "genuine or real within."

Susan Wolf introduces the Real Self View by first discussing and contrasting it to the so-called Autonomy View. The Autonomy View requires that to be responsible, the agent must not only control her behavior via her will, but must have "ultimate control": "her will must be determined

by her self, and her self must not, in turn, be determined by anything external to itself" (Wolf 1990, 10). This requirement of "ultimate control" implies that a self can govern her will if and only if it is completely unaffected by external forces: not only the environment and our biology, but also random occurrences or brute inexplicable facts must not affect our will. The Autonomy View requires a prime mover unmoved, whose self can endlessly account for itself and its behavior, which "seems incoherent or, at any rate, logically impossible" (Wolf 1990, 14). Hurley (2003, 39) points out that the Autonomy View is ultimately committed to the regression requirement.

By contrast, the Real Self View does not require ultimate control and accountability, which seems to be an advantage:

> It is required that an agent have a real self, and that she be able to govern her behavior in accordance with it. But it does not matter where her real self comes from, whether it comes from somewhere else or from nowhere at all (Wolf 1990, 35)

On the first sight, this might seem an attractive position, but it is on reflection unacceptable for Wolf, since a fully developed agent could act irresponsible even when she acts in a way that is clearly attributable to her real self (1990, 37). People with certain mental disorders or a deprived childhood illustrate this possibility. Wolf's famous case of JoJo, the favorite son of an evil and sadistic dictator, questions "whether anyone with a childhood such as his could have developed into anything but the twisted and perverse sort of person that he has become" (Wolf 2003, 379–380). People can be wholly responsible for their actions, even though they are not responsible for the real selves from which their actions result, which Wolf finds incoherent (1990, 38–39). Real/Deep Self Views cannot account for situations where the conditions for moral responsibility are met, but a person still cannot be considered free and be taken responsible for her actions. Advocates of the Real Self View ultimately confuse, says Wolf (Wolf 1990, 40–43), causal responsibility with deep moral responsibility. In sum, for the Real Self View, an individual is responsible if and only if she is able to form her actions on the basis of her particular values, which is for her an insufficient condition for responsibility.

Wolf's own view, which she calls the "Reason View," is, according to her, not subject to the flaws of the Autonomy View and the Real Self View. From this view, responsibility requires something more: "an individual is responsible if and only if she is able to form her actions on the basis of her values *and* she is able to form her values on the basis of what is True and Good" (Wolf 1990, 75). The difference between the Real Self View and the Reason View amounts to, as Hurley (2003, 42) has indicated, "the difference between subjective and objective reasons."

From this perspective, it also becomes clearer why Wolf rejects the Autonomy View. Besides the commitment to ultimate, regressive control, the Autonomy View also is committed to the ability-to-do-otherwise requirement (Wolf 1990, 68). Autonomy requires radical freedom, which implies that the autonomous agent who does the right thing could have done otherwise, even holding constant all the possible reasons for acting. For Wolf, this autonomous ability to act irrational is unacceptable. She opts for the Reason View because she believes that praise does not require an alternate possible sequence: the ability to do the right thing for the right reasons "is compatible with the inability to do anything else" (1990, 69). Like the Autonomy View, the Reason View also requires that the responsible agent has the ability to do otherwise, except in those cases in which the agent does the right thing for the right reasons.

3.2 Freedom and Determination in the Context of Social Psychology and Neuroscience

Opponents of the existence of autonomy and freedom often find Kantian ethics too formal, that is, they believe that Kantian ethics unjustly evaluates and insufficiently appreciates the concrete, empirical environment that to a great extent determines our decisions and actions as well as our biological nature that strongly influences our behavior. I limit myself to two currently influential approaches that challenge the idea of an autonomous self: situationists and brain-determinists. I will also briefly discuss some arguments in favor of the existence of free will in the context of neuroscience. Later, in this part of the book, I will hook into free will debates in the context of neuroscience and highlight and discuss the increasing role of brain imaging technologies in attempts to prove or disprove free will.

Situationists

Situationists believe that autonomy and free will are an illusion because our actions are determined by features of the environment of which we are not aware, or of whose influence we are not aware (see Darley and Latane 1968; Baron 1997; Williams and Bargh 2008). For a situationist, external factors rather than internal character and motivations define behavior and personality.

Stanley Milgram's Yale University Study (1960), the Philip Zimbardo's Stanford Prison Experiment (1971) and the bystanders experiments (Latane and Rodin 1969) are often cited as supporting evidence for situationism. In the Milgram experiment (see Milgram 1974), an experimental assistant asked each subject to administer "electric shocks" to another subject-accomplice each time the subject-accomplice made a mistake on a word-learning exercise. After each mistake, the subject was

asked to administer a shock of higher voltage, which resulted in audibly "apparent" increasing distress from the subject-accomplice. Over 60% of the subjects were willing to administer the highest voltage to the other subjects-accomplice. Many replications of Milgram's experiment showed similar outcomes to the original study (see Doliński et al. 2017). The Zimbardo experiment (1971) was carried out in a mock prison setting. Undergraduate volunteers were randomly assigned to be either guards or prisoners for a two-week experiment. After less than a week, the experiment was stopped as the subjects were transformed by the situational conditions, resulting in behavior that included sadism, brutality, lying and extreme stress. In the case of bystander studies, Latane and Rodin (1969) and Latane and Darley (1968, 1970) attempted to show that the presence or absence of others has the most significant contribution to the subjects' willingness to help or not to help a subject-accomplice. Situationists use these experiments to illustrate that it is not internal intentions and motivations, but situational factors that are the most influential determining qualities for behavior, morality and personality.

Citing these and other experiments, Harman (1999, 2002, 2003) argues that there is no such thing as "character": given the right circumstances, all people are capable of behaving viciously. In his 2002 work *Lack of Character*, philosopher John Doris stated that robust character traits are very rare: it is not character traits but situational factors that determine our behavior. Doris (2015), in his latest book, presented a large amount of evidence showing humans often act in unreflective and downright self-ignorant ways. When "the causes of her cognition or behavior would not be recognized by the actor as reasons for that cognition or behavior, were she aware of these causes at the time of performance," Doris argues, "these causes are defeaters" (2015, 64f). Agency is defeated because reason is "overwhelmed" or "bypassed" (52). Along similar lines, Stocker (1979), Brink (1997), Svavarsdóttir (1999) and Zangwill (2003), sometimes labeling themselves as "motivational externalists," argue that also our moral judgments have no motivational efficacy in themselves, and that when they motivate us, the source of motivation lies outside the moral judgment.

Brain-Determinists

In recent years, leading neuroscientists have, partly inspired by older studies, argued that the actions that we ascribe to our (real) selves are in reality caused by unconscious, automatic operating processes in the brain. Since we relate who we are to our mind or certain contents of our mind, and our mind cannot control our brain, what we do is not up to us. Therefore, autonomy and free will is, for the most radical opponents of free will, impossible. It must be clear that critics of free will come in different shades and that attempts to disprove free will come in very

different forms and degrees (Nisbett and Wilson 1977; Libet et al. 1983; Wegner and Wheatley 1999; Hassin et al. 2005; Soon et al. 2008; Lamme 2010; Swaab 2010; Fried et al. 2011; Harris 2012).

Benjamin Libet (see Libet et al. 1983) has become famous for his finger wagging experiment. This experiment shows that the moment at which respondents indicate awareness of their decision to move their finger is slightly less than half a second after the so-called "readiness potential" is observed in an EEG. On the basis of this observation, the (contentious) conclusion is drawn that before respondents are aware of their intention, the brain has already made the decision to move the finger, which seems to undermine the concept of free will.

Following Libet's observations, more attempts have been made to conduct increasingly sophisticated experiments and confirm that not our conscious selves but our brains determine our actions. John-Dylan Haynes' (in Soon et al. 2008) MRI set-up provided improved spatial resolution and could be used to survey the whole brain, whereas Libet's EEG technique had only recorded a limited area of brain activity. In this study, participants could freely decide if they wanted to press a button with their left or right hand, but were also asked to remember exactly when they felt they had made up their mind, which was registered by a computer. Haynes and his team claimed to be able to predict with 60% accuracy which button subjects would choose six seconds before they were consciously aware of their decision.

Fried and his team (2011) studied individuals with electrodes implanted in their brains as part of a surgical procedure to treat epilepsy. This enabled them to record the activity in single neurons using microelectrode systems, which provided a more precise picture of brain activity than both EEG and fMRI. In their study, they showed that there was activity in a small group of neurons in particular brain areas before the subject made a conscious decision to press a button. With about 700 milliseconds to go, the researchers assert that they could predict which decision would follow with 80% accuracy.

Daniel Wegner's research combines situationism and brain-determinism. It can be best framed by describing an experiment that was conducted in the 1970s. In a study by Richard Nisbett and Timothy Wilson (1977), people in a department store were asked why they had chosen a particular type of stockings. The respondents provided various seemingly well-thought through reasons for their preference: a smoother fabric, nicer color and so forth. But what does the experiment show? The stockings were exactly the same. The actual reason why respondents chose a certain type of stockings seems to be dependent on their place: due to a particular brain mechanism, stockings that are lying to the right on the shelf were chosen more often than stockings that are lying to the left. Inventing reasons for our actions after an event has taken place is called "confabulation," and is confirmed by many experiments.

Wegner (Wegner and Wheatley, 1999) regenerates this kind of experiments, but also investigates the conditions under which people experience the feeling of free will. If our intention precedes a particular action (the "priority principle"), if there seems to be a substantive relation between our intention and the action (the "consistency principle") and if our conscious intention is the only apparent cause of the putative effect (the "exclusivity principle"), we (often wrongly) tend to think that the act is caused by our conscious intention. Wegner's experiments try to show that the sense of free will is a mental construct and not a direct representation of the real causes of a particular act.

Wegner's results are part of a larger set of experiments that has become known as the "New Unconscious" (Hassin et al. 2005). This approach attempts to show that the vast majority of the actions that we ascribe to our conscious intentions, including those conscious intentions themselves, are in reality caused by unconscious, automatically operating processes in the brains or elsewhere in the body.

Brain-Non-Determinists

Situationists and brain-determinists believe that autonomy is an illusion because our actions and personality are not determined by our consciousness or mind, but by features of respectively the environment and our brain, which we cannot control. Various positions have been developed to show that autonomy and free will are not or are not entirely an illusion. These positions attempt to illustrate that conscious intentions and decisions not only can have causal efficacy, but that they also in one way or another can intervene in and control both brain inclinations and environmental influences, and are therefore not necessarily subject to unconscious and/or external forces (Mele 2006; Libet 2011; Baumeister et al. 2011; Slors 2012, 2015). I will again limit myself to a couple of examples (see Klemm 2010 for a more extensive overview) that illustrate how brain imaging technologies are used to prove the existence of free will.

Some advocates of free will refer to another experiment of Libet (2011), in which respondents are given a second assignment: when they are aware of their intention to move their finger, they should try to go against that inclination. On the EEG, one can see again that the readiness potential precedes the conscious intention by about half a second. However, just before the expected action, we see that the readiness potential discontinues. Libet and some contemporary advocates of free will argue that it might be true that we mistakenly believe that our actions are caused by conscious intentions, but that does not exclude the ability to intervene in an impulse, and consciously veto and stop the action that our brain has unconsciously prepared. We do not have a "free will," but a "free won't."

Trevena and Miller (2010) challenge the assumption that evolving brain activity prior to conscious awareness of an intention to act is unambiguously associated with preparation for movement. According to these authors, there is no evidence that electrophysiological signs before "a decision to move" are stronger then electrophysiological signs before "a decision not to move," which would indicate that these signs are not specific to preparation of movement.

Koechlin and Hyafil (2007) have pointed out that the brain regions most frequently studied in the context of free will debates, namely the (pre-)supplementary motor area and the anterior cingulate motor areas of the brain, may be only involved in the later stages of motor planning. However, this does not exclude the possibility that other parts of the brain, which are not yet taken into account, are responsible for decision-making and exerting will.

According to Mele (2006), the readiness potential depicted by brain imaging technologies does not necessarily exclude the influence of conscious decisions on behavior. Forming certain habits could initially be the result of a conscious choice. Once habits are initiated, they may be executed with little or no conscious involvement. This explains why an intention to act in a certain way can arise without being actively formed from a decision process, but does not exclude the possible influences of consciousness on our decisions and actions. This also does not exclude the possibility that, as Libet suggested, a habit may be vetoed consciously.

3.3 The Inside-Outside Distinction as a Blind Spot in Free Will Debates

With respect to Kant, the inside-outside distinction could be hardly called a "blind spot," since it is at the heart of his ethics. Kant strongly adopts, in his own ingenious way, Descartes' distinction between a thinking substance (*res cogitans*) and a material substance (*res extensa*), highlighting the decisive role of an impartial, autonomous and rational self in moral deliberation. Descartes argued, as we have seen, that in methodical self-contemplation, the self can clearly and distinctly fathom that besides being a body, it is first of all a mind that can exist completely independent of the material world, including its body. Since I have this privileged access to my mind, but not to my body and the world, I can primarily identify myself with my conscious, rational thought. For both Descartes and Kant, the distinction between mind and world/body is primarily the outcome of an epistemological investigation. However, Kant further develops it within a moral context: his radical distinction between a rational, inner world and an empirical, external world is a necessary condition for securing the possibility of moral autonomy and responsibility.

The idea that the possibility of autonomy and responsibility requires a kind of independent inner self continues to appeal to a range of philosophers and philosophical approaches. Pauen engages with and tries to debunk different compatibilist and incompatibilist approaches, and attempts to show that determination and free will do not contradict one another; freedom is, as we have seen, translated in terms of self-determination (2004, 2007). The absence of pure chance entails, says Pauen, "authorship," whereas absence of compulsion involves "autonomy." Taken together, these conditions are, he thinks, not only necessary but also sufficient (2007, 458). For Pauen, an action is free if (and only if) it springs from the inner desires, wants and beliefs of an agent: if action A is determined by X's personal preferences, then A is self-determined and hence freely willed (2004, 163). The task of neuroscience is, he believes, to establish if that is really the case: "neuroscientists would have to ascertain whether individual actions really are determined by central personality traits or whether they are reliant on external factors" (Pauen 2004, 47).

Although Pauen's view reverberates Kant's idea of an inner self that determines human actions and decisions and hence shows traces of the inside-outside distinction, it must be clear that he does not want to adopt a thick Kantian dualism of a noumenal world where an inner self resides and a phenomenal world of sense experience. However, the question is whether his notion of "self-determination," which, as I will discuss later on, I do not reject as such, can be preserved and used to develop a non-essentialist and non-dualist view of the self with sufficient explanatory power. With his great emphasis on autonomy, authorship and personal preferences, Pauen takes a very individualistic view of the self, which seems to make it difficult to go beyond thinking in terms of the opposition between the individual and the external world. However, I fully agree with Pauen's assertion that it "is completely unclear what it means to act in a self-determined way, as long as it remains to be spelled out what the self is" (2007, 5). This study attempts to contribute to that endeavor, stressing that fundamentally reflecting on and challenging the inside-outside distinction is pivotal for developing a non-essentialist and non-dualist self.

The Autonomy View, The Real/Deep Self View and the Reason View also engage in different ways with the Kantian notion of autonomy as self-legislation. They have in common with Kant's view that a decision or action is *mine* if and only if it can be ascribed to *my* will. For Kant, this condition was secured by his strong epistemological distinction between an inside noumenal world and an outside phenomenal world: by identifying itself with its unaffected, pure and rational inside world, the self could have ultimate control over its will and overcome the gap between subjective and objective reasons, which was for Kant the most important ethical imperative. Irrespective of external and biological influences and

contingencies throughout its past and present, the self could this way be completely taken accountable for its decisions and actions.

The Autonomy View, The Real/Deep Self View and the Reason View recognize that both our surroundings and/or our biology influence the way we determine ourselves and consequently constrain our autonomy and freedom. However, they respond to those influences in different ways, articulating which aspects or principles with regard to autonomy and responsibility they find nonnegotiable. Their responses also reveal what is at stake in the debate about autonomy and responsibility.

For the Autonomy View, it is pivotal that decisions or actions attributed to me are all the way *mine*, which requires, besides the "absence of compulsion" and the "absence circumstantial chance" (see Pauen above), also the ability to even go against what is considered reasonable. The Reason View overlaps with the Autonomy View, except that it rejects the commitment to the ability to go against reason because that would ultimately amount to a kind of willed irrational behavior. The Real Self View overlaps with the Autonomy View, except that it does not require that the self is endlessly unaffected by external forces. Moreover, it accords with the Reason View, except that it does require decisions and actions being based on subjective reasons that also necessarily coincide with objective reasons.

The important question here is to what extent these views can uphold these aspects or principles and render them nonnegotiable in our present world without bringing in the categorical inside-outside distinction through the backdoor. The commitment of the Autonomy View to ultimate control and tamper-proof *mine-ness* of decisions and actions attributed to me seems to only be able to avert regression and incoherency if it embraces a prime mover that is completely independent of the outside world. This way, the Autonomy View disregards the fact that our agency and autonomy cannot be placed outside particular environments but are shaped within them.

The commitment of the Real Self View to the existence of a real self that is the basis for and the governing agency of its decisions and actions, irrespective of where this real self comes from, seems to allow for a self that has at least partly formed itself through interaction with its social and material environment. However, there still seems to be a strong distinction between a self that has been formed in the past and the current external environment that it can detach itself from. This way, a horizontal or historical inside-outside distinction is introduced, disregarding that what we consider our real self cannot be isolated from current influences of its environment. In addition, it remains unclear how a particular self can improve its real self and develop itself into a more (socially) admirable self.

Finally, the commitment of the Reason View to an objective Good and True disregards the fact that the qualification of a decision as True,

an action as Good and conduct as reasonable is always done within a particular, historical context, and a self is formed and continuously influenced by divergent social and material factors. Though, it must be noted that recognizing this does not entail that the framework I will propose includes no guidelines that enable making a distinction between a more and less well-formed self. Hence, all three positions contain, on the one hand, remnants of the Cartesian-Kantian inside-outside dualism, and, on the other hand, highlight different elements with regard to autonomy, reason and responsibility that they find negotiable or not negotiable.

As indicated earlier, this criticism does not entail that I completely reject the idea of freedom as self-determination. I will rather argue that subscribing to the intuition that understanding autonomy as self-determination, as well as to the idea that the orientation toward certain values and ideals make up our "real self," is fruitful. However, developing this alternative requires overcoming the inside-outside distinction more thoroughly and persistently; shaking off this distinction will shed a different light on both the notions of "(real) self" and "determination" in "self-determination" as well as on the role that technology plays in self-formation.

Situationists reject the possibility of autonomy that is either defended or assumed in Kantian views, including the view of a real self that can determine its conduct. They believe that external factors determine our behavior and personality, including what we consider our "real" and "deepest" self. However, endorsing this determinist view does not overcome the inside-outside distinction but only preserves it. Rather than defending the idea that an independent, real inside is able to control the outside, the determinist proposes that the outside controls the inside or renders it into an epiphenomenon. Reverting the importance or power dynamic of inside and outside does not overcome the essential relationship of the self to the world. The idea of the self as a passive "agent" that is dependent on contingent circumstances is very much dependent on the inside-outside distinction. In addition, situationism does not or does not want to provide a balanced account of, on the one hand, recognizing that we are greatly shaped by our environment, and, on the other hand, showing how this does not render self-formation impossible, which seems to go against our basic intuition that, at least potentially, we possess some kind of control with regard to our conduct.

This objection is to a certain extent also applicable to brain-determinism. Brain-determinists who consider the brain, instead of consciousness/mind, as the part within the human organism that is really in charge – as the "real self" that instigates cognitive processes, moves the body and manipulates the external world – do not overcome the inside-outside distinction either. Instead of the mind, it is now the brain that is envisioned as a kind of *causa sui* that exerts control over our

behavior. Although Descartes would never have situated cognition in or reduced it to brain processes, since brain and body are part of the *res extensa*, the division between "internal" and "external" that breaks along the line of an independent agent that is in control (brain or mind) is, as Rouse (1996) has pointed out, a Cartesian legacy.

This Cartesian-Kantian view not only represents a dualist framework, but also evokes particular notions of "control" and "freedom," which inform (and, as I will show, also misinform) the debates about autonomy, freedom and responsibility. Although advocates and opponents of autonomy and free will do not agree on the question of whether and to what extent consciousness/mind, the brain or the environment can be attributed causal efficacy and control, they often share the views that a distinction should be made between opposing realms and that freedom can only be attributed to an agent that is not fundamentally affected or determined by something else, which is a precondition to be able to cause and control actions, make decisions and utilize the outside world.

In the next chapter, I will, building on the intuitions and concerns highlighted in the discussed positions regarding autonomy and free will, attempt to go beyond the inside-outside distinction more radically. I will do that by challenging the elementary idea that mind is an inner realm detached from our increasingly technological, external environment, which will have bearing on the related assumption that the self primarily identifies itself with this inner, mental realm. From the proposed interactionist perspective, the mind, which was for Descartes, Kant and other essentialists and dualists the mark of the internal, will prove to be "external," whereas the material, which was for them the mark of the external, will turn out to be intelligible. Because in our current society the material environment has become increasingly technological, I will elaborate in what sense the mind has become technological and the technological has become intelligible. In a constructive criticism of the "Extended Mind Thesis," I will develop what I will call an "Artifactual Mind Thesis" that not only argues that the technological environment is increasingly becoming part of the mind and body of the self, but also that a further radicalization of this view debunks the very idea of an inner self that controls and manipulates an external environment.

4   The Artifactual Mind

The (real) self as an "inner agent" that can detach itself from the outside world and govern its corporeal behavior is one of the most dominant and, at the same time, one of the most challenged ideas in Western philosophy. The idea of independent "within" has been criticized on various fronts, especially by so-called externalists. From their viewpoint, our mental contents, including moral judgments, desires and cares, are influenced by external factors. However, "externalism" is not a unified philosophical "school" but rather represents diverse views that challenge in different ways "internalist approaches." For example, situationism could be seen as an externalist theory, but externalists are not necessarily situationists.

Already in the 1940s, Ryle (1949) argued that both self-knowledge and knowledge of other people are achieved by drawing inductive conclusions on the basis of externally observed behavior. Other authors have argued that all knowledge of mental states, whether they are our own or other people's, is dependent on sensory information that requires interpretation (Carruthers 2011). As we saw earlier, "motivational externalists" attempt to show that our moral judgments have no motivational efficacy in themselves, and that when they motivate us, the source of motivation lies outside the moral judgment (Stocker 1979; Brink 1997; Svavarsdóttir 1999, Zangwill 2003). "Extended Mind theorists" go one step further than the mentioned positions and argue that external objects, and especially technologies, can play such a significant role in aiding cognitive processes that they can count as part of the machinery of our thought and our "self" as such (Clark 2003, 121; see also Clark and Chalmers 1998 and Clark 2008). This view seems to be a good candidate for more radically overcoming the inside-outside dualism in a technological era: not only does it attempt to integrate the material in the mental, but it also recognizes the increasing role that technology plays in realizing cognitive processes.

In their seminal essay "The Extended Mind" (1998), Clark and Chalmers challenge the view that the demarcation of skin and skull determines the boundaries of cognition. They object, in other words, to the idea that what is outside the body is necessarily outside the mind. The separation

of mind, body and environment is, according to them, not evident but rather an assumption that requires justification. We cannot simply point to the skin and skull boundary as justification for the boundaries of cognition or the self, Clark and Chalmers say, since the legitimacy of that boundary is precisely what is at issue (see also Clark 2003, 2008).

In this chapter, I want to show that, although the Extended Mind Thesis (henceforth: EMT) is a very interesting, fruitful and bold attempt to demonstrate how our cognitive system can contain external objects and technical artifacts, it is still parasitical on the "inside–outside" dualism like the more conventional internalist views and more moderate externalist views of cognition, even though this is the very Cartesian worldview that EMT wants to overcome. Although cognition can be, according to EMT, extended by technical artifacts, Clark and many of his friends, including the so-called "second-wave EM theorists," preserve, as I will show, an inner-outer dualism by ascribing to cognition an original starting point in an internal biological core, an inside that utilizes the outside world in order to fulfill certain cognitive tasks that it has set for itself.

Inspired by Charles S. Peirce's view of mind and his semiotics, I will propose a perspective that offers a more successful way out of this thinking in terms of "inside" and "outside." Peirce's view of mind and his reinterpretation of it in terms of signs is a further specification of his category of Thirdness, which I have discussed in part one of the book. In that respect we also have seen that for Peirce, thinking only apprehends and does not create thought; thought is not an inner, subjective faculty but can have an objective reality. From a Peircean perspective, mind is, as I will show, not extended by objects and artifacts, but rather unfolds through and is shaped by them. In this respect, I introduce the term "artifactual mind" and what I call the Artifactual Mind Thesis (henceforth, AMT). Acknowledging that our thinking has an artifactual character means recognizing that external objects and technical artifacts, rather than being utilized by an inside world, have shaped and are continuously shaping the very fabric of our thinking, of what we take to be our "inside world." And there is more: not only are thoughts exosomatically embodied, but the specific physical characteristics of artifacts also can activate new modes of thinking.

The primary goal of this chapter is to explain how the very *cogito* that Descartes rendered independent and pure is shaped by the objects and artifacts that surround it. Especially in our current era, technology is recognized as a major force that not only is increasingly shaping our world, but also is shaping our mind and self. In addition, I will indicate how awareness of the artifactual dimension of mind enables a critical attitude toward reductionist claims that ascribe to thinking certain original and irreducible features. I will show how AMT addresses normative questions regarding the profound influence of artifacts and technologies

on what we presume to be our deepest "inside," which will pave the road for reflecting on how self-formation is possible and what good self-formation could entail.

In the following sections, I will first discuss in more detail Clark's EMT and his parity principle. By debunking the notion of "introspection," I will then explain why I believe Clark's EMT as well as "second-wave" EM theorists still presuppose a fallacious distinction between "inside" and "outside." Next, inspired by Peirce's view of mind and his semiotics, I will elaborate how this "inside–outside" fallacy could be overcome by an artifactual account of mind. In the concluding section of this chapter, I will indicate how AMT can make us more aware of the normative impact of technology on our mind and self and contribute to the development of the notion of critical self-formation.

4.1 The Extended Mind Thesis and the Parity Principle

EMT should not be confused with the more modest view that external features of the environment can have a causal influence on cognitive processing in the brain. Clark and Chalmers claim that the human organism can be "linked with an external entity in a two-way interaction, creating a coupled system that can be seen as a cognitive system in its own right," which implies that "[c]ognitive processes ain't (all) in the head" (Clark and Chalmers 1998). Certain objects in the external environment are utilized by the mind in such a way that the objects can be seen as extensions of the mind itself. External tools can, according to Clark, become transparent in use, meaning that our intentions can "flow" in such a way through technical tools that we feel as if they are part of us (Clark 2003, 123). Clark not only holds that cognition is not all in the brain, but he also asserts that we have to revise our view of the individual as such: our selves do not coincide with our bodies but can be expanded to include nonbiological parts and circuits (Clark and Chalmers 1998).

To explain when a coupled system can be conceived as a cognitive system, Clark and Chalmers (1998) formulate a so-called "parity principle":

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

This parity principle has prompted different responses. Some critics of EMT (Adams and Aizawa 2001; Rupert 2004; Dartnall 2005) argue that the parity principle indicates a similarity relation between external and internal processes: if external processes are sufficiently similar to internal ones, then they can be considered as cognitive. It is easy to show, these critics claim, that external processes are not only very different,

but are also not susceptible to the same explanations as internal cognitive processes. Law-like regularities such as "primacy effects, recency effects, chunking effects and others" apply, according to Adams and Aizawa (2001, 61), only to processes found in the brain. "Primacy effects" and "recency effects" refer to the mechanism that given a list of items to remember, the first items (primacy effects) and the last items (recency effects) are remembered better than those things in the middle. "Chunking" is a phenomenon whereby individuals group items when performing a memory task, for example, by linking them to something that they easily can recall. When retrieving information via extensions such as notebooks or PDA's, these characteristics are not displayed, which indicates that external processes are very different from internal processes. External processes should, therefore, not be considered as cognitive and the boundaries of cognition could be drawn at the skin and skull of a person. EMT's coupling principle oversteps, according to Adams and Aizawa, the "mark of the cognitive," that is, the fact that cognitive content is intrinsic and not based on any sort of social convention (2010, 73).

Friends of EMT (Menary 2006, 2009; Hurley 2010; Sutton 2009; Sutton et al. 2010) as well as Clark (2010) himself, argue that the parity principle does not say that an external process is cognitive if it is sufficiently similar to internal processes. The point of coupling is not to show that external objects have the same properties or as such are cognitive, but rather that objects can become part of a larger cognitive system, such as a human agent. Pivotal for a process to be considered as cognitive is not its location or properties but whether it fulfills a certain cognitive function.

4.2 Challenging the Notion of Introspection and Blurring the Criteria of What it Means to Be "Internal"

The opponents of EMT as well as many other advocates of modern (neuro)biological approaches, clearly uphold a distinction between an inside world of cognition and an outside world of material objects, but often claim that this "inside world" is physically realized or constituted. The criticism addressed to brain-determinists also applies, as indicated earlier, to many opponents of EMT: it must be clear that Descartes did not situate cognition in or reduce it to physical processes, since the body is also part of the *res extensa*; however, the division between "internal" and "external" that breaks along the line of the knowing subject is a Cartesian legacy (Rouse 1996). Moreover, this Cartesian inheritance cannot only be recognized in the positions of opponents of EMT but also the advocates of EMT have, I believe, not sufficiently succeeded in escaping it. To demonstrate this and make preparations for my own

thesis, I will review how the distinction between "inside" and "outside" has been justified and why it is problematic. I will also explicate why in the shift from a Cartesian dualism to a (neuro)biological framework, which is endorsed by most internalists as well as externalists, a problematic "inside–outside" distinction remains.

I have elaborated in Chapter 2 how the distinction between *res cogitans* and *res extensa* is strongly based on the assumption that we have a direct and privileged access to the contents of our thoughts, which we lack toward the external world of material things. Since introspective beliefs about our mind are radically different from our beliefs about the outside world and our mind can be clearly and distinctively grasped independently of anything else, our mind, Descartes claims, must be a completely different substance than the outside world, including our body. This would prove his dualism true (Descartes 1641/1985).

The debate about introspection, especially in analytic philosophy, is complex and the differences in opinion are great. Providing a full account of the different perspectives is in this context impossible, not least because this issue is intertwined with questions concerning self-consciousness and the qualitative aspect of our subjective experiences. Offering a complete outline is clearly not my intention. I will instead briefly discuss a couple of relevant positions that problematize the notion of introspection in order to show which necessary steps have to be taken to radically challenge the "inside–outside" dualism.

I will begin with a characterization of the term "introspection" in the context of contemporary discussions. "Introspection" usually refers to special methods or means that enable us to obtain knowledge of our own mental states. Different features are, as we have seen in Chapter 2, ascribed to this faculty of introspection (see Macdonald 2007): knowledge of our own mental states seems to be authoritative, in the sense that if we think we are in a particular mental state that cannot be challenged. Knowledge of our own mental states seems also to be privileged, meaning that we know the contents of our own minds always better than we know the contents of the minds of other people. Another important feature that is related to introspection is immediacy. This notion implies that introspective beliefs, as opposed to perceptual beliefs, are non-inferential and nonevidence based.

The special, authoritative, privileged and immediate access to the contents of my own mind would, according to some authors, ensure infallibility, indubitability and/or transparency (see Gordon 2007). Beliefs regarding (some of) my own mental states would be, for example, immune from error because they are not inferred from any other beliefs or based on any other evidence (Russell 1910; Lewis 1946; Burge 1988).

Various philosophers from different perspectives have criticized both the features of introspection and the conclusions that are drawn from recognizing them. Some authors have attacked the alleged epistemic

specialness of our self-knowledge by showing that our introspective judgments can be distorted by different confusing factors. Churchland (1981), for example, argues against the infallibility thesis by showing that we frequently make mistakes in our introspective judgments because of (false) expectation, method of presentation and memory effects. I will briefly discuss a few other positions that, in their critique of the idea of authoritative, privileged and immediate self-knowledge, focus on how mental contents are determined by environmental factors.

Gilbert Ryle, who has strongly criticized the Cartesian idea of the mind as an independent entity, argues that introspective knowledge consists of applying words that we have derived from observations of behavior of others to our own conduct: achieving self-knowledge requires, therefore, external cues. Self-knowledge and knowledge of other people are achieved in the same way, namely by drawing inductive conclusions on the basis of observed behavior. Dretske (1996) extended the externalist idea to phenomenal content, arguing that the phenomenal content of experiences could depend on external factors, though the experiences themselves remain in the head. Ryle's critique of introspection results in rejection of privileged access (Ryle 1949). According to Dennett's theory of consciousness (1991, 2003), introspective expressions are the primary data gathered from research into consciousness and not what these expressions seem to refer to. There is no difference between consciousness as such and how we think about consciousness. Even first-person "reports" of conscious episodes are gathered "from the outside." From Dennett's heterophenomenology, there is no inner perspective from which we can compare self-conscious experiences with our judgments about them. Along similar lines, Carruthers (2011) argues that the idea of having authoritative and privileged access to our own mental states is untenable. Against this idea, he poses his "Interpretive Sensory Access" theory of self-knowledge: all knowledge of propositional attitudes, whether those thoughts are our own or other people's, is dependent on sensory information that requires interpretation. Knowledge of our own thoughts is not different in kind from knowledge of the thoughts of other people.

Although Wittgenstein – the last author I want to discuss a little more extensively in this context – should be chronologically mentioned before Ryle, he represents perhaps the most relevant position regarding introspection for our thesis. Hutchinson and Read (2005) show in a fruitful reading of §1 of Philosophical Investigations – in the context of a discussion of Christopher Nolan's movie *Memento* – that positing an inner realm does not provide the necessary conditions for understanding. A conception of mind as an inner realm is precisely that which is being represented externally in *Memento*. The film dissolves our appeal to an inner realm by showing how such an appeal does not give us sufficient insight into what is going on; nothing is gained from the move to an inner realm, just as we experience in Wittgenstein's therapeutic exercise

in §1 (84). This view is also in line with Gordon Baker's (1998) thera-
peutic reading of the "private language argument." Wittgenstein doubts
if it is possible to develop in isolation a language that refers to one's
own sensations (see Wittgenstein 2001, §244–§271). Baker attempts to
show that the notion of a private language should not be read as intel-
ligible but false, but rather as nonsense masquerading as an important
possibility. According to Conant, the point of the exercise is not even
to get us to see that there is something determinate to imagine, which
we are then supposed to see as a sort of thing that cannot be, but rather
"to get us to see that there is nothing for us to mean by the locution
"private language" that corresponds to what we, under the pressure of
certain philosophical perplexities, want to mean by it" (Conant 2004,
187). This also indicates that in contrast to Ryle, Wittgenstein's princi-
pal activity was not the clarification or correction of category mistakes.
In conceiving thought as an inner mental activity, we already go astray,
and in our attempt to try to think it all the way through, we ultimately
discover that thinking and understanding are something different than
we thought, namely: "To understand, having understanding, is being
able to fully participate in the world with others" (Hutchinson and Read
2005, 86). Without certain rules or language games, which regulate so-
cial practices, our own thoughts would neither arise nor have meaning.

If both first-person "reports" of conscious episodes and the knowl-
edge of the thoughts of other people are dependent on external cues
(Ryle), "the outside" (Dennett) sensory information (Carruthers) or
rules/language games that regulate social practices (Wittgenstein), then
the very distinction between "internal" and "external" becomes opaque
or the whole notion of "inner" dissolves on us (see also Kirchhoff 2012,
292). And if what is internal to a cognitive system cannot be secured by
referring to a special, authoritative, privileged or immediate access, then
also the individualist assumption of cognitive agency becomes question-
able (see also Hutchins 2011).

Since the very notion of "inside" is justified by the idea of introspec-
tion, challenging this idea will also have bearing on the view of the brain
as a separate inside world. Taking the brain, instead of as a Cartesian
spiritual mind, as some kind of inner homunculus (or a multitude of ho-
munculi) that instigates and regulates cognitive processes is difficult to
justify, not only because the relation between brain processes and mental
states is still a confusing enigma for us (this is the so-called "mind-brain
explanatory gap"), but also because ascribing autonomous agency to
the brain requires clear criteria that ensure the brain (together with its
physically realized mental states) is a kind of isolated "inside" that can
be considered the locus of our self. The lack of clear criteria to estab-
lish what it means to be "internal" to a cognitive system also greatly
weakens the assumption that an individual's brain is an inner initiator
of cognition, a kind of *causa sui*, that is, independent from and should

be categorically detached from external objects and artifacts. In the next chapter, I will return to this theme in the context of usage of brain imaging technologies and show how in that context this lack of clear criteria is ignored and the assumption that the brain could be depicted as an independent agent is sustained and strengthened.

4.3 EMT, Second-Wave EM Theorists and the "Inside–Outside" Distinction

EMT and the "Inside–Outside" Distinction

Internalist opponents of EMT not only assume that the mind is physically realized, but also believe that in principle, brain processes can sufficiently explain mental processes. Externalists claim that external observations are needed to know what we think, and thus attribute a "broad content" to mental states. Although in these views the mental contents of a subject and his/her access to these mental contents are dependent on aspects of the environment, those aspects are still, many externalists believe, situated outside his/her cognitive processes. Dretske's "phenomenal externalism" is a clear example of this type of externalism (1995; see also 1999). EMT defends an active externalism and states that technical artifacts can become a genuine part of a cognitive system.

If there are, as we have seen above, no clear criteria to substantiate what it means to be "internal" to a cognitive system, then also granting an inside status to mental processes becomes highly dubious. This argument seems, on first sight, only to pose problems for internalists and modest externalists. I will, however, try to show that EMT has not sufficiently overcome the inside–outside dualism, which indicates that it also does not recognize the bearings of challenging the notion of introspection for the inside–outside dualism.

First of all, Clark's parity principle models cognition as something that goes on in the head. Although the outside world can take part in cognitive processes, external processes, in order to qualify as cognitive, have to correspond to or image mental processes that would normally take place in the brain. This condition is strengthened by the additional criteria that Clark and Chalmers formulate. Processes in the external world can only be included as part of an individual's cognitive process, if: (1) the resource is reliably available, (2) the retrieved information is automatically endorsed and (3) the information contained in the resource is easily accessible (Clark 2008, 79; see also Gallagher and Crisafi 2009, 46). These criteria are reminiscent of the features that are traditionally ascribed to introspection: infallibility/indubitability, transparency, immediacy. Although cognition can be extended by an external object or process, this external object or process can only be included as part of an individual's cognitive process if it mirrors "unextended" cognition

(more about this notion later on), that is, if the external resource or retrieved information is infallible/indubitable, transparent and/or immediate. This way the parity principle does not challenge, but rather asserts the inside–outside dualism.

Clark further prioritizes the inside brain world by taking what goes on in the individual's brain as the genuine cognitive agent. The idea that cognition arises from an inside world of brain processes is explicitly stated in Clark's 2008 book:

> in rejecting the vision of human cognitive processing as organism bound, we should not feel forced to deny that it is (in most, perhaps all, real-world cases) organism centered. (…) it is the biological human organism that spins, selects, or maintains the webs of cognitive scaffolding that participate in the extended machinery of its own thought and reason (Clark 2008, 123); the organism (and within the organism, the brain/CNS) remains the core and currently the most active element. Cognition is organism centered even when it is not organism bound (Clark 2008, 139); Concerning the process of recruitment, it is indeed the biological brain (or perhaps some of its subsystems) that is in the driver's seat (2008, 122).

Cognition has an original starting point in an internal biological core that then can employ external objects in order to realize certain cognitive tasks. Clark indeed partly decentralizes the cognitive agent by allowing external objects and processes to become part of the cognitive machinery. However, his individualist assumption of cognitive agency and admiration of the exceptional role of the brain in cognitive and physical activities largely prevent him from overcoming the "biochauvinistic prejudice" that he has set for himself to overcome (Clark 2008, 77). Clark prioritizes the biological brain as the center of coordination and control. He isolates its activities from social dynamics and architectures, and identifies it as the spider that "spins, selects, or maintains the webs of cognitive scaffolding." Clark even acknowledges that "conscious mental states might well turn out to supervene only on local processes inside the head" (Clark 2008, 79), which brings him close to the internalist perspective that he criticizes.

Clark's prioritizing of an "internal biological realm" is already expressed in the very notion of "extended mind," which implies that brain processes are a kind of *causa sui* that utilizes, structures and manipulates external artifacts in order to fulfill certain cognitive tasks that it has set for itself (also Kiran and Verbeek 2010 pose this critique). The notion "extended mind" indicates a movement from "inside" to "outside." "Extended mind" presupposes that there is also something like original or primary "unextended mind." The content of brain processes is granted an original, non-inferential, unextended status, an

assumption that is parasitical on the idea that we have the capacity of introspection, that is, cognition can be localized in an isolated inside sphere that we can access immediately and that is not fundamentally affected by external influences. This isolated inside sphere is further secured by assuming that the relation between biological brain and world is in essence unidirectional. The brain is not substantially affected and altered by external material influences (see Clark 1997, 198).

However, there are various studies that indicate that with regard to brain and world, there can also be influence in the opposite direction: sociocultural practices can reshape certain cortical areas of the brain or transform the brain's representational capacities (see Näätänen et al. 1997; Dehaene et al. 1999; Wheeler 2004). These empirical results confirm that the idea of an isolated, "initial," non-derived inside realm is difficult to uphold.

In short, EMT commits the "inside–outside" fallacy by unjustifiably presuming that our brains are a kind of *causa sui*, which grants them a completely different status than outside processes. This presumption also prompts the view that an individual's brain is an inner initiator of cognition and, therefore, can be considered as the locus of our self. Since we lack clear criteria for the establishment of what it means to be "internal" to a cognitive system, the very distinction between "internal" and "external" becomes opaque. The view of the brain as an isolated, inner initiator of cognition is also being disputed by empirical research that shows that sociocultural influences can alter certain areas of the brain and the way it functions, which seems to prove the challenges to introspection right (I will further nuance this inference in the next section). By upholding the idea of a separate inside brain world, Clark does not overcome but rather rehabilitates and modernizes the Cartesian mindset (see also Zahavi 2008).

Second-wave EM Theorists and the "Inside–Outside" Distinction

Although some so-called "second-wave EM theorists" have defended a more nuanced interpretation of, in particular, Clark's parity principle, often an elementary distinction between inside and outside is maintained in their theories. Many of them emphasize that the intention of the parity principle is not to eliminate the distinction between the characteristics of internal and external processes, but to signal that internal and external processes can complement each other in the fulfillment of a certain cognitive task. Under certain conditions, both can become part of the same cognitive process.

Menary uses the term "cognitive integration" for his view, which, according to him, expresses that "external vehicles and processes are integrated into a whole" (Menary 2006, 329). He problematizes the

notion of "intrinsic/non-derived" cognitive content (Menary 2006, 334–336), and stresses (referring to Peirce's Continuity Principle) that inside and outside cannot be discrete categories (Menary 2009). Menary also recognizes that external processes such as writing transform our cognitive processes (Menary 2007). However, he does not sufficiently explain in what sense then one can maintain the notion "inside"; how overcoming the categorical distinction between "inside" and "outside" alters our view of mind (and matter); and under which conditions the features that today are attributed to this presumed "inside" have emerged.

Sutton, another important representative of the "second-wave EM theorists," speaks of the "complementarity of disparate inner and outer resources" and focuses on how both contribute to more or less intelligent thinking and acting (Sutton 2008; Sutton et al. 2010, 524). Although he also recognizes the influence of external objects on cognition, the distinction that he makes still seems to presuppose that it is possible to account for an unmediated "within."

In another paper, Sutton does anticipate a so-called "third wave EMT," which "dissolves individuals into peculiar loci of coordination and coalescence among multiple structured media […]. Without assuming distinct inner and outer realms of engrams and exograms, the natural and the artificial" (Sutton 2010; see also 2009). AMT's starting point is the dismissal of this distinction between inner and outer realms, since the criteria for what it means to be "internal" are unclear, and might be part of this "third wave EMT." For this reason, it may be argued that AMT, rather than only contesting or refuting EMT, tries to make a positive contribution to the EMT debate by showing why taking a step further is important and necessary.

Overcoming the idea that cognition cannot be localized in an isolated inside brain world that functions as some kind of *causa sui* does not only require recognition that "[c]ognition leaks out into body and world" (Clark 2008, xxviii), but it also entails acknowledgment that cognition leaks in from the world into the individual's brain. Although AMT is sympathetic to approaches that validate this view and finds in them some confirmation, it does not only endorse the reciprocity between brain and world, but also argues for their interdependency: the "leaking out into the body and world" and "leaking in from the world into the individual's brain" must not be conceived as a cognitive process traveling from one autonomous sphere (brains) to another (world) or vice versa. Without certain external rules or mechanisms, we would, as Gallagher and Crisafi (2009) have argued, not be able to accomplish our thoughts. Cognition should be understood as a self-organizational process in which brains, bodies and world simultaneously participate and depend on one another. Take away one of these parts of the cognitive process and "the system's behavioral competence will drop" (Clark and Chalmers 1998, 9).

The neuroarchaeologist Malafouris illustrates this intertwinement of human and world by showing, for example, on the basis of

the practice of pottery, how the categorical distinction between the mental and the physical disappears. Intentionality, according to him, is not the internal quality of the subject. Rather, the intentions of the potter are co-determined by the physical properties of the clay and the way they react to his actions (Malafouris 2008a, b). A different line of thought will bring me, as I will elaborate below, to a similar conclusion.

In the next section, I will use Peirce as my guide and positively elaborate on his philosophy that overcoming the idea that thoughts emerge from an original inside, which then utilizes the outside world, not only discloses a radically different view of mind but also makes it possible to recognize the fundamental, constitutive influence of technical artifacts on mental processes.

4.4 The Artifactual Mind

Peirce on Mind as an Inherent Dimension of the World of Objects and Artifacts

Peirce's characterization of actual conscious states goes with a kind of reversal of the Cartesian dictum: introspective impressions are, he states, not the most immediate and transparent form of knowledge, but rather the most hidden. I quote: "Few things are more completely hidden from my observation than those hypothetical elements of thought which the psychologist finds reason to pronounce 'immediate'" (CP 8.144). By depicting immediate introspective impressions as "completely hidden," Peirce challenges the seemingly self-evident belief that we have a direct and privileged access to a separate "thinking world." If there are no special methods to secure a separate "thinking world," then the quasi-Cartesian propensity to situate mind somewhere inside our heads, in Peirce's words, the psychological tendency to consider "each of us is like the operator at a central telephone office, shut out from the external world, of which he is informed only by sense-impressions" (CP 8.144), becomes also implausible.

According to Peirce, thinking is not instigated by such internal impressions, but rather everything starts with what he sometimes calls "percepts," which are "out in the open." Peirce repudiates the idea that we have immediate access to our "inside realm" and a mediated access to the "outside world." That is why he can say: "It is the external world that we directly observe" (CP 8.144). Thinking is not instigated by "introspection" but by "extrospection." Although in this reversal Peirce still uses an inside–outside distinction, his argument ultimately culminates in a kind of collapse of that distinction. By depriving introspective knowledge of "immediacy" and "transparency," the criteria for what it means to be "internal" are blurred and, hence, the very distinction between "internal" and "external" becomes opaque.

To further clarify Peirce's perspective, I quote another passage from the same text:

> We first see blue and red things. It is quite a discovery when we find the eye has anything to do with them, and a discovery still more recondite when we learn that there is an ego behind the eye, to which these qualities properly belong. Our logically initial data are percepts (CP 8.144)

The (phenomeno) logical sequence is here of great importance. The percepts that instigate the cognitive process precede for Peirce every possible distinction, including a distinction between "inside" and "outside." That we attach to percepts a subject with a brain is something that logically (Peirce makes a distinction between a psychological and a (phenomeno) logical perspective) takes place afterwards. We should not reverse this sequence and presume that meaning and thought are located in, let alone produced by, an inner brain world. Peirce's repudiation of introspection has a phenomenological basis. All other less basic disciplines, including ethics, logic, metaphysics and the sciences, are dependent on the elementary findings that phenomenology offers.[1]

Building on his category of Thirdness, Peirce stresses that, and this is crucial, these percepts have a mental character. This, however, does not mean that they are products of individual brain processes. Indeed, according to Peirce:

> [t]hought is not necessarily connected with a brain. It appears in the work of bees, of crystals, and throughout the purely physical world; and one can no more deny that it is really there, than that the colors, the shapes, etc., of objects are really there (CP 4.551; see also Watts 2008, 190, 191, 199, 201)

Our world of objects and artifacts does not only consist of matter, but also of mind. This is almost hyperbolically expressed in a footnote of a text from 1884: "just as we say that a body is in motion, and not that motion is in a body, we ought to say that we are in thought, and not that thoughts are in us" (Peirce, CP 5.289, n.1; see also Aydin 2007a, b). If we are immersed in thoughts and the world is saturated with mind, the very distinction between an "inside world of human mind" and an "outside world of material objects" becomes obsolete.

A question that could be raised in response to this view: if mind is so fully pervasive of the universe, what is the principle of individuation that allows us to talk of minds and selves in the plural? It must be clear that Peirce's perspective does not allow, as we saw in Chapter 2, for a unique, pre-given inside as the principle of individuation. In his most anti-individualist writings, the individual is for Peirce no more than a privation:

"Psychological analysis shows that there is nothing which distinguishes my personal identity except my faults and my limitations" (CP 673). We discover ourselves as distinct selves if we are not able to sufficiently connect to our environment. Peirce's aim is, however, not the destruction of the individual, but rather the situation of the individual within a broader whole. A self develops a particular identity by virtue of his or her very empirical interactions and continuous attempts to govern and regulate them by certain social habits, which enable him to adequately respond to his or her environment. Although I cannot further elaborate this argument here (see Aydin 2009 for Peirce's view of individuality), it indicates again Peirce's belief that we are not detached, atomistic egos living in a separate inside world but "cells of a social organism" (CP 673), who discover and develop themselves in an interaction with their environment.

To get a better grip on Peirce's not easily accessible view of mind as an inherent dimension of the world of objects and artifacts, I will briefly discuss another philosopher whose ideas bear an intriguing resemblance to Peirce's view of mind, namely Karl Popper. According to Peter Skagestad (1993, 1999), Popper's account was directly influenced by Peirce's philosophy.

Popper on Exosomatically Held Knowledge in Objects and Artifacts

Popper explained the process of human evolution as a development of new organs outside our bodies, organs such as tools, weapons, machines and houses (also recall what I have said about Erns Kapp in the introduction of this book). He was especially interested in memory enhancing artifacts and writes in that context:

> Yet the kind of exosomatic evolution which interests me here is this: instead of growing better memories and brains, we grow paper, pens, pencils, typewriters, dictaphones, the printing press, and libraries. (…) The latest development (used mainly in support of our argumentative abilities) is the growth of computers (Popper 1972, 238–239)

Popper held knowledge to reside not in mental states but rather exosomatically in books, articles and the like. He emphasizes that it is a fundamental mistake to understand books and articles as only the outwards expressions of knowledge that is "really" residing inside the human mind.

Exosomatic knowledge is for Popper inherently dispositional: material objects have the potentiality to bring about certain states of mind. He writes:

> It is its possibility or potentiality of being understood, its dispositional character of being understood or interpreted, or misunderstood or

misinterpreted, which makes a thing a book. And this potentiality or disposition may exist without ever being actualized or realized (Popper 1972, 116)

Although a book is only a book if it is readable, it would potentially still be a book even if it were never to be read.

Objects and artifacts can be, in Popper's view, potential carriers of knowledge and meaning (Popper 1972, 107–108). If this is recognized, then it becomes easier to except that objects can shed more light on the human mind than the psychological features of mental states. In Popper's words: "Contrary to first impressions, we can learn more about production behavior by studying the products themselves than we can learn about the products by studying production behavior" (Popper 1972, 114). The products of the human mind teach us more about its evolution than its psychological features.

Popper stresses, like the proponents of EMT, the importance of the mind-expanding function of artifacts. In that context, he writes:

> we use, and build, computers because they can do many things which we cannot do; just as I use a pen and pencil when I wish to tot up a sum I cannot do in my head. 'My pencil is more intelligent than I,' Einstein used to say (Popper 1972, 225, n. 39)

Different from EMT, Popper emphasizes that the collective knowledge that dwells in objects and artifacts is relatively independent of particular individuals. The knowledge that resides in books, libraries and museums is not an extension of what goes on in an individual's brain but rather a cultural creation that has become independent of its particular creators. This way, Popper is able to situate knowledge in objects and artifacts.

Unfortunately, Popper's view does not sufficiently overcome the distinction between an "inside world of human mind" and an "outside world of material objects," as is evident in his Three Worlds Theory. The objective, abstract world of knowledge and ideas (third world), which is situated in and carried by material objects (first world), is only relatively autonomous, meaning that it is independent from separate particular human minds. Although the development of scientific theories can lead to unintended consequences, it remains the world of the products of a collective of human minds (second world). In addition, the human mind is a prerequisite for actually interpreting and grasping the third world of objective knowledge.

Although the knowledge that resides in material objects can become independent of its creators, Popper's material objects seem to function merely as passive knowledge storage. However, if mind would be an integral dimension of objects and artifacts in Peirce's sense they would be able to function as genuine epistemological agents. What is not

sufficiently accounted for by Popper is the dependence of thought on its carriers and their very physical characteristics. Popper's viewpoint is for this reason, as Paul Levinson claims, still "too primarily ideational" (Levinson 1988, 79): it is the human mind that generates and is the ultimate source of knowledge, which then can be stored in material objects.

To explain how the knowledge that resides in material objects is not the sheer product of a separate world of human minds and how its material carriers can play an active role in the generation of cognitive processes, I will in the next subsection, after having explained the relation between percepts and signs, briefly elaborate what signs are and how they work. It should be clear that my purpose here is not to provide a general description of Peirce's semiotics – which would be a very complex, maybe impossible enterprise, since Peirce continuously revised his thoughts on signs (see Robert Marty's "76 Definitions of The Sign by C. S. Peirce" 1997). In the last subsection, I will finalize my account of the artifactual mind.

The "Factual" Mind: Peirce on Percepts and Signs

Earlier we have seen that for Peirce, building on his category of Thirdness, the cognitive process is initiated by percepts, which are "out in the open." Percepts are "the data of all knowledge" (CP 8.300). As "the first premises of all our reasonings" they prompt a certain reaction, which is, according to Peirce, "utterly beyond control" and even "cannot be called in question" (CP 5.115f). We see here again a kind of reversal: the cognitive process is not generated by what traditionally is called the "subject," but rather what traditionally is called the "object." If the very concept of introspection is unsustainable, Peirce believes, we have to understand thought as something "factual." In Peirce's words: "It appears (…) that there is no reason for supposing a power of introspection; and, consequently, the only way of investigating a psychological question is by inference from external facts" (CP 5.249).

Although percepts are "factual," our access to them is always limited and designates a certain provisional stage in our comprehension. We only have certain representations of these "facts," which Peirce calls "signs." Without signs, cognitive processes could not be brought about. I quote a longer passage:

> If we seek the light of external facts, the only cases of thought which we can find are of thought in signs. Plainly, no other thought can be evidenced by external facts. But we have seen that only by external facts can thought be known at all. The only thought, then, which can possibly be cognized is thought in signs. But thought which cannot be cognized does not exist. All thought, therefore, must necessarily be in signs (CP 5.251)

Only through signs do we have access to the world. That which cannot be represented by signs, does not (yet) exist for us. Although Peirce still uses, as mentioned earlier, the distinction between internal and external, his concept of "sign" is an attempt to overcome the mind-matter dualism and, hence, the Cartesian inner-outer distinction. He sometimes uses in this context the notion "objective idealism" for his view (see CP 6.24; 6.25; 6.163).

It must be clear that percepts and signs are analogue (see also Hausman 2007). If we want to understand how percepts generate cognition, we have to interpret them in terms of signs. That is why Peirce can say: "Percepts are signs for psychology" (CP 8.300). Although Peirce understands cognition, in a similar fashion to Popper, as something potentially residing in the "factual" world, he does not conceive it as something that is produced by a separate world of human minds and can be stored in objects and artifacts. Cognition as "semiosis" is for Peirce a process that is brought about in and by virtue of our engagements with the objective world around us. To better understand this "bringing about," we have to briefly reflect on what signs are and how they work.

In 1897, Peirce defined a sign as: "something which stands to somebody for something in some respect or capacity" (CP 2.228). This sign has a triadic character: (1) "It addresses somebody, that is, creates in the mind of that person an equivalent sign or perhaps a more developed sign." Peirce calls this "the interpretant of the first sign." (2) "The sign stands for something, its object." (3) "It stands for that object, not in all respects, but in reference to a sort of idea" or "quality," which Peirce sometimes calls "the ground" of the sign (CP 2.228; see also 5.283).

To illustrate this triadic character of a sign, I give a simple example: one day I discover that one of my suits has holes and infer that there are moths in my wardrobe. The holes are a sign that represents the moths (the object) and instill a certain idea (an interpretant) in me, for example "damn moths." The set of particular qualities present in the sign (the ground of the sign) determine which interpretant is produced. They schematize the object so that it can be interpreted by some sign agency in a certain way. Signification is, in other words, dependent on in this case the very physical elements of the holes: the scattered shape and the small size of the holes are primary to the sign's ability to mediate between object and interpretant. The interpretant is, in turn, a sign, capable of further sign production in myself ("my 500 euro suit is ruined"), or should I want to utter it, in other selves. If I utter it, it could, for example, produce in the mind of my wife the interpretant "these are not moth holes, you silly fool, but cigarette burns; not moths but *you* have ruined your suit." This example also shows that each sign is what it is by virtue of its possible future "interpretants," which implies that "the meaning of a thought is altogether something virtual" (Peirce, CP 5.289).

There are two aspects in the sketched semiotic process that are relevant for our discussion. The first aspect is that the sign process is not generated by an inner realm of a subject but is determined and constrained by an object. The term "interpretant" should not be confused with "(human) interpretation." That a sign has to be suitable for interpretation means that the sign itself must have such a shape and structure that it can produce a specific interpretant in a certain (in this case a human) context. It does not mean that the interpretant is dependent on human consciousness, let alone that it is produced by it. Indeed, to prevent this possible misinterpretation, Peirce sometimes uses the term "quasi-mind" to describe the locus of the interpretant, and omits any reference to human being and human consciousness in his definition of a sign. I quote:

> a Sign has an Object and an Interpretant, the latter being that which the Sign produces in the Quasi-mind that is the Interpreter by determining the latter to a feeling, to an exertion, or to a Sign, which determination is the Interpretant (CP 4536)

The sign production is instigated by an object, which functions as a kind of *causa finalis* that prompts investigation. This investigation consists of properly recognizing and collecting the signs that the object produces and disclosing a certain structure or regularity, which enables ever better understanding of the object.

The claim that, even without some person interpreting a sign, thoughts could be produced is part of Peirce's non-anthropogenic view of thinking and might be too extreme and difficult to prove. It does, however, maximally stress the tenable view that the world around us plays an active role in instigating thought. We find meaning in the world rather than creating it. Thought and meaning are not brought about by a presumed private inside but are generated in our engagements with the objects around us. Even knowledge of ourselves is not acquired by a presumed "introspection" but requires collecting signs and figuring out what they represent. The following passage from Wittgenstein, which could have been written by Peirce, expresses this very nicely:

> But if you say: 'How am I to know what he means, when I see nothing but the signs he gives?' then I say: 'How is he to know what he means, when he has nothing but the signs either?' (2001, §504)

For both Peirce and Wittgenstein, thinking is the activity of operating with signs (Wittgenstein 1958, *The Blue Book*; see also Baker 1998).

The second aspect that is important for our discussion is that thoughts are not only exosomatically embodied, but that the specific physical characteristics of this embodiment also determine to a great extent which (type of) thoughts are induced. How a sign is brought into connection

with its object and which interpretant is being produced as a result of that is greatly dependent on its specific physical embeddedness. This brings me to the last part of this section.

AMT

The idea that mind is an inherent dimension of the world as such makes its specific material embodiment not less but more relevant. In a famous, and still very relevant, passage from 1905, Peirce tries to show how thoughts are determined by the worldly objects around us:

> A psychologist cuts out a lobe of my brain (nihil animale a me alie-num puto) and then, when I find I cannot express myself, he says, 'You see, your faculty of language was localized in that lobe.' No doubt it was; and so, if he had filched my inkstand, I should not have been able to continue my discussion until I had got another. Yea, *the very thoughts would not come to me.* So my faculty of discussion is equally localized in my inkstand (Peirce, CP 7.366; my italics)

The point that Peirce wants to make here is of course not that ink is the only possible means to utter his thoughts. More importantly, Peirce does not want to show – contra EMT – that he utilizes ink to reach certain goals in a more functional way. Thought is not something that emerges from an inside and, by functionally utilizing external objects, extended to an outside. Thinking is an activity in which the brain participates along with the eyes and the hands and a multitude of devices, including in this case ink.

The proposition "the very thoughts would not come to me" in the quoted passage indicates that, and this is crucial, ink plays not a passive but an active role in inducing certain thoughts. In and through the act of writing with ink and by virtue of the specific physical characteristics of writing with it, certain thoughts are triggered. Without ink, those thoughts would not have emerged. Conceiving thought not as an internal state of mind but as a process that takes place "out in the open" allows for ascribing to objects a thought generating capacity. In his article "Logical Machines," Peirce claims along this line that "it is no figure of speech to say that the alembics and cucurbits of the chemist are instruments of thought, or logical machines" (Peirce 1887, 168). Artifacts such as alembics and cooking pots evoke a very specific way of thinking in the chemist. The mind itself is evoked through and shaped by the very material characteristics of its carriers. The mind has an artifactual character. Although from our human perspective thinking would always require some engagement between an artifact and an organism, Peirce's non-anthropogenic view that all reality is saturated with mind necessitates him to entertain the possibility of semiosis without humans.

Here again, we find striking similarities with Wittgenstein, who writes:

> It is misleading then to talk of thinking as of a 'mental activity'. We may say that thinking is essentially the activity of operating with signs. This activity is performed by the hand, when we think by writing; by the mouth and larynx, when we think by speaking; and if we think by imaging signs or pictures, I can give you no agent that thinks. If then you say that in such cases the mind thinks, I would only draw your attention to the fact that you are using metaphor (1958, 6f)

In AMT, the external world does not consist of dead matter that is utilized by thoughts that emerge from an inside world. The repudiation of the inside–outside dualism implies that there are no purely physical objects that are operated by a separate, purely mental faculty. That is why the term "artifactual mind" is, I believe, not only more suitable than "extended mind, "but also more appropriate than "external mind": "external" still presupposes "internal," which is a distinction that the proposed perspective wants to overcome. Moreover, I prefer "artifactual" to "artificial" because "artificial" is too strongly connoted with "unreal" and does not express the "out in the open" character of mind.

"Artifactual Mind" expresses that our thinking is not pre-given or naturally present in a presumed inside world, but that it unfolds itself by virtue of and through objects and artifacts (Wittgenstein 2001, §16). It is crafted and shaped by physical things. For that reason, mind can be an object of archaeological or anthropological research (see Schick and Toth 1993, 49; see also Hilpinen 1993, 156f). From an artifactual perspective, thought is located in a world of objects, which are no less mental for being "out in the open" and no less real for being mental (see also Skagestad 1999, 557).

Our thinking is facilitated or impeded by specific physical features of objects and artifacts. We do not instrumentally use objects and artifacts to express ourselves or reach certain goals. Rather, objects and artifacts enable us to induce and develop certain thoughts. Dependent on their physical features, objects and artifacts, which are in our current time evermore technical artifacts, can "disclose," to borrow a Heideggerian term, the world in a certain way. The physical characteristics of the artifacts that we work with also determine to a great extent which goals are pursued and how they are pursued.

This view is also neatly expressed by Engelbart:

> Brains of power equal to ours could have evolved in an environment where the combination of artifact materials and muscle strengths were so scaled that the neatest scribing tool (equivalent to a pencil) possible

had a shape and mass as manageable as a brick would be to us – assuming that our muscles were not specially conditioned to deal with it. We fastened a pencil to a brick and experimented. (…) How would our civilization have matured if this had been the only manual means for us to use in graphical manipulation of symbols? (…) The concepts that would evolve within our culture would (…) be different, and very likely the symbology to represent them would be different (…) It thus seems very likely that our thoughts and our language would be rather directly affected by the particular means used by our culture for externally manipulating symbols (Engelbart 1962, 26).

Technical artifacts, such as pencils, books and computers, are not neutral tools for the more efficient execution of intellectual projects antecedently conceived and adopted; rather, technical artifacts inform our choices of which intellectual projects to adopt and thereby transform our conception of what we are doing and even of who we are (see also Ihde 1990; McLuhan and Fiore 1996). Artifacts are not neutral tools that are functionally utilized by an internal biological core, as expressed by EMT; rather, they shape to a great extent what we consider as our "inner," mental realm of goals, aspirations and ideals.

4.5 AMT, Normative Technologies and the Question of Self-Formation

EMT has contributed a great deal to the view that the external environment not only can influence cognitive processes but that external objects and artifacts, which are, especially in our modern, Western culture, ever-increasingly technical artifacts, can become a very part of those cognitive processes. According to EMT, objects can be utilized by the mind in such a way that they become extensions of the mind itself.

Instead of immediately adopting or rejecting EMT, I have tried to show that, despite significant differences, both opponents and advocates of EMT have in common that they do not sufficiently challenge the pivotal distinction between inside and outside. This reflection has led to an analysis of the concept of "introspection," since the distinction between inside and outside is often (implicitly or explicitly) justified by the presumed capacity of introspection and the ability to depict the contents of our mind as a separate realm. Different arguments against the epistemic validity of introspective judgments ultimately culminated in an overall rejection of the idea that thinking should be situated in an inside world – whether that is in the *res cogitans*, a realm beyond the empirical world, or, as in more modern materialist approaches, in the brains. Inspired by especially the philosophy of Peirce, I have argued that thinking has not only an exosomatic character, but rather is instigated by and unfolds through external objects and technical artifacts.

AMT evidences that a persistent inside-outside fallacy often lies at the heart of the failure to recognize a more fundamental, constitutive influence of external objects and technologies on cognition. External objects and technological artifacts are not dead matter utilized by a separate living, inside sphere with the aim to realize certain pre-established goals, but rather mind itself is induced and shaped by these artifacts. Mind is not extended by an inner biological cognitive core, but rather unfolds through objects and artifacts.

This perspective allows maximally recognizing and, henceforth, investigating how old and new technologies shape and "norm" thinking. Opponents of EMT believe that mental activities are expressions of organic brain processes, which display specific and characteristic mechanisms, such as primacy effects, recency effects and chunking effects. These typical features indicate, they believe, that cognitive processes are intrinsic and not reducible to external, social processes. The inner world of cognition has, they say, features that are completely different from external processes, which indicates that cognition is brain-bound. AMT can reveal that this "inside–outside" distinction does not take into account the history of the terms attributed to cognition and, hence, reflects a "short-term memory." A genealogy of concepts such as "sequence," "listing" and "grouping," which are presupposed in notions like "primacy effect" and "chunking," can provide insight into how these terms can be traced back to technologies, such as reading, writing and calculating, and the artifacts employed for those activities, such as pens, calculators and computers. Without the development and ubiquity of these technologies, today we would not interpret presumed "internal" cognitive processes exclusively in terms of processing information and "processing information" would not be understood in terms of the above-mentioned concepts. It is also no coincidence that notions like "primacy effect" and "chunking" are studied in the context of more or less well-*functioning* memory. The functionalist character of modern technologies determines to a great extent our view of memory (see Donald 1991). My goal is not to deny that our thinking is subject to particular mechanisms, but to illustrate how old and new technologies frame these mechanisms and have continuously inscribed their "grammar" into what is considered the natural, non-derived mark of cognition.

Not only internalists but also advocates of EMT have this "blind spot" with regard to the inside-outside dualism, which prevents them from being able to sufficiently recognize the technological dimension of cognition. Clark's insufficient liberation from the "inside–outside" dualism prevents him from seeing how much his view of cognition is determined by current technological idiom. Although he partially locates cognition outside the brain, he, at the same time, upholds this dualism by ascribing to cognition an original starting point in an internal biological core, an inside that utilizes the outside world in order to fulfill certain

cognitive tasks that it has set for itself. Within this framework, Clark interprets the mind in terms of all the way down functionalist terms, without taking into account that this characterization is already an expression of the current technological vocabulary (Selinger and Engström 2008). "Computation," "processing," "modularity" and "functionality," which for him ultimately characterize what cognition is, are not neutral, self-evident and irreducible notions but part of the technological mindset that we have incorporated in our modern era.

If our "inner self" is shaped by the technologies that we have invented, to which we have adapted and that have become a structural part of our modern life (see also Tenner 2003), are we then not determined by technologies? This question allows us to take up again more explicitly our leading theme of whether self-formation is possible: if technologies not only shape our environment but also our mind (which was for Descartes the faculty that secured independence), to what extent is self-formation then still possible? Have we not replaced one determining force with another? Instead of brain-determinism, we now seem to have to endorse technology-determinism, which can be seen as a kind of "invasive situationism."

First, AMT displays that in both internalist views and in EMT, the radical influence of technologies is rendered invisible. By locating particular functionalist features of cognition in an alleged independent and irreducible inside brain-world, it becomes possible to understand mind-as-such exclusively in terms of instrumental problem solving or the enhancement of information processing functions. In this way, the technologies that have generated this type of thinking are kept out of sight. They become indeed, as Clark states, "transparent." However, this transparency comes with the risk of disregarding their profound influence. As a consequence, a fundamental criticism of particular framing of cognition becomes impossible: for example, it becomes no longer possible to criticize the functionalist framework, since all technology as well as the thinking that it facilitates is solely evaluated on the basis of the greater efficiency with which it enables us to carry out familiar tasks.

Second, by depriving cognition of its original, non-derived and independent features (the "mark of the mental"), situated in a presumed inside, it now not only becomes possible to recognize how the mechanisms attributed to it were influenced by certain technologies, but also *to make visible and evaluate* those technologies and the thinking that they are generating. It becomes, for example, possible to question whether it is adequate to conceive technologies and the thinking that they are generating exclusively in terms of efficiency. AMT does not exclude normative questions, but rather places them at the center of our attention: what is the value of particular technologies in terms of the type of thought that they evoke for our lives? How should we evaluate the mental realm of

goals, aspirations and ideals that those technologies trigger and shape? In short, how can technologies contribute to forming a good self?

   Third, it must be clear that the question of self-formation is now addressed within a different, more suitable framework, namely one that does not separate an inside-world from an outside-world and attempt to establish which realm controls the other. Answering this question requires further elaborating how the proposed non-essentialist and non-dualist perspective enables not only recognizing the technological dimension of the self (AMT), but also reinterpreting what it then still can mean for a self to form itself. If technology is increasingly shaping my mind and self, how can I then still justify that it is *me* that forms my self? In the next chapter, I will further elaborate how self-formation beyond the inside-outside dualism is possible by challenging a particular usage of a type of technology that presumes being able to scientifically depict the relation between mind, brain and self and potentially explain whether the self is free, namely brain imaging technology. In this respect, I will also elaborate the notion of "critical self-formation."

Note

1 Peirce termed, probably independently from Husserl, his overall approach as "phenomenology." Later, he uses the notion "phaneroscopy" for his method. See Aydin (2007a) for a comparison between Peirce's and Husserl's phenomenology.

5 Brain Imaging Technologies and Critical Self-Formation

The Artifactual Mind Thesis (AMT) not only shows that the "inside-outside" fallacy obscures the fact that our cognition and self are shaped by particular technologies, but also indicates that overcoming this fallacy and recognizing the technological dimension of mind and self can reframe and bring about a notion of self-formation that goes beyond autonomous-determined twin traditions. Today brain imaging technologies such as Positron Emission Tomography (PET) and functional Magnetic Resonance Imaging (fMRI) are increasingly used not only to diagnose diseases and lesions, but also to correlate brain activation with psychological states and traits. Brain imaging technologies are used to find neuronal networks and brain regions that are specific to the functional realization of aspects of the self (e.g., Greene et al. 20014; Legrand and Ruby 2009; Christoff et al. 2011; Haggard 2017). These studies inform debates in philosophy of mind as well as philosophical, psychological and legal conceptions of, for example, freedom and responsibility (e.g., Klemm 2010; Vincent 2013). Neuroscientists, psychologists and philosophers who draw a direct and causal relation between brain states and cognition even claim that through brain research, we could gain "immediate" access to our thoughts. These claims and expectations make brain imaging technologies of special importance for our current investigation: data and images provided by brain scanners are used to draw causal relations between brain processes and cognitive functions, and to potentially resolve the question of whether the self is free or determinated, and, hence, self-formation is possible.

The relation (type-identification) that is drawn between processes in the brain and mental states is, of course, very problematic, as extensively discussed in the philosophy of mind. But that problem is not my focus here. In the light of AMT, I will rather argue that the interpretation and usage of brain imaging technologies as functional evidence strongly relies on the disputable idea of the existence of an autonomous self that is located in the brain. Although methodological objections have been uttered against the view of brain images as valid and reliable representations, these objections neither sufficiently challenge nor explicate how brain imaging technologies are able to instigate a framework that allows

the brain to be approached as an agent that produces cognition and regulates behavior.

Not only will I show that this framework can be traced back to and is implicitly justified by the Cartesian notions of self and freedom, but I will also argue that these particular notions were influenced by an older technology, namely the camera obscura. It must be clear that I neither claim that the camera obscura can exclusively explain nor that it is the only source of Descartes' views of self and freedom. However, this reconstruction will illustrate how technologies, which often remain virtually unrecognized, co-constitute seminal ideas in the history of philosophy, including notions of self and freedom. Arguing that even the notion of an internal, pure self that is independent of the material world was co-constructed by the metaphor of a particular technology can be seen as my most radical attempt to go beyond the inside-outside distinction.

Although this comparison suggests a continuity between the camera obscura and brain imaging technologies, I will also point out a strong discontinuity: in contrast to the Cartesian view, the usage of brain imaging technologies in this regard presupposes that the self and its functions can be externally accessed by (neuro)scientific means. Brain imaging technologies challenge the idea that privileged access to the self merely can be obtained through immediate, subjective access to our ideas. Although I will not endorse the idea that brain images represent the brain and can explain psychological functions, I will argue that "extrospective knowledge" provided by brain images allows us to explain how brain imaging technologies could play a role in the process of self-formation from the proposed interactionist perspective.

From the proposed view, the brain is neither regarded as the locus of the self nor are brain imaging technologies conceived as offering representations or photographs of the brain. Instead, the self will be further elaborated as continuously forming itself in the world by pursuing certain goals and ideals and developing matching habits. The idea of orientation toward goals and ideals, inspired by in particular Peirce's pragmatism and his view of Thirdness, will prove itself fruitful in illustrating how the self does not have to be completely determined by its brain, without denying the role of the brain, as depicted by brain imaging technologies, in the process of self-formation. Brain imaging technologies can contribute to this process because they could potentially display to what extent our biological constitution is on par with or poses limits or constraints to our personal or social ambitions. The extrospective relation to and interpretation of brain imaging technologies allow critically reflecting on whether our habits are consistent with the long-term goals or ideals that we pursue. Also taking up again the idea of freedom as "self-determination," self-formation will be further explained as "critical self-identification." Building on AMT and Technological Mediation

Theory (TMT), the role of artifacts and technologies in self-formation processes will be further elaborated.

5.1 Brain Imaging Technologies as Functional Evidence in Free Will Debates

In Chapter 3, I briefly discussed brain-determinism and brain-non-determinism. We have seen that brain-determinists reject the existence of free will because they deny that my actions are caused by conscious intentions or decisions, whereas brain-non-determinists argue that conscious intentions and decisions can intervene in and influence both brain inclinations and environmental influences. What I did not highlight in that chapter is that both opponents and advocates of free will often draw upon the same brain imaging technologies in supporting their positions. Again, it must be clear that my aim here is not to directly contribute to the free will debate as such. Engagement with this debate is rather yet another steppingstone in the process of developing my view of self-formation in a technological world.

Although opponents and advocates provide very different and contesting interpretations of brain images, both groups often assume that data and images provided by brain scanners in one way or another depict causal relations between brain processes and cognitive functions. This usage of brain images raises the question of what exactly brain images represent or visualize and how they are supposed to provide functional evidence. Brain images are colorized pictures of alleged brain activity, which can be measured in different ways. I will focus on fMRI, which is a dominant technology in brain mapping research and is increasingly used in psychological explanations. fMRI measures the so-called Blood Oxygenation Level Dependent (BOLD) response. When a brain area is more active it consumes more oxygen, which is delivered by an increase in blood flow in the activated area. Oxygenated and deoxygenated hemoglobin in the blood have different magnetic properties. An MRI scanner can read changes in magnetization and translate them into a computerized image (Matthews and Jezzard 2004).

Neuroscientists and philosophers who draw a direct identity relation between brain states and cognition use data and images provided by brain scanners to infer something about the role of particular brain areas in cognitive functions. Brain imaging data are also increasingly used, in particular in free will debates, to make a so-called "reverse inference": brain imaging data are taken as evidence that a given brain region plays a particular causal role during the performance of a cognitive task. There are "neuro-optimists" who go one step further and assert that brain images will permit immediate access to our thoughts. They presume that brain images will open up the "black box" of the mind to direct observation (Camerer et al. 2005).

A general objection against these kinds of assumptions is that brain images are not simple pictures of BOLD signal differences. It is not possible to directly measure quantitative signal magnitudes. General mapping from BOLD signal to significant neural activity is lacking (Nair 2005). In addition, the BOLD differences that are associated to brain activity are small, noisy and multifaceted. According to Klein (2010), brain images do not provide quantitative information, but rather display *statistical* significant differences in BOLD signal between different task conditions: inferential statistics are used to contradict the so-called "null hypothesis" that indicates that an experimental condition had no real effect on the observed MR signal. Brain images do not directly display activation in brain areas, but are rather maps of places where we can assume with a certain probability that the presumed correlation (speaking of "causation" is too premature) between data and a stereotyped pattern of activation is unlikely to be the result of chance fluctuations from a true zero signal (Klein 2011).

Roskies (2007) draws a similar conclusion by arguing that brain images should not be conceived as photographs of brain activity. Brain imaging does not let us directly see properties carrying visual data of the brain. Unlike photography, brain imaging is belief-opaque: the information crucial to its correct interpretation is, Roskies argues, not contained in the image itself. The assumed functional decomposition, the tasks involved in producing an image and the chosen statistics are necessary for interpreting the image but not recoverable in information inherent in it. Inferences are, therefore, dependent upon the beliefs and experience of the researchers about how to analyze their results. What further complicates matters here is the question of whether "direct representation" as such is possible. One can acknowledge that certain conventions or beliefs might always be involved in recovering information. However, it still seems to be possible to make a distinction between more and less common or accepted conventions and beliefs shared by a cultural community.

No less controversial is reverse inference: because brain regions can perform multiple functions and be involved in a variety of different processes, a causal relation between activation of a certain brain area and a particular cognitive function cannot be drawn (Klein 2010). The many conceivable mappings between brain regions and cognitive functions allow for a combinatorial explosion of possible explanations of observed brain activity (Klein 2011). Discovering more systematicity in the visual field might strengthen certain explanations (see also Klein 2017). However, currently reverse inference provides virtually no evidence and is deductively invalid (Poldrack 2006, 2011).

These objections against the validity of brain imaging indicate that brain images are not direct representations of brain activity and that they provide very weak functional evidence. However, for most critics these concerns about validity and reliability do not render brain imaging worthless. According to Roskies (2007), "imaging data is, within

its limits, a reliable indicator of brain activity" (2007, 867). Cacioppo et al. (2008) believe that brain imaging can help generate hypotheses in an exploratory way, which can be tested in future studies. In a similar vein, Klein argues: "neuroimages do not confirm functional hypothesis, but they do show brain areas in which the imaging data might be further used to confirm functional hypotheses" (2010, 275).

These methodological considerations and evaluations have two relevant implications within the scope of our investigation. On the one hand, they clearly show that brain images do not directly display neural activity but are, in fact, very much dependent on theoretical beliefs and statistical suppositions. On the other hand, these methodological assessments regarding validity and reliability neither challenge nor sufficiently explicate the particular way brain imaging technologies frame the relation between brain, cognition and world.

By focusing only on methodological issues concerning functional evidence, the more basic framework that *allows* the brain to be approached as an agent that determines cognitive functions and utilizes the external world is neither revealed nor questioned. The objection that brain regions can perform multiple functions and be involved in a variety of different processes does not challenge the idea of the brain as an agent that causes cognitive functions. Although the discussed methodological reflections indicate that brain images might not be reliable and valid pictures of brain processes, they do reinforce the view of the brain as a *causal agent* by framing the brain as an isolated realm that – if certain methodological problems are solved – *could* provide functional hypotheses. Brain imaging technologies frame in a particular way the relation between brain and self, and bring about a particular notion of "free will."

5.2 Technological Inscription from Camera Obscura to fMRI

AMT has been developed in a constructive criticism of Extended Mind Thesis (EMT) and highlights the technological dimension of mind. I have indicated that technology not only shapes our mind, but also our self. In order to further nuance AMT in the context of self-formation, I will reflect on a theory that exhibits similarities to AMT, namely TMT. Then, I will illustrate how a particular technology, namely the camera obscura, has contributed to inscribing the idea of an independent, inner inside to the self and how today this idea is both sustained and transformed by brain imaging technologies.

TMT, AMT and Resilient Technological Inscription

The idea that technologies influence how we experience both the world and ourselves forms the core of TMT (Ihde 1990; Verbeek 2005). This

approach analyzes technologies as "mediators" of the relations between humans and world. Rather than seeing technologies as material "objects" that need to be understood in opposition to human "subjects," the mediation approach understands technologies as constitutive for human practices and experiences.

Don Ihde (1990) has provided four different ways in which technologies can mediate the relation between humans and the world as well as human intentionality. The first type of human-technology relations he calls "embodiment relations" refers to technologies that become part of our bodily constitution; the prime example of this type is a pair of glasses that becomes part of our sense organs and that we do not look *at* but *through*. The second type Ihde calls "hermeneutic relations" refers to technologies that we "read" in order to acquire information about the world; the prime example here is a thermometer that does not give a sensation but a number that represents the temperature. The third type he calls "alterity relations," which includes technologies that we interact with as quasi-others; the prime example of this type is an ATM that we interact with in order to draw money from our bank account. The fourth type Ihde calls "background relations" refers to technologies that influence our experience of the world without getting noticed; the prime example of this type is an air conditioning system that is not always perceived as such but still influences how people experience their environment.

Verbeek (2008) complemented TMT by illustrating how technologies not only mediate the world, but also the way we experience ourselves and our intentions to act as well as our ethical decisions and actions (Verbeek 2005, 2011). He has illustrated how sonograms mediate the intentional relation between expecting parent(s) and the fetus: it constitutes the fetus as a separate person, as gendered and as a potential patient; in addition, it transforms expecting parents to parents who have to make technologically informed decisions, moving pregnancy from the realm of fate to the realm of human responsibility (Verbeek 2008). This case clearly shows in what sense technologies are not neutral tools: technologies disclose the world in particular ways as they organize and reorganize human routines, habits and morality, and channel and re-signify social relations (Rosenberger and Verbeek 2015).

AMT shares with TMT that technologies influence the way we experience the world and ourselves. What makes the scope of AMT different, which can also be seen as complementary, is its focus on how technologies gradually inscribe certain *notions* or *categories* in our mind that, if becoming dominant, durably shape our mind and self, and resiliently structure and regulate how we experience the world and ourselves and how we interact with our surroundings. In the previous chapter, I have discussed how terms like "computation," "processing," "modularity" and "functionality," which were derived from computers, have become

so prevalent that they currently predominantly characterize how cognition as such is conceived. Now, I will argue that the conception of "freedom" assumed in many contemporary debates about freedom and responsibility has been strongly influenced by particular technologies.

It must be clear that my aim is not, as I have stressed repeatedly, to disclose a more "original conception" of mind, but rather the opposite, namely to recognize and reveal how mind has always been shaped in a particular way by external factors and technologies. Recognition of and insight into the technological dimension of mind enables in particular debunking reductionist claims that ascribe to thinking presumed original and irreducible features. Instead of revealing a presumed "original content," AMT enables recognizing and making visible how particular technologies have generated a particular conception and form of thinking and, as we will see, to evaluating to what extent these particular framings foster "good" self-formation.

Technological inscription in and framing of our mind and self can have different degrees of solidification. Some inscriptions are changing as we speak. An example: although "health" is theoretically defined in different ways in dictionaries and medical handbooks, in our everyday life encounters and interactions, it greatly derives its meaning from technologically generated measurements of variables like weight, heart rate, blood pressure and insulin level. However, today we witness more and more people using apps to "measure" and watch their health: running distance trackers, calorie counters, nutrition assessors, sleep cycle analyzers and so forth. These technologies are not neutral tools that merely measure "health"; in fact, "health" is co-determined by what these tools measure and by what is left out and not disclosed. They provide new norms for what is considered as "healthy" and influence whether people consider and experience themselves as healthy. In the long run, these new health technologies could gradually become dominant and alter and possibly surpass older conceptions of "health."

There are also framings and inscriptions that have been engraved in our mind, self and world as a result of a particular dominant technological grammar to such a degree that they have become "ontologized," are attributed "originality" and have become part of our cultural and ethical canon. This robust solidification often runs parallel with obscuring the historicity and genealogy of the concepts and forms in question. One could say that brain imaging technologies frame the relation between brain, self and world within the free will debate, in a similar vein as health apps frame "health"; they strongly influence how freedom and control are approached and understood through what they apparently measure and leave out. However, on closer inspection, framing the brain as a kind of a *causa sui* and attributing to it the capacity to control behavior, determine cognitive functions and utilize the external world appears to have a much longer history. Hence, brain imaging

technologies have not initiated but have rather rehabilitated this particular inscription, which took root and germinated to such a degree that it durably and resiliently structured our very thinking and became "ontological." Revealing how the idea of freedom has been influenced by a much older technology, namely the camera obscura, will contribute to understanding the impact of brain imaging technologies on this concept.

The Camera Obscura, the Self and Freedom; or How Even the Idea of a Pure, Unmediated Self was Technologically Mediated

In previous chapters, Descartes' notions of self and freedom have been explained as outcomes of an epistemological investigation. In methodical self-contemplation, the self discovers that our access to the contents of our thoughts (introspection) radically differs from how we encounter the external world of material things: we can have an authoritative, privileged and immediate access to the contents of our thoughts, which we lack toward the external world of material things (Macdonald 2007; see also 2001). Because introspective beliefs about my mind are radically different from my beliefs about the outside world, and I can clearly and distinctively conceive both as separate realms, my mind must be a completely different substance than the outside world, including my body. We have seen that this argument has also another implication: since we have this privileged access to our mind and not to our body and the world, which makes us doubt their existence, we identify ourselves primarily with our conscious thought: I think, therefore I am (*Meditations* 1641/1985). Moreover, the mind is in one way or another able to exert control over brain, body and environment via the pineal gland, whereas the body is a machine that is subject to mechanistic laws (Lokhorst 2008). The sense of being an autonomous agent is derived from the *cogito*, which is ontologically separated from the external world but in some way capable of controlling it.

The imperative prejudice that Descartes did not put in brackets in his methodical doubt experiment is that it is altogether *possible* to fundamentally detach oneself from the world and evaluate different ideas on the basis of clarity and distinctness. Although this presumed ability has been extensively challenged throughout the history of philosophy, especially from phenomenological perspectives (intentionality, being-in-the-world, embodied, embedded, extended enactive cognition etc.), not much attention has been given to the historical *material-technological* conditions that contributed to the emergence of this view and made it appear cogent.

How an idea, view or theory emerges is in traditional philosophy of science considered to be part of the "context of discovery" and is not seen as important or relevant for its validation. However, from the proposed

perspective recognizing the "context of discovery" can help to uncover relevant presuppositions of the theory in question and indirectly provide cues that could be used to challenge its "context of justification."[1] Recovering how the camera obscura played a role in Descartes' "context of discovery" will not only help to uncover how certain notions of self and freedom have been preserved and transferred but also illustrates how even our most fundamental and theoretical-speculative philosophical notions have a material-technological grounding.

I am not the first person to highlight the connection between technology and Descartes' view of the self. Lee W. Bailey (1989) has argued that the camera obscura is one of the root metaphors that helped generate the image of the mind as an inner realm, ontologically separated from the external world, an image that is characteristic of our modern suppositions of subjectivity and autonomy. Technological mediation theorists Petran Kockelkoren (2003, 2007) and Don Ihde (2007) have explored this view along similar lines.

In the 16th century, the camera obscura was widely used for perspective drawing. Descartes was very familiar with this piece of technology. In his *Optics*, he has described it and used it as a metaphor to explain how the world enters our mind through our senses:

> The objects we look at do imprint very perfect images on the back of our eyes. Some people have very ingeniously explained this already, by comparison with the images that appear in a chamber, when having it completely closed except for a single hole, and having put in front of this hole a glass in the form of a lens, we stretch behind, at a specific distance, a white cloth on which the light that comes from the objects outside forms these images. For they say that this chamber represents the eye; this hole, the pupil; this lens, the crystalline humour, or rather, all those parts of the eye which cause some refraction; and this cloth, the interior membrane, which is composed of the extremities of the optic nerve (Descartes 1965, 91)

This comparison shows – as also Bailey and others have pointed out – that the central perspective of the camera obscura makes the idea of an aloof spectator position, as opposed to a geometrical *res extensa*, a *plausible starting point* for reasoning. It makes it altogether possible and visual to fundamentally detach oneself from the world and evaluate different ideas on the basis of clarity and distinctness. From this perspective, the *camera obscura* is thus not an innocent and neutral metaphor that Descartes used to explain his epistemology but mediated the idea of the mind as an internal representation of a world that can be observed (and acted on) by a homunculus situated somewhere outside the world; the autonomous subject is modeled after a spectator inside a camera obscura.

This particular framing of the relation between mind, world and self has also far-reaching consequences for conceptions of freedom. The

metaphor of the camera obscura reflects the idea of freedom as a quality that is characteristic of an agent that, being sheltered in an impenetrable "black box," is independent of the world and the laws and mechanisms that determine it and, at the same time, has control over it on the basis of the images that are reflected on the screen of its (the agent's) internal control room.

The Cartesian framework, influenced by the *camera obscura*, does not mark a historical event that we have left behind, but one that has been transported and rehabilitated by other and different technologies. In the context of the medical field, which is increasingly spilling over to the psychological and philosophical realm, this continuation is facilitated by brain imaging technologies. In the intersection of neuropsychology, neurophilosophy and neuroethics, brain imaging technologies are the dominant mediating technologies on the basis of which the self is attributed certain psychological functions and capacities, such as the capability to will, decide and act freely. In current brain research, there is a strong tendency to conceive the brain as constitutive of the "real self" or as the "control room" of the realization of aspects of self-related experiences, as an agent that instigates cognitive processes, moves the body and manipulates the external world (e.g., Northoff et al. 2006; Legrand and Ruby 2009). Although Descartes would never have situated cognition in or reduced it to brain processes, since brain and body are also part of the *res extensa*, the division between "internal" and "external" that breaks along the line of an independent agent is, as Joseph Rouse has pointed out, a Cartesian legacy (1996). Instead of the mind, it is the brain that is envisioned as a kind of *causa sui* that exerts control over the way (aspects of) the self are realized.

What is relevant here is that brain imaging technologies mediate and transfer similar Cartesian notions of self, freedom and responsibility. By focusing on the inside of the head and (interactions between) certain regions of brain activity, an fMRI image depicts the brain as an isolated realm and allows approaching it as an independent inner agent (see also Doris 2015, 37f). This way, a categorical distinction between "inside" and "outside" is preserved. The brain is attributed a privileged status because of its presumed capacity to detach itself from the outside world, process information that is collected by the senses and determine, as a kind of *causa sui*, our decisions and actions.

Advocates and opponents of free will do not agree on the question of whether consciousness can be attributed causal efficacy (see also the entry "epiphenomenalism" in Stanford Encyclopedia). However, they often share the view that freedom can only be attributed to an agent (consciousness or brain) that is detached or has the capacity to detach itself from its environment or other external influences. They believe that freedom can only exist in a separate, autonomous realm, that is, a realm not fundamentally affected or determined by external factors. Only then is it possible to process external input by virtue of which behavior can be controlled and the external world can be utilized.

Thus far, the presented analysis has a descriptive and not an evaluative character. If technologies influence and mediate to a great extent our conceptions and experiences of the world and ourselves, what then is wrong with the conception of self and freedom/control suggested by brain imaging technologies in the wake of the camera obscura? Answering this question enables revealing the ultimate paradox in the Cartesian notion of self and autonomy. It must be clear that the problem is not that the Cartesian conception is technologically mediated. The problem lies in (1) the *particular* self-identification and self-interpretation that brain imaging technologies encourage and (2) the insufficient acknowledgment of and accounting for the mediating role of these technologies in self-identification and self-interpretation processes. Both impediments are, as I will attempt to show, related.

The *particular* self-identification and self-interpretation that brain imaging technologies generate is problematic for reasons that I discussed earlier. Brain researchers and philosophers who identify "our real self" with the brain and attribute to the brain causal efficacy and control conceive the brain, relying on fMRI-images, as some kind of *causa sui* that is detached from the outside world and can determine our decisions and actions. Clear criteria that ensure the brain (together with its physically realized mental states) as a kind of isolated "inside" that can be detached from the "outside" world are, as I discussed in Chapter 4, lacking.

The view of the brain as an autonomous agent underlies our second impediment: the effect of insufficiently taking into account the mediating role of fMRI technologies in self-identification and self-interpretation processes. The idea of the brain as an autonomous agent – a detached *causa sui* – is tenable insofar as the mediating role of the technology that co-instigates that idea remains in the background and is disregarded. It validates itself insofar as the technical conditions that have generated it remain invisible.

The proposed reconstruction of the genealogy of the Cartesian self and the influence of the camera obscura on its emergence shows that even the notion of a pure, unaffected self was co-produced by the metaphor of a particular technology. This genealogy highlights the impossibility of detaching our self-conceptions and self-identifications from influences of our surroundings. These influences, which are increasingly technological, are overseen especially when they become part of commonly accepted practices. They then become, as Andy Clark (2003, 123) (uncritically) states, *transparent*.

How Brain Imaging Technologies Compete with Introspective Knowledge

There is continuity between the camera obscura and brain imaging technologies insofar as they mediate a particular view of self and freedom, but there is also an important and relevant difference between them.

In Descartes' epistemology, the *cogito* is a "black box" that is only accessible through subjective introspection, whereas brain imaging technologies promise objective knowledge of how (aspects of) the self are realized. For Descartes, the *cogito* could not be scientifically objectified and investigated (in the modern sense of the word). In contrast, fMRI has – given the status attributed to it in contemporary debates about free will (e.g., Lamme 2006; Soon et al. 2008) – the potential to eventually display the interior of the "dark room" and to disclose not only *what* really determines the realization of our self and our behavior, but also *how* it is able to do it. From this perspective, knowledge of the "real agent" of our actions and decisions, that is, of the brain, is no longer based on speculative, subjective assumptions, but can be scientifically investigated and validated.

The assumption that fMRI can provide objective knowledge of how the brain causes decisions and actions does not weaken the camera obscura framework of freedom, but rather reinforces it and makes it more credible. The idea of freedom "not being determined by something else but capable of determining behavior and utilizing the world," which is now attributed to the brain – and instead of "free will" often is reframed in terms of "control" – appears more credible because it is investigated in a scientific setting.

However, the credibility of approaching the self as something that can be objectified in a scientific setting also breaks with the Cartesian framework of the mind with regard to conceptions of self-experience. In the Cartesian framework, the self is located "within," and privileged, authoritative knowledge of it can only be obtained through introspection. By definition then, reliable knowledge of the self is subjective and cannot be outsourced to processes in the external world. Within a strictly Cartesian perspective, neuroscientific insights based on brain imaging technologies cannot be used to make authoritative knowledge claims about self-experiences.

This picture changes drastically when the human brain is thought to be constitutive of the human self. The self that allegedly is located "within" is no longer a private realm. Visualizing (aspects of) the self on brain scans supposes the possibility of making inferences about what the subject is going through (Davey et al. 2016). In fact, introspective knowledge and knowledge that is based on brain scans may contradict each other, and studies suggest that no clear factor to determine which of the two knowledge claims is correct (e.g., Legrand and Ruby 2009; Christoff et al. 2011). For example, attributing an action to oneself is no longer something that can (just) be introspectively determined, but is (also) something that can be externally inferred on the basis of observed patterns of brain activity (Farrer and Frith 2002). The truth-value of introspective claims such as "I believe I have executed action X" thus might be effectively challenged by inferences made on the basis of

brain imaging technologies. These views reminisce older speculations: Armstrong anticipated the idea that a "brain technician" with perfect understanding of brain processes would be able to correct the mental states that someone ascribes to himself (Armstrong 1968/1993, 109; see also Haynes and Reese 2006).

The view of the self as mediated by brain imaging technologies is thus not one in which self-knowledge can exclusively be obtained "from within." If brain scans can indeed be used as means to settle debates about (in)voluntary action and free will, the privilege of accessing experience is no longer limited to the experiencing subject, but the self and its actions become potentially accessible for external informed observers. Brain imaging technologies offer the prospect of allowing scientists to make predictions about experiences of people without relying on re-portability. It must be clear that relying on scientific predictions only make sense when a certain authority is ascribed to them, which makes them epistemologically competitive to introspection.

This discontinuity instigated by brain imaging technologies further debunks the idea that self being primarily determined by mind/brain (as inner) and world (as outer) are fundamentally detached from one another. While brain imaging technologies provide an image of the brain as detached from the world, which is continuous with a Cartesian framework, they undermine the Cartesian idea that introspection is a source of privileged, authoritative knowledge of the self. Through brain imaging technologies, knowledge of the self can be obtained *in* the external world (*res extensa*), thereby challenging the assumption that there is an ontological divide between mind/brain and world. However, actually bridging this divide is dependent on how radically the inside-outside model is debunked: considering the brain as a new, independent "inside" on the basis of brain imaging is recapturing this view through the back door. Nevertheless, the view that the self can acquire knowledge of itself in and through this world can be used to challenge the model that our self is categorically detached from the world and opens up, as I will demonstrate, a different framing of the role of brain imaging technologies in self-knowledge practices.

5.3 From Self-Determination to Self-Formation

If the self can acquire knowledge from itself through brain imaging technologies, it can also relate in a certain way to "itself" or at least an objectified dimension of itself. This knowledge can inform the self about its inclinations, habits and conduct, and raise the question of whether its present conduct coincides with its desired conduct. Put differently, this information can allow the self to evaluate its present conduct and, if needed, form itself in another or more preferred direction.

This alternative approach to the self as something that can form itself in and through information acquired from the-self-in-the-world can

break with the idea that brain imaging technologies can be understood as representations of an inner self. From this perspective, the self (or brain for that matter) is, in accordance to EMT and AMT, not to be found "in the head." In fact, this framework aims to surpass the discussion of whether the self should be understood as residing "inside" or "outside" the human body by making obsolete the question of where to locate the self. Rather, the self is understood as an ongoing formative process that takes place in a world mediated by existing institutions, scientific developments, technologies and the biological functioning of the human body. The formation of the self is conceived as a *process* of developing habits that are on par with long-term goals or ideals, which are set against the (environmental, psychological, biological) constraints and opportunities encountered during the process of formation.

When understanding the self in this way, not only is the question concerning self-localization surpassed, but the interest also shifts to *how* the self can form itself in relation with the world in which it is unfolding. Now, also the question that has been addressed in the previous chapters can be taken up again: does blurring the distinction between self and world not destroy the autonomy of the self? If there is no independent self that can be separated from its interactions (neither as an isolated mind nor as an isolated brain) and if the self is rather *formed* by its interactions, would the self then be nothing more than a plaything of random and/or enforced influences? And if technologies inscribe their forms to human cognition to such a degree that the self interprets itself in technological terms, does the self, instead of an agent, not become a patient? I have indicated in Chapter 3 that I subscribe to the intuition that autonomy should be understood as self-determination as well as to the idea that freedom and control requires orientation toward certain values and ideals. However, now that the inside-outside distinction has been more radically soaked off, not only the "self" but also "determination" in "self-determination" as well as the role that technology can play in that process can and must be reframed. How is it possible to understand "determination" in "self-determination" from the proposed interactionist perspective?

The Different Meanings of "Determination"

My reflection on the notion of "determination" in "self-determination" in this context does not aim, at least not directly, to contribute to the determinism-indeterminism debate, nor does it, as indicated repeatedly, attempt to do justice to all the subtleties of the complex and multilayered questions regarding freedom. My aim is rather to investigate how overcoming the inside-outside distinction and recognizing the technological dimension of "freedom" can reframe the question of whether free will is possible in terms of self-formation. First of all, I second Philippa Foot's

view that the use of "determinism" does not necessarily imply universal determinism as well as that causation is not necessarily deterministic in the strict sense of the word: "an action said to be determined by the desires of the man who does it is not necessarily an action for which there is supposed to be a sufficient condition" (Foot 1957, 441). Furthermore, it should be noticed that "determination" in "self-determination" is an ambiguous notion. Not only is it related to different concepts (like "causality" and "control"), but its meaning is also dependent on its subject: "determines" in "the brain determines our actions" does not have the same meaning as "determines" in "conscious thought determines our actions" or in "our environment and development determine our actions." Some of the indicated "causes" may even go back before the birth of an agent; for example, heredity causes. What complicates things further is that the meaning of "determines" is also dependent on a *particular* interpretation of its subject, of consciousness, brain and environment; for example, there are recent studies that claim that particular brain processes are indeterministic (see Balaguer 2015), which indicates that ascribing causal agency to the brain does not necessarily exclude the possibility of the existence of free will.

Without being able or needing to go into all the complexities related to neural, mental and/or societal causation and control in regard to an organism's behavior (different meanings and types of causality and causation, contesting views regarding mental states and properties, exclusion argument, nature-nurture debate etc.), I only notice (and that is sufficient for our scope) that there are many contributing causes, in different meanings of that word, involved in the determination of any decision or action, including preconscious or nonconscious determinants.

What is crucial here is that these considerations in relation to my externalist critique of the idea of a "real, deep or genuine self" as something "within" implies that we cannot isolate one factor (be it neurological inclinations, conscious intentions or other dimensions of a person and her material and social surroundings), identify it as the real and exclusive source and cause of our decisions and actions, and take it as the locus of our self. There is no detached "real self" as a *necessitating* cause behind our decisions and actions. We cannot single out one factor that determines, in the sense of "necessitates," our behavior.

"Determination" as Formation

If neither consciousness, brain or environment nor another dimension of a self can be conceived as a kind of "detached, inner source" that sufficiently determines its behavior, how then should we understand "determination" in "self-determination?" It must come as no surprise that I propose to reinterpret "determination" as "formation." Peirce offers, as I have discussed extensively in Chapter 2, a general characterization of

the conditions for self-formation (CP 8.320; see also Aydin 2009). First of all, according to Peirce, self-formation requires self-control. It is only possible to give ourselves a certain form if we have some control over our interactions. Self-control presupposes in its turn self-criticism. Self-criticism means in this context that an actor consciously reviews each of her (important) actions and compares them with certain longer-term goals and ideals that she wants to achieve. The decision to commit to an ideal and to try to realize it will eventually provoke the agent to adjust her (natural or acquired) tendencies or to form new habits.

Peirce proposes, as we have seen, a kind of reciprocal reinforcement between ideals and habit formation: the more we devote ourselves to a certain ideal, the more we will be able to form and cultivate matching habits; and the more we form and cultivate certain habits, the more we will be sensitive to the attractive power of the ideal that we wish to incorporate. However, certain habits could be also the result of inclinations and contingent interactions. Longer-term goals and ideals enable assessing and adjusting those (unquestioned) habits.

There are some similarities between Peirce's view of the self and Frankfurt's (1971) concept of a person. Frankfurt elaborates how people can have different first order (wanting something) and second order desires (desires *about* desires to do this or that). If someone not only wants to have a certain desire but also wants that desire to be effective, to be her will, she has, according to Frankfurt, a "second order volition." A person is somebody who deliberately and decisively identifies herself with a particular desire, who, in Peirce's (and my) terminology, commits herself to an ideal. Along similar lines, I believe that a person is able to discover in herself a discord between instant inclinations and higher-order volitions that are guided by longer-term wishes and goals. However, I depart from Frankfurt, as I will show later on, in my view regarding the "source" of those longer-term wishes and goals.

That we consciously can set distal goals and ideals for ourselves can hardly be contested. We take this as phenomenological evidence based on everyday life experience. It is made possible by our ability to reflect on ourselves, which is often instigated by a certain discord between how we act and how we would like to act. However, what is less evident and of vital importance for a self to have a degree of freedom, is whether conscious long-term intentions can really influence our behavior. Some opponents of free will argue that proximal conscious intentions and decisions have no causal efficacy and power because they are produced by preceding unconscious processes in the brain. Other opponents of free will claim that they have no causal efficacy and power because they are necessitated by situational influences. Although these views are controversial, I can disregard here whether and in what sense proximal intentions and decisions necessitate behavior, since they do not affect the alternative perspective that I attempt to develop. What is a relevant

question for the proposed thesis is whether the arguments used by opponents of free will also apply to distal or longer-term intentions.

Slors (2015), for one, attributes causal efficacy and control to distal conscious intentions. On the one hand, he points at the wealth of empirical evidence (see Baumeister et al. 2011) that demonstrates how consciousness can influence our actions in the long run. On the other hand, he appeals to simple everyday life experience that strongly indicate that distal intentions are efficacious: "When I form the intention, while checking the airline website and making the relevant reservation, to take the 2:15 flight to London next Wednesday, I usually find myself in that airplane on that day" (2013, 98; see also Monroe and Malle 2010 and Stillman et al. 2011). These examples resonate in everyday life intuitions that self-formation is not impossible: in everyday life there might be indeed people who are to a great extent formed by contingent influences; at the same time, we tend to make a distinction between people who are more and people who are less successful in forming themselves. In other words: self-formation seems to be conditional but not impossible.

Slors further sophisticates his argument by borrowing from Dretske (1988, 43) a distinction between two kinds of causation: "triggering causes" and "structuring causes." He argues that effective conscious intentions are indirect and should be seen as structuring causes of our actions, while our actions are possibly always triggered unconsciously by internal or external stimuli. Conscious distal intention formation is, according to him, a structuring cause, which could be conceived as a form of "self-programming": "it causes us to be 'programmed' to be responsive in specific ways to specific stimuli in specific circumstances" (2015, 106). Slor's notion of "self-programming" fits well within the Peircean framework: deliberate habit-formation could be understood as a form of programming. A steady orientation toward certain preferred goals and ideals enables us to regulate our behavior and ultimately form certain preferred habits, that is, to program ourselves.

So, we seem to have good reasons to believe that we not only can set distal goals and ideals for ourselves, but also that these goals and ideals could affect our future behavior and enable us to form a self that we would like to identify with. Longer-term goals and ideals provide orientation and prevent us from fully disintegrating in contingent interactions. They enable us to deliberately identify ourselves with some urges and inclinations that we find in ourselves and distance ourselves from others. I can, for example, set for myself the longer-term goal to live healthier. This goal can help me to regulate my behavior: it enables me to monitor and control my eating habits, for example. I will be responsive to urges (e.g., "mmm, I would kill for a hamburger") and short-term conscious decisions (e.g., "let's drive to Burger King for the last time this week") that do not fit my health program, find ways not to succumb to

these urges and intentions and gradually adjust my habits in correspondence with my long-term goal.

Now, it also becomes clearer how biological inclinations as well as conscious short-term intentions and decisions can be reinterpreted from the proposed perspective. Biological inclinations and short-term conscious intentions, which might be instigated by the brain and/or our surroundings, can influence or even greatly determine our actions, but we do not necessarily have to identify ourselves with them. By virtue of longer-term goals and ideals, we can reflect on and effectively adjust our inclinations and short-term intentions and form ourselves in a more preferred direction.

The proposed approach is sympathetic to the idea of freedom as "self-determination," but it also reframes its meaning. "An action is free if (and only if) it can be attributed to *my* (and not to an enforced or a random) will" is translated as "I can only consider an action free if (and only is) I have the possibility to identify with it and approve it." However, this "my" does not reside in an inner realm that is unaffected by external factors. From the proposed perspective, the self is not an isolated, *a priori* existing entity that "immediately" determines its decisions and actions from something "within": the self is never a completely autonomous author of its decisions and actions but is greatly heteronomous and receptive. The notion of the "mediated self" neatly expresses that there is no "direct" or original self-conception and self-experience but that the self rather "indirectly" conceives and experiences itself. What we consider our most "real" or "deep" self, including our most personal goals and ideals, is affected by a conglomerate of "external" influences; hence, the very idea of an original, unmediated self is repudiated. The notion of self-formation contributes to this view that the self also forms itself by virtue of certain identifications. The self is not an "entity" but an "identity." The extent to which this self can be considered "our real or deep" self depends on how well we impose a particular form onto our interactions by virtue of an orientation toward certain longer-term goals and ideals.

5.4 Brain Imaging Technologies, Extrospection and Self-Formation

Brain Imaging Technologies and Self-Formation

The proposed notion of the self as self-formation attempts to offer an alternative to the problematic view that the self is realized by the brain that acts as a *causa sui*. Yet, at the same time this does not exclude the possibility that brain imaging technologies could play a significant role in this process of self-formation. Brain imaging technologies can potentially explicate how our neurophysiology both sets constraints and offers

opportunities for the development of specific habits that help or fail to help us realizing our long-term ideals.

The use of brain imaging technologies breaks, as argued earlier, with the Cartesian idea that privileged, authoritative knowledge of the self can only be acquired through introspection. The proposed potential of brain imaging technologies to externally acquire even more reliable knowledge of the self seems to make them very special or unique. However, also earlier psychological techniques such as Rorschach tests or Intelligence quotient (IQ) tests were used as external sources to attain knowledge of the self. What then makes brain imaging more unique than these older techniques? More specifically, if brain scans should not be understood as direct representations of the brain, how do they differ from earlier psychological techniques to attain knowledge of the self?

First of all, it must be acknowledged that with this respect there is a continuum between brain imaging technologies and earlier psychological techniques: both types of techniques attempt to define or shed light on people's personality as well as on how people relate to themselves, thereby providing scientific discourses that offer an implicit normative foundation of the self by setting standards that constitute what is "normal" or "good" (e.g., Hacking 1986; Danziger 1994). Analogously, we can say that both types of techniques embody principles that dictate what are good processes of self-formation, and offer normative guidelines for pursued goals and ideals. Similar to Rorschach and IQ tests, brain imaging technologies can be understood as, in the words of Ian Hacking, a way to "make up people" (1986). That is, they (among other factors) can constitute, facilitate and constrain the possibilities through which we can shape and interpret ourselves.

Additionally, methodological challenges against the idea of brain scans as direct representations of the human brain/self relativize the difference between imaging technologies and, for example, Rorschach tests. We have seen that Roskies, for example, shows that brain scans should not be understood as pictures of the brain because the information needed to interpret brain scans is for the largest part external to them (2007). We have also seen that the reading of brain scans cannot be cut loose from the experimental circumstances in which several interpretational choices are already made. Klein indicates that brain regions and networks are involved in multiple processes and cannot be tied to particular functions. This suggests a variety of possible explanations for the activity that is visualized on a brain scan (Klein 2010).

Both earlier psychological techniques and brain imaging technologies seem to offer a certain view of the self, which, when uncritically accepted, competes with other possible descriptions of human behavior and self-formation. A Rorschach test conceptualizes the self in terms of different personality traits, whereas brain scans offer a view on the self that is grounded in neurophysiological differences. This continuity

between brain imaging and older technologies suggests taking a critical approach to the omnipresence of the neuroscientific discourse (e.g., Dumit 2004; Slaby 2010).

However, the fact that brain imaging technologies are not representations of the brain/self does not imply that there is no significant difference between them and older technologies. In contrast to techniques, like Rorschach tests, characteristic for brain imaging technologies is that they *foreground* physiological responses. Instead of uncritically accepting the neuroscientific narrative that brain imaging technologies offer direct representations of cognitive functioning, neurophysiological measurements can be critically related to because they allow us to establish a new physical relation to our own process of self-formation. Brain imaging technologies foreground neurophysiological processes, which allows us to make sense of – and possibly influence – how our neurology co-constitutes the habits that we develop and, in turn, enable us to orient toward our long-term ideals. This foregrounding can be seen as a new form of technological mediation that opens up new ways of self-formation, which highlight the pragmatic contribution of brain imaging technologies.

In the case of earlier psychological techniques, neurophysiological interactions were silent processes implicitly running in the background, yet of significant influence on our actions and habits. If, however, as suggested, our biological and neurological possibilities and constraints significantly influence our self-formation, foregrounding those provide the opportunity to more explicitly and deliberately relate to them. For example, this allows individuals to respond to physiological processes, and see the physiological results of their interventions in real time. This opens up the opportunity to critically reflect on the relation between the urges and habits that we are aware of, and the physiological processes that correlate with them. Making this correlation explicit allows individuals to observe that the opportunities and constraints they are confronted with can be pragmatically formed not only on a psychological, but also on a physiological level.

Extrospection as Critical Self-Identification

Current neuroscientific developments promise allowing individuals to interact with patterns of brain activity, which could contribute to form preferred or more preferred selves. Especially the "neurophenomenology" program that aims to align first-person and third-person perspectives on conscious experience proposed by Varela seems promising in this regard (Varela 1996). Neurophenomenology attempts to integrate phenomenological qualitative experiences into existing quantitative research in the neuroscience in order to (1) allow neuroscientists to better interpret neurophysiological data relevant to conscious experience (Ibid.), and (2)

open up new possibilities for individuals to relate to physical processes correlating with their conscious experiences (Lutz and Thompson 2003). Thus far, primarily the former has been discussed, focusing on how phenomenology can be used to epistemologically warrant neuroscientists to scientifically investigate conscious experience (e.g., Gallagher and Varela 2003; Lutz and Thompson 2003; Garrisson et al. 2013). In the present context, especially the latter is of interest because it takes seriously the potential to use measurements of brain activity for self-formation purposes.

The proposed view endorses the aim of neurophenomenology to move beyond a computational theory of cognition that interprets brain scans as representing how human cognition is functionally realized (Varela et al. 1991). However, the emphasis on *critical* self-formation distinguishes this proposal from the project of neurophenomenology. The aim here is not to develop a methodology grounded in phenomenology that allows for the scientific investigation of conscious experience. Instead, the leading question is how brain imaging technologies can provide knowledge that allows relating the feasibility of habitually realizing our goals and ideals in terms of our neurophysiology. This knowledge is not understood as obtained through introspection, and neither, as would be in line with neurophenomenology (Gallagher and Varela 2003; Varela 1996), as the product of a phenomenological reduction through which our habitual relation with the lifeworld is bracketed. From the proposed perspective, brain imaging technologies are interpreted as pragmatically allowing individuals to critically interact with processes that influence their self-formation that inevitably takes place in the lifeworld. Labeling this process *extrospection* indicates that this knowledge is – although potentially of direct influence on our self-formation – always mediated by both brain imaging technologies and the pursued goals and ideals. Brain imaging technologies instantiate the existence of new perceptual experiences that can be used to deliberately intervene in our self-formation (Aydin 2015).

Understanding the knowledge acquired through brain imaging technologies in terms of extrospection allows for critical reflection on whether our habits are consistent with our pursued long-term goals or ideals. By making explicit whether our habits resonate with our ideals in terms of their neurophysiological origin, brain imaging technologies offer insight in how our neurological set-up affects and is affected by the way we form ourselves. This allows for establishing an explicitly critical relation with the way our conscious and deliberate self-formation interferes with – and *is* interfered by – our neurophysiology. Self-formation is the continuous shaping and being shaped by the constraints and opportunities that the world and our biology offer. Brain imaging technologies allow establishing this critical relation by foregrounding our neurophysiology, which is an aspect of our world that was previously out of sight. Consequently, brain imaging technologies can be incorporated in the proposed

Peircean framework. When understanding ourselves in terms of self-formation, we are constantly positioned between who we currently are and who we eventually want to be. By foregrounding our neurophysiology, brain imaging technologies offer insight in how we may more effectively be able to pursue our long-term ideals, because they reveal how our habits relate to our neurophysiological make-up.

From this perspective, brain imaging technologies are not understood as tools to represent our inner realm, but are seen as ways to help us in orienting to our ideals by offering extrospective knowledge of the factors that conflict with them. This new form of knowledge is a form of *critical* self-identification, because it makes explicit the neurological constraints and possibilities that affect our behavior, which would implicitly remain of influence on our self-formation without us being aware.

Self-Formation and Neurofeedback

The proposed framework can be further clarified by applying it to *food craving*, a phenomenon closely associated with developing obesity (e.g., Weingarten and Elston 1990; Bowell and Kober 2016; Volkow et al. 2011). The neural pathways that give rise to *food craving* could display the neurological constraints and opportunities in the attempt to form ourselves. Food craving can be defined as an intense desire to consume a particular food or food type that is difficult to resist (Weingarten and Elston 1990). From the perspective of self-formation, such short-term desires function as constraints on the long-term ideal to form oneself as a healthy individual. Food craving correlates with patterns of brain activity that potentially undermines the realization of a particular long-term ideal. If these patterns are the consequence of (earlier) "bad" habits, they potentially also can be altered by developing new habits. Studies suggest that noninvasively stimulating specific regions of the brain significantly reduces food craving by modulating neuronal constraints (e.g., Goldman et al. 2011; Ljubisavljevic et al. 2016). Apparently, our neurology not only functions as a constraint on our long-term ideals, but also offers the opportunity to develop new habits by altering neural pathways. Brain imaging technologies could potentially help us to develop new habits that counterbalance patterns of brain activity that instantiate food craving.

Recently, it has been suggested that the paradigm of "neurofeedback" can have a unique function in the treatment of eating disorders because it allows individuals to voluntarily regulate localized brain activity (e.g., Bagdasaryan and Le Van Quyen 2013; Bartholdy et al. 2013). While noninvasive brain stimulation still explicitly relies on external intervention, neurofeedback is closely associated with neurophenomenology and explicitly directed at giving individuals the possibility to relate to their neurophysiology (Bagdasaryan and Le Van Quyen 2013). Respondents receive via real-time displays of EEG or fMRI feedback on the way they

respond to certain situations. Relating to these displays enables participants to actively influence their brain activity in order to attend in a more focused manner or respond more sociably using video or audio. The clinical merits of neurofeedback on food craving are still in a trial phase, but positive results of neurofeedback were found in the treatment of also other disorders. For example, the use of neurofeedback reportedly had a positive effect on the ability of children diagnosed with Attention-Deficit/Hyperactivity Disorder (ADHD) to concentrate (Gevensleben et al. 2009), the capability of individuals to control addiction-related impulses by managing dopamine levels (Sulzer et al. 2013), and helped increase the mood of people suffering from Major Depressive Disorder (MDD) by self-regulating amygdala responses (Young et al. 2014).

These studies support the proposed idea that brain imaging technologies can potentially inform processes of self-formation because they offer extrospective knowledge that allows establishing a relation with one's neurophysiology (e.g., Micoulaud-Franchi et al. 2014). They illustrate how brain imaging technologies could contribute to the formation of a self that better satisfies our ambitions and life projects. Since technologies mediate our self-identification, they can be integrated in our practices in such a way that they can help to bridge possible tensions between how we act and how we would like to act. Neurofeedback makes the discrepancy between short-term urges and long-term goals physiologically accessible, bringing to the fore a dialectic between the process of self-formation and self-identification. When being confronted with the neurological constraints that constitute our habits, at the same time we are confronted with the opportunity to actively manipulate them and turn them into ways to more effectively realize our long-term goals. It must be stressed that it is still unclear to what extent neurofeedback is effective (see also Thibault et al. 2018). Yet, although one can be agnostic about its scientific status, neurofeedback does illustrate how brain imaging technologies *could* be used in self-formation processes.

Note

1 Since I reflect on the "context of discovery" only as a means to point at "blind spots," I disregard Kuhnian challenges to the distinction between "context of discovery" and "context of justification."

6 How Critical Is Critical Self-Formation?

In the previous chapter, I argued that brain imaging technologies, instead of displaying how the self is functionally realized, allow attaining "extrospective knowledge" that can be used in processes of self-formation. The notion of "extrospection" highlights that brain imaging technologies can contribute to self-formation by foregrounding both the neurological constraints and possibilities that underlie our (unwanted) habits and allow us to directly intervene in or develop new habits in the light of desired longer-term ideals.

Building on externalist views of the self entails that the self discovers and forms itself by virtue of an orientation toward goals and ideals provided by society, its institutes and its infrastructure. An issue that I have so far left implicit is to what extent those longer-term goals and ideals could be subjected to deliberate and critical judgment. If that would be impossible, it might greatly harm the proposed approach. I have challenged the view of the self as something that is determined, in the sense of necessitated, by an entity "within" – either consciousness or the brain – by arguing that longer-term goals and ideals enable adjusting urges and short-term conscious intentions. Longer-term goals and ideals also enable, we have seen, regulating and governing our interactions with our surroundings. Now, I seem to state that those longer-term values and ideals are ultimately determined by what a society finds admirable. Have I not simply replaced one necessitating determinant with another and rendered impossible free will and, hence, deliberate self-formation? In addition, if society provides the goals and ideals by virtue of which self-formation is possible, does that not imply that anybody who had been socially formed well in, for instance, Hitler's Germany would have supported Nazism, and that the same problem occurs in less perverse cases. In short, is calling upon longer-term goals and ideals in order to soak off determining factors, variables and antecedents not ultimately a detour to yet another version of situationism?

This problem seems to become even more stringent in the context of Extended Mind Thesis (EMT) and Artifactual Mind Thesis (AMT); understanding the self as a formative process in the world runs the risk

of falling into yet another determinism, which is also expressed with respect to the appropriation of EMT by Gallagher and Crisafi (2009). They propose (as we indicated in Chapter 4) that scientific, social, legal and educational institutions can be understood in terms of an extended mind at work: "we take one of these cognitively produced things and we use it for further cognitive production. Our subsequent acts of cognition are facilitated or enhanced or made possible by particular tools or institutional mechanisms" (Gallagher and Crisafi 2009, 51). While they think that it is productive to think about cognition as distributed over humans, nonhumans and institutions, there also seems to be a downside: "the extended mind can come back to bite us; it can place limitations on our thinking, as easily as it can enable great and wonderfully extended cognitive performances" (Ibid., 51). Social and institutional forces not only offer opportunities for self-formation, but also put constraints on how self-formation can take place. They further raise the concern of "whether thinking itself, as a human enterprise, and as an individual practice, has changed (…) because of the particular means that we have invented to facilitate or enhance cognition" (Ibid., 51). From AMT, the answer to this question is definitely positive: we have seen that in a critique of EMT, AMT holds that our thinking has no original starting point in an internal, unextended biological core that spins, selects or maintains the webs of cognitive scaffolding but rather unfolds itself by virtue of and through objects, artifacts and institutions. Hence, according to AMT, thinking, and also our brain, has always been influenced by environmental factors and is increasingly affected by institutions, technologies and technological infrastructure.

In sum, how is critical self-formation possible if the pursued goals and ideals that enable that critical relation are greatly influenced by society? And how is it possible from an active externalist view to critically relate to our self-formation? Are we still free to relate to and shape our mind and self, if that self is inherently embedded in socio-institutional mechanisms and structures? How critical is critical self-formation?

6.1 Are our Longer-Term Goals and Ideals Ultimately Determined by Society?

Are our longer-term goals, values and ideals ultimately determined by what a society finds acceptable? First of all, it can hardly be denied that our goals and ideals are strongly derived from what a society finds admirable. Therefore, I second Taylor (1992) who argues that all our goals, including our most personal preferences, are greatly determined by our society, history and tradition. We are born and discover ourselves in a social environment that imparts on us certain values, goals and ideals. How and in which direction we constitute ourselves is largely dependent on what is available in the culture in which we are raised. From this

perspective, Frankfurt's "higher order desires" are also to a great extent generated by what a society finds preferable (see also Taylor 1976).

Often, there seems to be, on a basic level, no apparent reason to challenge those societal goals and ideals. If we want to live a healthier life, we often try to achieve it by pursuing goals and ideals that are generally accepted: eating less junk food, doing more exercise, not taking drugs, not smoking and so forth. These values and ideals enable us to resist internal urges and conscious intentions, which might be instigated by our biology or particular surroundings, and regulate our behavior. Technologies like brain imaging technologies can, as argued in the previous section, foreground our physiology and help more effectively realize our goals and ideals. Usually we do not challenge the legitimacy of those goals and ideals, even if we do not live by them.

Nevertheless, this does not mean that it is impossible to challenge goals and ideals that a society finds admirable. Situationists who believe that free will does not exist because our actions are determined by features of the environment of which we are not aware, or of whose influence we are not aware (see, e.g., Williams and Bargh 2008), commit a fallacy similar to neuroscientists who claim that the deterministic nature of brain processes rules out free will (see also the example of JoJo above). We have seen that, for example, brain imaging technologies can make neuronal patterns underlying tendencies and urges visible and subject to ratification, intervention or alteration. Dependent on our longer-term ambitions, those tendencies could be rejected but also endorsed. Analogously, the fact that a preference is strongly determined by societal factors does not necessarily make it enforced, and, therefore, not-mine and, therefore, unfree. Rather, the extent of freedom depends on the degree of critical identification with our conduct. Longer-term goals and ideals enable us to question immediate urges and conscious intentions. However, those longer-term goals and ideals could also be subjected to deliberation, critique and so forth. The self's ability not to completely coincide with itself and anticipate possible alternative future behavior enables deliberate self-formation. The self is, in other words, divided, and is able to relate to itself and adjust its future conduct; in the words of Peirce, the actual self can engage in a critical dialogue with "that other self that is just coming into life in the flow of time" (Peirce CP 5.421). This dialog can, as I will further elaborate below, also be conceived as a narrative that can be more or less consistent.

It must be clear that challenging societal goals and ideals will not come about spontaneously; in order to make that possible, favorable conditions need to be created. The incentive for "critical identification" cannot only come from individuals, but should also be embedded at a societal level. A healthy, open society is characterized by built-in mechanisms and tools (democratic regulations, informed consent, the possibility to opt out etc.) that enable people to deliberately and *critically*

identify themselves with certain needs and desires, including the ideals that society finds preferable. The society we live in might, for example, strongly favor both spouses working. However, there are men and women who after careful and critical deliberation reject this ideal, often by referring to "other" or "higher" goals and ideals. In their turn, the more critically acquired ideals are also subject to criticism and so *ad infinitum*. Even our most cherished goals and ideals, both on a personal and societal level, could be challenged.

I admit that there is here the potential for an endless sequence of evaluations, the so-called threat of an *infinite regress*. However, this is only a big problem if the final goal is to reach a completely independent and autonomous decision (I disregard here whether this "problem" undermines Frankfurt's view of free will and personhood). From the proposed perspective, the possibility of an endless sequence of evaluations is not a problem, but rather a positive challenge: how "real" or "deep" the self is depends on the extent and intensity of its actual engagements in identifying with ever-inclusive goals and ideals. Reaching a stage in which goals and ideals can no longer be contested is not only impossible but also undesirable (see Peirce CP 1.405, 1.173, 1.121; see also Aydin 2009). The belief in a Peircean ultimate ideal has, as we have seen, a regulative aim to enable genuine criticism and prevent absolutist claims that bring further analysis and improvement to a halt.

This remains a delicate issue: society can contribute to deliberative and self-critical thought but can also have a manipulative impact, which could render critical self-formation, and, hence, freedom, virtually impossible. In a worst case scenario, it is even possible that people who, to take up again the mentioned example, are in favor of both spouses working are manipulated by their society and people who are against it are deceived by other forces; what both groups consider their "own" choice is then, in reality, determined by factors that remain hidden for them, which would render critical identification impossible.

However, taking this objection to the extreme means returning to the assumption that there is a purely autonomous self. It is by virtue of longer-term goals and ideals that a self is able to form itself in a particular direction and is not the result of contingent biological inclinations and contingent interactions with its surroundings. Without an orientation toward longer-term goals and ideals, the self would fall apart and never reach a sense of "wholeness." So the question is not whether our decisions or actions spring from pure inner preferences or whether they are reliant on external factors but rather whether and how much room there is for critical identification. An action can be considered "really" and "deeply" mine, and hence free, if it is an expression of a critically formed self. Although the degree of critical thought that societies allow, and the measures that they take to foster it, greatly varies – and is an empirical issue – there seems to be no strong arguments that render it

impossible. Societies do often build in mechanisms that make room for other, foreign or deviant views. Hence, critical identification comes in different degrees and is at least *possible* on different levels.

6.2 Are our Cognitive and Neurological Means to Form Ourselves Determined by Institutions?

How can the posed threat to the possibility of critical self-formation be thwarted more specifically with regard to the externalist positions that I have defended, that is, AMT and the pragmatic appropriation of brain imaging technologies? How can we tackle the possibility that our very thinking has been formed by the particular institutions "that we have invented to facilitate or enhance cognition?" (Gallagher and Crisafi 2009, 51). Have EMT, and even stronger AMT, "come back to bite us?" (Ibid., 51). Can we critically relate to our self-formation, considering the external forces that influence the development of our mind and brain? This basically refers us to the question of to what extent we can form ourselves critically if we, at the same time, acknowledge that the technologies and the technological environments that are subject to such critical scrutiny and evaluation are greatly shaping the very fabric of ourselves. In the third part of this study, this question will be further investigated in terms of critical appropriation of (particular) "intrusive" technologies.

From her perspective, Catherine Malabou (2008) has articulated this threat by arguing that there is a close connection between the understanding of ourselves in terms of the plasticity of the brain and the societal demand to be flexible citizens and employees. Plasticity refers to the idea that neuronal interactions in our brain are not fixed once and for all, but can change in relation to both societal demands and our own preferences. According to Malabou, when uncritically accepted, the plasticity of our brain functions as a scientific legitimization of the societal status quo, as a neurological legitimation of societal demands. She holds that the idea of the plasticity of our brain conforms to the demand that to be flexibly capable to easily switch between different tasks, thereby denying the importance of individual variability: "In effect, anyone who is not flexible deserves to disappear" (Malabou 2008, 46).

Even if we would not fully agree with Malabou's diagnosis, it points out that it is possible that the institutional frameworks through which we are shaped render critical self-identification and self-formation impossible: how can the self critically identify itself with the very same forces and constraints that have greatly shaped it? Malabou's analysis brings to the fore that the idea that our actions, goals and ideals are predetermined by our neurophysiology can be understood as the effect of institutional forces that put constraints on our own process of self-formation.

Along the lines of the reasoning in the previous section, we have, first of all, to acknowledge that institutions in very closed and totalitarian societies could strongly engrave patterns in our brain, determine our practical thinking and render critical self-formation virtually impossible. For that reason, I have stressed that critical self-formation is not only an individual but also a societal challenge: critical self-formation entails societies building in mechanisms and tools that foster that capacity. Moreover, the plasticity of the brain not only allows society imposing its demands on it, but it also enables assertive people actively engaging with and manipulating it. It is not either-or. Malabou recognizes this "affordance" of the brain too: plasticity does not only allow the environment to shape the brain but it also makes it possible for people to form neuronal interactions relative to their own preferred ideals (2008, 11). For example, someone who desires to become a piano player could counterbalance an environment that does not encourage developing musical skills by practicing regularly on a piano available in, for example, train stations (which these days has become fashionable in the Netherlands), which would shape the neuronal and synaptic interactions in the brain through which piano skills are co-shaped. Forming these interactions is the consequence of the desire to develop oneself in this way. However, this indicates in its turn that this counterbalance, in this case making pianos available in train stations, is dependent on societal organization. Accordingly, Malabou (2008, 11) argues that the structure of society is not a mere reflection of what is going on in the brain anyway, but that it is possible to actively form and adjust brain and self, also in dissonance with societal preferences. What needs to be added to this "way out" suggested by Malabou is that it is always possible to highlight certain societal measures that allow a counterbalance, for example, making pianos available in train stations, in their turn as determinants that restrict critical self-formation. As indicated in the previous section, it must be acknowledged that critical identification comes in different degrees and is at least *possible* on different levels.

By foregrounding our neurophysiology and offering extrospective knowledge, brain imaging technologies can disclose possibilities to form the brain more effectively and critically. Instead of offering theoretical-conceptual knowledge of the plasticity of our brain – what seems to be Malabou's proposal – brain imaging technologies allow the proposed perspective of the practical-perceptual visualization of one's neurophysiology. Doing so, they can offer insight into how we more effectively can pursue our longer-term ideals, because they reveal how our habits relate to our neurophysiological make-up. Brain imaging technologies can help us to gradually make the transition from who we are and who society want us to be to who we eventually want to be; we can, for example, do something about our urges to give in to instigations from the market to consume fast food. Moreover, they can reveal that

a change in our habits is also accompanied with a neurophysiological change. This takes Malabou's narrative one step further: by foregrounding the neurophysiological plasticity of the brain, brain imaging technologies reveal that we can change the brain processes that both enable and constrain the habits that we develop.

6.3 Reinterpreting Freedom as Self-Formation

From the proposed perspective, "self-formation" is not a given but a challenge. Decisive for the "realness" or "deepness" of the self is, as I have argued, not whether its actions spring from inner preferences, and are, therefore, not reliant on external factors, but rather the degree to which that self was able to form itself in a process of critical identification with various (biological, neurological, social, institutional) influences. This alternative approach could, I believe, contribute to reinterpreting current neuropsychological findings regarding free will and overcome some of the limitations of more established approaches in free will debates. I will briefly touch on how those findings could be integrated within the proposed framework, without pretending to provide a detailed argument, which I leave to possible future research.

Own Actions and Odd Actions

If neuroscientists and social psychologists like Libet et al. (1983) and Wegner (Wegner and Wheatley 1999) claim that our actions are often determined by unconscious causes that are brought about by neurological processes, they are absolutely right. Neuroscientists, who claim that unconscious causes co-shape our "true" or "real" self, could be also right. They are, however, wrong when they claim that we usually *only* identify ourselves with our conscious and not with our unconscious part. We do not necessarily have to identify with our unconscious urges and habits, but often we do. Driving a car is for the most part something that we do unconsciously. Indeed, if we think too much about how to operate the car, it can actually hinder our ability to drive safely. However, we usually do not respond to appraisal or blame after preventing or causing a car accident with "it was not me driving the car." On other occasions, we might identify less strongly with our urges: I might respond aggressively to my children's inappropriate behavior, while not wanting to accept that I am an impatient or aggressive father. A person can identify less or more with her unconscious, biological inclinations. How "true," "real" or "deep" her self is depends on the degree of deliberate and critical identification with those urges she feels in the process of self-formation. In other words, the extent a person can be attributed "real" or "deep" agency is dependent on the degree of self-critique in the process of self-formation.

Further, we have seen that conscious longer-term intentions could indirectly influence our actions. This idea integrates well with the proposed framework. Take the example of the impatient and aggressive father: every time I respond aggressively to my children's inappropriate behavior, I might experience a certain discrepancy between how I act and how I would like to act. After repeatedly reflecting on my behavior, recurrently realizing that responding more calmly has a much better effect on my children's conduct, listening to the feedback of my wife on my behavior and trying out different techniques (from "counting to ten" to neurofeedback) in order to prevent the same mistakes, I might finally succeed in changing my behavior (recall Peirce's account of self-control, discussed earlier). The measures outlined would then have helped me to adjust my tendencies and habits, and to gradually develop a disposition that better corresponds to my ideal. It must be clear that I may not only *identify* myself with unconscious habits and tendencies that cause my actions (that is, if they fit my longer-term goals and ideals), but also could *not-identify* with conscious decisions, which I may make anyway (recall the Burger King example above).

To what extent the ideals that I pursue are really *my* ideals instead of my wife's or societies' remains a difficult question to answer in general. Excessive biological inclinations, a very dominant wife or an extremely repressive society might leave me virtually no room to critically form myself in pursuit of (ever more) preferable goals and ideals. However, in general it is possible to intervene in biology – not all wives are (equally) dominant and not all societies (equally) repressive. Moreover, without intuitions engraved in my biology, the feedback of people around me and the goals and ideals society provides me, I would not be able to acquire the critical attitude needed to form myself into a person that is not the plaything of contingent circumstances. Freedom should, therefore, not be sought outside those determinants. However, even though those determinants form who I am, there is no determinant that is singled out as *necessitating* my actions. Longer-term goals and ideals as well as technologies that could foreground different aspects of my personality (physiological or psychological) enable me to display the gap between who I am and who I would want to become and gradually work toward a preferred self.

This view of the relation between short-term and longer-term intentions sheds a different light on what I consider "my actions." I consider an action as "my action" if it fits the goals and ideals that I want to accomplish. Whether an action can be considered as "my action" does not depend on whether it can be attributed to a necessitating determinant "within" (conscious or unconscious), but to whether it fits the character traits and dispositions that others and I find typical for me, whether it reflects the longer-term goals and ideals that define me.

Of course, this view needs further refinement. Simply considering any action that conflicts with my ideals as "not mine" is too blunt. I need to

convince others and myself that I really have devoted myself to the goals and ideals that are in discord with the action in question. In general, this view is not at odds with every day lived experience: if somebody does something that is completely dissonant with what she usually personifies, questions could be and often are raised: was it really Mary? Is this the result of a misunderstanding? Did she act under pressure? Has she had a really bad day? Is she on drugs? And so on.

There might be other situations when a person continuously displays actions that seem to be in discord with her ideals. We even might call these shortcomings "in character," even if the ideals she fails to meet might ultimately be her target. This again raises questions and we often seek structural reasons: peer-pressure, weakness of will, addiction, neurological proclivities and so forth. Whether her actions are really "odd," and, therefore, could not be (completely) attributed to her, depends on the plausibility of explanations offered for this discrepancy.

We see here again a strong relation between the "realness" or "deepness" of the self and critical self-formation: whether an action is an expression of my "real" or "deep" self (and, hence, can be attributed to me) depends on the degree to which that self was able to form itself in a process of deliberate and critical identification with various influences (biological, neurological, social etc.). However, evaluating to what extent my actions are on a par with my "real" or "deep" self is not only considered in prospect, but often also in retrospect.

Successful and Unsuccessful Self-Interpretation and Self-Formation

Taylor (1992) offers an "instrument" to evaluate whether my actions are on a par with my "real" self. If there is no tension between my actions and the long-term goals and ideals that I pursue, which form to a great extent (depending on the degree of critical identification) my "real" self, I can also often manage to fit my actions into a comprehensible *story*. And I could manage to do that even if those actions were not the result of conscious decisions. I can still persuade others and myself that the actions can be given, as Taylor says (1992), a meaningful place in a coherent narrative about myself.

From this perspective, the writings of Wegner and Wheatley (1999) and others on "unconscious actions" and "confabulation" can be reinterpreted. If I intend to live a healthier life and somebody asks me why I bought "healthier" cookies, I will be tempted to give reasons that fit into my project of living a healthy life. That does not mean that choosing those cookies (at that time) was a conscious choice. If I have been successfully practicing living a healthy life for a long time, I will buy healthy food out of habit and without thinking consciously. In this instance, my choice would be the result of an unconscious cause that I explain

afterwards. So even if my unconscious choice harmonizes with my ideals, it has to be characterized as what Wegner calls "confabulation": I reconstruct a certain intention for something I did unconsciously and then pretend I consciously experienced that intention prior to the action.

This example shows that "confabulation" does not have to be sheer fabrication. It can make complete sense for others and myself why I acted so and not otherwise, even though I reconstructed a reason for my action afterwards. The question again is not whether the action immediately springs from conscious inner intentions or decisions, but rather whether it is "really" or "deeply" on a par with my longer-term goals and ideals. The more I am able to convince others and myself – from a Peircean perspective: ideally an infinite (internal and external) community of critical minds – that the consistency between the action and my longer-term goals and ideals is not accidental but the result of deliberate and critical self-formation, the more I can attribute the action to my "real" or "deep" self.

It is quite possible that my unconscious choice to go for a healthy type of cookie has another reason than the one I have provided, even though I really believe in the given reason and attempt to give it a meaningful place in a coherent narrative about myself. It is possible that the real reason was not that those cookies were healthy, but simply that they were to the right on the shelves in the store; and due to a particular brain mechanism goods that are lying to the right on a shelf are chosen more often than goods that are lying to the left (I'm referring here to Nisbett and Wilson's (1977) experiment). "Confabulation" could in that case be qualified as unsuccessful self-interpretation, as suggested by Slors (2012, 142–161). Unsuccessful self-interpretation could be a sign of self-deceit, although that is not necessarily the case.

Often, one way to find out is to gauge whether other related actions are also on a par with my pursued long-term goals and ideals. Suppose that I not only took the cookies, but without thinking also pitched into my cart the cream that I always bought in my unhealthy period, which tasted so nice on top of my cookies. If I really do not know how that cream ended up in my cart and, for example, tell others and myself afterwards that I bought it to train myself not to give in to unhealthy urges, then I am probably fooling myself. I invent reasons to make my actions fit my ideal, while the fact is that I just have not succeeded in achieving that ideal. In this instance, failed self-interpretation is also a sign of failed self-formation.

This proposed framework for self-deception suggests that I do not or do not want to recognize that my short-term intentions (unconscious or conscious) and long-term projects are not "really" or "deeply" brought into line. There may be consistency between them, but this consistency is not the result of critical and deliberate self-formation but rather of accidental circumstances (in the first "healthy cookies" case, the result

of the influence of a neurological proclivity of which I am not aware). Brain imaging technologies in the proposed framework of self-formation could help to make visible to what extent our confabulations are expressions of contingent urges or urges that are in par with our long-term goals and ideals.

As mentioned earlier, Doris presents extensive evidence showing that we often act in a rash and downright complacent manner. When the causes of an actor's cognition or behavior are not recognized by her as reasons for that cognition or behavior, were she conscious of these causes at the time of acting, Doris characterizes these causes as "defeaters." Agency is defeated because reason is overwhelmed or circumvented (Doris 2015, 52, 64f). From my perspective, what is decisive is not whether I am consciously aware of all the causes of my actions, but rather whether they can be convincingly given a meaningful place in a coherent narrative about myself and are together expressions of a deliberately and critically formed "real" or "deep" self.

In response to his findings regarding defeated agency, Doris develops another, more positive argument, which is also relevant for the approach that I propose. In his collaborativist account of agency, Doris (2015) stresses that we adopt values in collaboration with others (our intimates, peers, culture). In addition, he suggests that we do not only tell stories to each other about who we are and why we acted; these stories also exert a powerful force over us and co-determine our agency. From this perspective, our self-ignorance and liability to confabulate about our reasons for action provide an opportunity for collaborative agent-formation. In social interactions, dissonance between our attitudes and our behavior can become visible, which often prompts us to avoid situations where this dissonance could be displayed or try and bring our attitudes in line with our behavior, or vice versa.

From the proposed interactionist perspective, I can completely endorse Doris' collaborativist account of agency: our self-formation is indeed co-determined by the contributions of others to the stories that we tell. We especially achieve diachronic identity in collaboration with others (in the next part of the book I will take this even further and argue that societal and technological interaction is a necessary condition for developing not only a social, but also a singular self). Doris rightly suggests that social interaction could facilitate better reasoning and the expression of individual agential values. That is, social interaction can reveal whether a person's actions and longer-term goals are on a par and whether this is the result of deliberate and critical self-identification and self-formation; in the process of negotiation, I could discover that the coherent narrative I "made up" is the result of accidental influences (recall the "healthy cookies" example). However, there are no guarantees: social interaction could facilitate better reasoning and decrease self-deceit, but perverted, manipulative groups or societies could actually increase self-ignorance

and render critical self-formation virtually impossible. We see here that not only brain images but also the responses of other people could function as "technologies" or "techniques" that make discrepancies between urges and long-term intentions visible. What is special about brain imaging technologies is that they make visible and enable intervention into neurological patterns underlying our conduct.

Free Will and Unfree Will

From the proposed perspective, I have neither directly shown that free will exists nor that it does not exist. This question has rather been reformulated and specified by interpreting "free will" in terms of the ability to determine and subsequently form oneself. I have developed this view of "freedom" as "self-formation" in response to, on the one hand, accounts that challenge the idea that our (real) self can govern our behavior (certain brain researchers and situationists), and, on the other hand, views that attach freedom solely to internal causation and claim that only actions that originate in factors *internal* to me can be morally attributed to me.

From the proposed view, I have argued that our "(real) self" does not correlate with something "within." There is no detached inner "real self" as a *necessitating* cause behind our decisions and actions. Several factors contribute to self-determination in an intrinsic interaction: biological inclinations, conscious decisions and social circumstances. However, we cannot single out one factor that determines (in the sense of "necessitates") our behavior. I have argued that longer-term ambitions and ideals enable us to make a distinction between who we are and who we want to be, and, hence, identify ourselves with certain *determinants* and not-identify with others. From this perspective, self-determination can be understood as a process of self-formation. The self is not conceived as something that is completely determined by neurological or social or even (shorter-term) conscious influences, but rather as a relational being that attempts to regulate its behavior by virtue of extrospective knowledge acquired from technologies, longer-term goals and ideals and possible discrepancies between the two with the aim to form a more attractive self, which is a better index of freedom.

This perspective sheds a different light on free will. If there is no *a priori* existing self that coincides with something "within" (consciousness or brain), then there also cannot be an *a priori* existing "free" or "unfree" self that is completely dependent on that "within." Acting freely, in the sense of acting in such a way that *I* act, may well not have the implication that my action is the immediate result of conscious intentions; it does, however, from the proposed perspective, mean that longer-term goals and ideals allow me to work on my undesirable habits and to gradually become the self that I want to be. This view also allows us to see

that our will is not free or unfree, but rather to a greater or lesser degree free. Acquiring freedom requires practice and can succeed to varying degrees.

This perspective acknowledges that biological inclinations and short-term conscious intentions might be to a great extent instigated by the brain and/or our surroundings. However, the argument that we often act unconsciously or are greatly influenced by our surroundings by no means excludes the possibility of the existence of free will. An action can be considered as *my* action and, therefore, an expression of a free will, if it fits with the longer-term goals and ideals that I pursue and consider worthwhile and desirable. Then, I can also adequately explain why I acted in the way I did.

I have argued that freedom comes in different degrees: the degree of freedom depends on the degree of critical identification with our conduct. Longer-term goals and ideals enable us to question immediate urges and conscious intentions. However, those longer-term goals and ideals could also be subjected to deliberation and critique. Whether we have free will on "higher levels" is dependent on the degree of being able to critically endorse the goals and ideals by virtue of which we form ourselves. Freedom and unfreedom are expressions of the degree of success or failure of critical self-formation.

Moreover, "critical identification" is not only a responsibility of individuals, but also a societal challenge. Since society greatly influences the goals and ideals by virtue of which people are formed, and even the criteria those people use to evaluate whether they *ought* to identify themselves with those goals and ambitions, society has a great responsibility to set conditions that allow people to critically appropriate those goals and ideals.

From the proposed perspective, there is a strong relation between the "realness" or "deepness" of the self and critical self-formation: whether an action is an expression of my "real" or "deep" self – and, hence, of free will – depends on the degree to which that self was able to form itself in a process of deliberate and critical identification with various influences (biological, neurological, social). So the question is not whether our decisions or actions spring from inner preferences or whether they are reliant on external factors, but rather whether and how much room there was, is and will be for critical self-identification and, hence, critical self-formation.

Conclusion Part II

Is self-formation in a technological world possible? In this part of the book, I have investigated this question by setting the stage with a brief discussion of positions pro and contra the existence of freedom relevant for my theme: Kantian notions of autonomy and responsibility, situationist and brain-determinist accounts of freedom and positions that endorse the existence of free will in the context of neuroscience. I have argued that despite sophisticated and relevant differences, proponents and opponents of autonomy, freedom and responsibility often assume a false categorical distinction between an "inside" and an "outside" realm, which prevents them from overcoming an essentialist view of the self and a dualist view of (real) self and world. This dualist framework evokes, as I have shown, particular notions of "control" and "freedom," which inform and also misinform debates about autonomy, freedom and responsibility. Although participants in these debates do not agree on the question of whether consciousness/mind, the brain or the environment can be attributed causal efficacy and control, they often share, as I have shown, the view that freedom can only be attributed to an agent that is not fundamentally affected or determined by something external. I have claimed that radically dissolving the inside-outside distinction has the potential to bring about a radically different view of autonomy and freedom, and shed a different light on the question of whether self-formation is possible.

Next, I have tried to radically challenge the inside-outside fallacy that underlies dualist views by demonstrating that the mind, which was for Descartes, Kant and other essentialists and dualists the mark of the internal, has a material dimension and that matter, which was for those authors the mark of the external, has a mental and intelligible dimension. I have elaborated this view of the mind as "artifactual" (and the material as mental) in a constructive criticism of the EMT. I have indicated that EMT is a very interesting, fruitful and bold attempt to demonstrate how our cognitive system can contain external objects and technical artifacts, but that it still preserves a pivotal distinction between inside and outside in its account of the relation between the human organism and the world of external objects and artifacts, a distinction which they proclaim to

have overcome. Inspired by Charles S. Peirce's philosophy of mind, in particular, I have tried to find a way out of this "inside–outside" fallacy. External objects, artifacts or processes should, I have proposed, not be conceived as inanimate and unintelligent matter utilized by a separately living, inner mental sphere that has set certain pre-established goals for itself. Instead of being extended by an inner biological cognitive core, mind, I have argued, unfolds itself through objects and artifacts; it has an artifactual character. Mind as such is, especially in our modern technological culture, shaped by virtue of and through technical artifacts. Recognizing this artifactual dimension of mind will, I have concluded, enable a more critical analysis of contemporary claims that ascribe certain original and irreducible features to thinking. AMT displays that in both internalists' views and EMT the radical influence of technologies is rendered invisible. By depriving cognition of its original, non-derived and independent features (the "mark of the mental"), situated in a presumed inside, it becomes possible to recognize how the mechanisms attributed to it were influenced by certain technologies, but also *to make visible and evaluate* those technologies and the thinking that they have generated. AMT enables addressing the question of self-formation within a framework that does not separate an inside world from an outside world and evaluate which realm controls the other but within one that recognizes the role of technologies in developing and becoming a self.

The notion of self-formation was then further elaborated by focusing on a particular type of technologies that are increasingly shaping our view of mind, brains and self, namely brain imaging technologies. Brain imaging technologies are, as I have shown, of special importance for this study because data and images provided by brain scanners are used to draw causal relations between brain processes and cognitive functions, and to potentially resolve the question of whether the self is free or determined, and, hence, self-formation is possible. In that respect, I have developed two lines of reasoning: the first line of reasoning shows that the usage of contemporary brain imaging technologies to understand how the self is functionally realized is continuous with a problematic Cartesian framework that relies on the idea of a fully autonomous self that is detached from the world, which was inspired by the *camera obscura*. From the proposed approach, it becomes possible to recognize the relevance of brain imaging technologies without having to endorse a dualist view of mind/brain and world. As we have seen, this requires, first of all, acknowledging that brain images are not "immediate" representations or photographs of brain processes. This is not only because of methodological difficulties concerning reliability and validity, but also because fMRI images necessarily frame the brain in a particular way, making the very notion of "direct representation" untenable. The proposed approach helps revealing this particular framing. Second, the proposed approach reveals that brain images do not show that the brain is the "locus of the

self" or our "real self" but rather they mediate a particular conception of the relation between brain and self: brain images, in so far as they are taken as potential functional evidence, bring about the idea of the brain as an agent – and hence as a real or deep self – that instigates cognition and controls behavior. Third, the proposed view illustrates that brain images do not prove the existence or nonexistence of free will, but rather mediate a particular conception of freedom: by framing the brain as an isolated region, the (Cartesian) idea of freedom as an independent realm that is unaffected by external influences is sustained and rehabilitated.

The proposed framework enables putting forward a different interpretation of what brain imaging technologies display, which takes us beyond Cartesian dualisms. If technologies mediate our conception of the self as well as the self's capacity for free will, then it does not make any sense to try to acquire an unmediated, original conception of free will. We should rather acknowledge that our view of free will is and always has been influenced by the technologies that we employ to identify it. Free will understood as "the capacity of an agent to determine itself and not be determined by someone or something else" disregards the fact that the medium (in our case, brain imaging technologies) that is expected to establish whether "free will" exists (i.e., whether we are able to determine ourselves) co-determines how we conceive "free will." However, brain images cannot provide us with direct access to a "within" – the brain – that is not affected by a "without" because our very notion of that "within" is co-determined by external influences. The brain as an independent realm that is unaffected by external influences can only be validated in the context of neuroscience by making the external, technical conditions that have generated it invisible.

The second line of reasoning displays a discontinuity between the use of brain imaging technologies and the Cartesian framework. We have seen that brain imaging technologies challenge the idea that introspection is a source of privileged, authoritative knowledge of the self. I have proposed a pragmatic understanding of brain imaging technologies. This approach shares with enactivist and neurophenomenological perspectives the idea that the self is not an identifiable entity, and that both subjective experiences and scientific insights on how aspects of the self are realized should be considered (Gallagher and Varela 2003; Gallagher 2007). In this respect, I brought to the fore the relevance of the knowledge obtained through brain imaging technologies for the experiencing subject. When being confronted with previously unobservable neurophysiological processes, the possibility to critically identify with these opens up. Brain imaging technologies can confront individuals with how their habits are co-constituted by neurophysiological processes. This, in turn, constitutes the possibility to develop new ways of self-formation.

From this perspective, brain images are not used to represent the "real" determinants of a person's behavior and show that she cannot control

her actions and is a powerless spectator, but rather practically confront the concerned person with her actions and enable her to change them in such a way that they correspond better with the way she would like to act. Neurofeedback was used as an example to show how current neuroscientific developments could be integrated in the proposed framework: using neurofeedback, individuals can critically reflect on whether their habits are consistent with the long-term goals or ideals they orient to.

This is where the two different lines of reasoning converge. Instead of debating brain imaging technologies in terms of their *representational* qualities, from the proposed framework they are understood in terms of how they mediate new opportunities for self-formation by *making present* how our neurophysiology co-constitutes our habits. By attaining extrospective knowledge of our physiology, brain imaging technologies potentially allow for intervening in the habits we form when pursuing our existing long-term ideals. Understood in this way, brain imaging technologies contribute to self-formation by making explicit the neurological constraints and possibilities that coincide with our (unwanted) habits, without relying on a Cartesian, representationalist framework.

In the last chapter of this part of the book, I discussed an important caveat. I challenged the view of the self as something that is determined, in the sense of necessitated, by an entity "within" – either consciousness or the brain – by arguing that longer-term goals and ideals enable us to not identify with and adjust our urges and short-term conscious intentions. However, I have also stated that those longer-term values and ideals are ultimately determined by what a society finds admirable. Have I not simply replaced one necessitating determinant with another and rendered impossible free will and, hence, deliberate self-formation? From EMT to an even stronger AMT, this problem becomes even more stringent, since it holds that our thinking unfolds itself by virtue of and through objects, artifacts and institutions, which implies that our thinking as well as our brain become even more dependent on technologies and institutions. In addition, if the structures or brain can be changed according to particular goals and ideals, ultimately derived from society, using brain imaging technologies, what does then guarantee that we are not a plaything of contingent societal systems and ideologies?

I responded to this difficulty by, first of all, recognizing that our goals and ideals are indeed strongly derived from what a society finds admirable. However, this, as I have contended, does not make it impossible to challenge societal goals and ideals. Facilitating this possibility requires creating favorable conditions, which cannot only be done by individuals but should also be embedded at a societal level. I have argued that critical identification comes in different degrees and is at least *possible* on different levels. The plasticity of the brain shows that not only the environment can shape the brain, but that it is also possible to form neuronal interactions relative to goals and ideals that diverge from particular

societal demands. Since the structure of society is not a mere reflection of what is going on in the brain anyway, it is possible to actively form and adjust it, also in dissonance with societal preferences.

This approach reframes free will. From the proposed perspective, acting freely does not have the implication that my action is the immediate result of conscious intentions. However, it does indicate that critical self-identification allows me to work on my undesirable habits and to gradually become the self that I want to be, which might be a better indication of freedom. From this perspective, the self is not an isolated, *a priori* existing entity that is "immediately" determined by something "within" – consciousness or brain – but rather continuously discovers and forms itself by virtue of its identifications and non-identifications with its various (biological, neurological, social, institutional) determinants. The self is always a "self in the making."

Part III

How Should We Technologically Form Ourselves?

After having elaborated why the self should be conceived as "self-formation" (part I) and in what sense self-formation is possible (part II), the question that arises now is: how then ought we to form ourselves? In which direction should we shape ourselves in a culture that is saturated with technology? And if technologies can contribute to self-formation, how then can we technologically form ourselves?

In this part of the book, I will investigate the question of how to form ourselves by first analyzing if technological self-formation is conceived best as technological "self-enhancement." Developments in nanotechnology, biotechnology, information technologies and cognitive science (NBIC; and more recently and specifically artificial intelligence) have fueled high hopes and expectations. Radical techno-optimists believe that these disciplines will converge (hence, NBIC-convergence) and radically transform and improve society and humans: discoveries in one field will serve for research in another and generate a synergy that will multiply the power of research and enable spectacular advances. Some advocates, labeling themselves as trans- or posthumanists (lately, additional brands and flavors have been introduced), forcefully defend harnessing this knowledge for radical enhancement of human capacities and functions. In the words of one of the transhumanists' spokesmen:

> Transhumanists view human nature as a work-in-progress, a half baked beginning that we can learn to remold in desirable ways. Current humanity need not be the endpoint of evolution. Transhumanists hope that by responsible use of science, technology, and other rational means, we shall eventually manage to become posthuman, beings with vastly greater capacities than present human beings have (Bostrom 2003b)

Transhumanists and other techno-optimists believe that after liberating ourselves from dogmatic views that have attempted fixing our nature, we can become masters of our own fate. Now that we have left all metaphysic and religious systems that attributed to us a static nature or an essence, we are able to form ourselves as we prefer. And if we are able to

form ourselves as we prefer, then why not boundlessly enhance ourselves? Instead of education through oral and textual means, now technologies enable us to enhance ourselves much faster and better. High-Tech bio-interventions (cloning, genetic alterations) and cutting-edge technologies and artifacts that can be internally or externally connected to different human faculties (auditory, visual and other bodily prosthetics, neural implants) will allow us, transhumanists believe, to improve ourselves in ways that go beyond anything we can imagine: the sky is the limit! From this perspective, technological self-formation is understood as techno-logical "self-enhancement," and technological "self-enhancement" is explained in terms of better, stronger and smarter.

Building on the developed perspective in this study, I will demonstrate that conceiving "self-formation" as "self-enhancement" is questionable, inadequate and untenable for two reasons. First, I will argue that trans-humanists are confused about their own conception of the posthuman: although they anticipate radical transformation of the human through technology and propose overcoming essentialist views of the human, at the same time they assume that the criteria to determine what is "nor-mal" and what is "enhanced" are univocal, both in our present time and in the future. This way they sustain the very essentialist view that they want to overcome. Since some transhumanists believe that there are strong similarities between the ideal posthuman that they propose and Nietzsche's Overhuman, I will develop my critique of transhumanists' notion of enhancement on the basis of a comparison between the two "ideals." Inspired by Nietzsche's notion of the Overhuman, I will argue that overcoming an essentialist conception of the "ideal human" requires copious variations of criteria, both diachronic and synchronic, of what can be considered "normal" and "enhanced." Paradoxically, the radi-cal and disruptive technologies on which transhumanists have pinned their hopes make sustaining a uniform criterion for enhancement un-tenable: new and emerging technologies are, as I have argued repeatedly throughout this study, not neutral means but often bring about different and, from our current perspective, foreign standards for determining what are "normal" and "enhanced" capacities.

Next, I will argue that the idea that technologies bring about new and foreign standards for what is considered as "enhancement" and affect our deepest and most cherished norms, values and ideals abolishes the assumption that we are masters and commanders of our own process of self-formation, which also affects our relation to ourselves. Hook-ing into the uncanny valley debate, I will interpret the uncanny feeling not so much as a response to a lack of humanness in human-like ro-bots, but rather as a response to the inability to fathom and appropriate what makes me "me": the alterity in human-like robots confronts me with some-thing within that also remains strange to me, which, at least partly, could explain why in our attempts to form ourselves (through

education or technologically) we come across some-thing within that resists complete appropriation, something "uncanny" that we cannot mold as we prefer. Contra to what many transhumanists claim, what is formed is not a "patient" that we can completely control and mold as we please. Acknowledging that the technological uncanny is ever-stronger, becoming a permanent structure of selfhood, signifies that technology cannot simply be externalized and conceived as an outside factor that can determine or liberate us, nor as something that can destroy or strengthen us. Employing Lacan's notion of "extimacy" and Nancy's view of being "closed open" for intrusive technologies, I will illustrate that technology as "other within" is increasingly becoming both the necessary condition for forming a stable social self and an obstacle that prevents it from reaching the singularity that it seeks. Acknowledging this view of the self being inherently compromised renders an ethics of enhancement unintelligible and calls for an alternative and much more nuanced view of self-formation.

In the remaining two chapters of this part of the book, I will provide this alternative, more nuanced view: instead of understanding technological self-formation in terms of making ourselves "faster," "stronger" and "smarter," we should, I will propose, conceive it in terms of technological sublimation, which will also redefine, as I will show, the notion of "human enhancement." The notion of sublimation will be developed through a constructive engagement with respectively Freud, Lacan, Nietzsche and Peirce, which will culminate in what I will call Technological Sublimation Theory (TST). This theory, as I will frame it, recognizes on the one hand that there are no invariable, uniform and independent criteria to establish what is "enhancement" and on the other hand resists embracing an "anything goes" view of self-formation, putting forward certain guidelines for what is "good self-formation." In the last chapter of this part, I will apply the developed theory to three technological fields, namely smart technological environments, brain imaging technologies and smart drugs.

7   Technological Self-Formation as Enhancement

Transhumanists believe that technology could transform the human into a so-called posthuman, a technologically enhanced future ideal being. Recent contributions to the human enhancement debate have wondered whether Nietzsche would be an advocate of transhumanism and whether his notion of the Overhuman (*Übermensch*) would strengthen transhumanists' notion of the posthuman.[1] Some authors claim that his philosophy fits well in the transhumanists' mode of thought. Although Nietzsche focuses on education, Sorgner (2009) holds that he would also have encouraged technological means to realize the Overhuman, since both procedures are, he believes, structurally analogous. More (2010) emphasizes the importance that both Nietzsche and transhumanists grant to critical thinking, scientific inquiry and self-transformation. He even states – referring to himself – that Nietzsche has directly influenced the founders of the transhumanist movement.

Other authors find it difficult to ally Nietzsche with transhumanist thought. According to Bostrom, transhumanism has, despite "some surface level similarities with the Nietzschean vision," as much or more in common with humanism, the Enlightenment and liberalism (Bostrom 2005, 4). Hauskeller (2010) believes that reconciling Nietzsche with transhumanism is impossible since Nietzsche has, he claims, contempt for all core humanist values and aspirations. Babich (2011) argues that transhumanism is but one more expression of the ascetic ideal, which is a life-negating belief that Nietzsche has contested throughout his writings.

Still, other authors relate the transhumanist movement to Nietzschean aspirations but reject both as unacceptable ideologies. Habermas, who rejects every type of genetic enhancement, identifies transhumanists as "self-styled Nietzscheans," who are "intoxicated by science fiction" and breeding fantasies and believes that they only succeeded in "staging a media spectacle" (Habermas 2003, 22).

My aim here is neither to settle the (anachronistic) question of whether Nietzsche is a transhumanist nor is it Nietzsche-exegesis for its own sake. Instead, I will use Nietzsche's figure of the Overhuman to analyze and evaluate the transhumanists' posthuman.[2] This investigation will reveal that transhumanists are confused about their own conception of

the posthuman: transhumanists anticipate radical transformation of the human being through technology and at the same time assume that the criteria to determine what is "normal" and what is "enhanced" are univocal, both in our present time and in the future. Inspired by Nietzsche's notion of the Overhuman, I will argue that the slightest historical and phenomenological sense undermines this essentialist perspective and discloses copious variations of criteria, both diachronic and synchronic, for what can be considered "normal" and "enhanced."

Radical transformation through technology does not simply enable us to become "stronger," "smarter" or "healthier," but it can and often will also change the very standard or yardstick with which we measure what counts as "stronger," "smarter" or "healthier." Put yet differently: new and emerging technologies are not neutral means but could bring about different and, from our current perspective, foreign standards for determining what are "normal" and "enhanced" capacities. Since the qualitative meanings of these terms are themselves not fixed, it is unintelligible and too reassuring simply to predict that new technologies will enhance human beings. I need to stress that my critique is directed at the conception of "enhancement" in transhumanism and affects affiliated approaches such as posthumanism or metahumanism (Ferrando 2013; Deretić and Sorgner 2016), insofar as they share or have not overcome the transhumanist's view of "enhancement."

7.1 Transhumanists and Their Ideal Human

Transhumanism

Although there are significant differences in preference, nuance and radicalness, most influential transhumanists (More 1990; Savulescu 2001; Stock 2002; Bostrom 2003a; Harris 2007) share a common ground. First of all, transhumanists believe that we should or even ought to enhance ourselves and/or humanity as such, which will require overcoming current limitations of the bodily or mental constitution of the human being (Parens 1998; Naam 2004; Garreau 2005; Wilsdon and Miller 2006). Transhumanists' examples of enhancement of physical capacities are increasing strength and stamina, being more resistant to illnesses and defeating aging. Mental enhancements include improvements of sensory perception, memory, imagination, mood and self-esteem. In contrast to therapy, Bostrom and Roache claim, "enhancement interventions aim to improve the state of an organism beyond its normal healthy state" (Bostrom and Roache 2008, 120).

Second, leading transhumanists think that transhumanism can be viewed as an extension of humanism, from which it is partially derived. According to More (1990), "[t]ranshumanism shares many elements of humanism, including a respect for reason and science, a commitment to progress, and a valuing of human (or transhuman) existence in this life

rather than in some supernatural 'afterlife.'" Transhumanists promote in compliance with humanism, says Bostrom (2003b, 4), "rational thinking, freedom, tolerance, democracy, and concern for our fellow human beings." Transhumanists do not abandon or leave humanism behind but attempt to realize humanist values by other means.

Third, transhumanists believe that in improving ourselves we are not limited to traditional humanist methods, such as discipline and education through oral or textual means. Rather, we can also use and should make widely available technologies that will radically modify and enhance our human nature. In the words of More:

> Transhumanism differs from humanism in recognizing and anticipating the radical alterations in the nature and possibilities of our lives resulting from various sciences and technologies such as neuroscience and neuropharmacology, life extension, nanotechnology, artificial ultraintelligence, and space habitation, combined with a rational philosophy and value system (More 1990)

Other technologies that transhumanists focus on are genetic engineering, prosthetics and powered exoskeletons and tissue engineering (Brey 2008 provides a categorization).

Fourth, many transhumanists expect that exponential technological progress will eventually enable us to move beyond what some would think of as "human." Quoting More (1990) again: "Transhumanism is a class of philosophies that seek to guide us towards a posthuman condition." According to Bostrom (2013), the idea of the posthuman being is based on the premise "that the human species in its current form does not represent the end of our development but rather a comparatively early phase." Some transhumanists also claim that so-called existential risks (global pandemic, comet impact, misuse of nanotechnology) compel us to evolve technologically, for the sake of our survival, into new and better humans (Bostrom 2002, 2013).

Transhumanists' Ideal Human

To acquire deeper insight into the transhumanists' ideal or more ideal human being requires further understanding of what they mean by "enhancement." "Enhancement" is, as indicated above, often contrasted to therapy. "Therapy" is understood as repairing a bodily or mental defect, whereas "enhancement" is conceived as improving a healthy or normal bodily or mental constitution. "Therapy" aims at a transition from less normal to normal, whereas "enhancement" implies transition from normal to better than normal. A "transition from normal to better than normal" can have at least two different meanings: (1) improving qualities in some individuals that already exist in average people; (2) introducing improvements that no human being has yet displayed (Brey 2008).[3]

These distinctions allow the categorization of human beings into four different, less and more ideal, types: the disabled, the enabled, the over-abled, and the differently-abled. The disabled lack certain bodily or mental functions, qualities, traits or capacities, which can be found in average people (e.g., being blind, missing a limb, having a memory deficit). The enabled possess average people's functions, qualities, traits or capacities (having an average intelligence quotient (IQ), living approximately 80 years, being able to appreciate good music). The over-abled have functions, qualities, traits or capacities that can be found in average people, but are much stronger, fine-tuned or further developed (e.g., being able to run 100 km per hour, playing a very difficult instrument with virtuosity, processing and remembering huge amounts of information). Finally, the differently-abled are characterized by functions, qualities, traits or capacities that no current human beings possess and that grant them an advantage over current human beings (X-ray vision, the ability to fly, seeing things in four dimensions and other capacities unimaginable for us).

Transhumanists mostly focus on the last two types: the over-abled and the differently-abled. In *Why I Want to be a Posthuman When I Grow Up*, Bostrom (2008) describes the process of becoming a posthuman being. The improvements that Bostrom mentions indicate a continuity between what we already are or have and what we eventually might become or realize:

> You have just celebrated your 170th birthday and you feel stronger than ever. Each day is a joy. You have invented entirely new art forms, which exploit the new kinds of cognitive capacities and sensibilities you have developed. You still listen to music – music that is to Mozart what Mozart is to bad Muzak. You are communicating with your contemporaries using a language that has grown out of English over the past century and that has a vocabulary and expressive power that enables you to share and discuss thoughts and feelings that unaugmented humans could not even think or experience (Bostrom 2008, 5)

In sum, we will still have birthdays, many birthdays. Strength and joy will still be qualities that we find important. We will still appreciate music and have developed a highly sophisticated receptivity for it. We will still use language to communicate with one another, albeit this language will enable a much more refined expression of thoughts and feelings.

Although Bostrom does recognize that our ability to imagine posthuman life is very limited, he believes that it is not completely impossible to imagine how posthumanism will emerge:

> As we seek to peer farther [*sic*] into post-humanity, our ability to concretely imagine what it might be like trails off … Yet we can at least perceive the outlines of some of the nearer shores of posthumanity, as we did in the imaginary scenario above (Bostrom 2008, 5f)

That there is a continuity between the over-abled and differently-abled is further stressed by Bostrom's conviction that a person does not have to change substantially in his transformation into a posthuman being:

> A person could obtain considerable increased life expectancy, intelligence, health, memory, and emotional sensitivity, without ceasing to exist in the process … these developments are not viewed as spelling the end of the original person … If most of what someone currently is, including her most important memories, activities, and feelings, is preserved, then adding extra capacities on top of that would not easily cause the person to cease to exist (Bostrom 2003a, 499)

Bostrom (2008, 16) compares the transformation into a posthuman being to migration, career change or religious conversion, and believes that we should not use unattainable criteria for technological self-transformation; in addition, he states that "[a] posthuman mode of being is one that includes at least one posthuman capacity" (Bostrom 2003b, 3). We will radically change and have better, other or more capacities but could "still embrace most traditional values and principles of personal conduct" (Bostrom 2003b, 7).

This view is shared by most transhumanists. Hans Moravec characterizes the posthuman robot as our evolutionary heir: "Intelligent machines, which will grow from us, learn our skills, and share our goals and values, can be viewed as children of our minds" (Moravec 1999, 126). Even Kurzweil's (2005) singularity, which marks a radical shift or conversion in the history of mankind, presumes continuity between the over-abled and differently-abled. Exponential growth of technological know-how and accumulation of technological (ultra)intelligence underpin the singularity. The resulting ultra-intelligent life forms, however, will share our moral values: "the nonbiological intelligence will be embedded in our society and will reflect our values" (Kurzweil 2005, 424).

7.2 Nietzsche's Overhuman

The "Over" in "Overhuman"

One of the first questions that Nietzsche's notion of the *Übermensch* raises for English writers is: how should one translate this word? In some translations *Übermensch* is translated as "Superman." I prefer "Overhuman" to "Superman" for two reasons: first, it preserves the allusions and word play Nietzsche intends with his recurrent references to going under (going down, downfall) and going over (overcoming, transform). Second, "Superman" conjures up images of the comic book and motion picture characters that possess supernatural powers (Superman, Spiderman, Hulk etc.). This association draws attention to certain extraordinary

capacities and functions and passes over the *über* in *Übermensch*, which is, I believe, of eminent importance for adequately understanding Nietzsche's proposed "ideal."

The "over" in "Overhuman" designates detachment, above all. Although the Nietzsche literature provides divergent answers to what exactly Nietzsche wants to overcome, a persistent motive that can be detected throughout his work is that he wants to overcome the Platonic-Christian worldview (Müller-Lauter 1971; van Tongeren 1989; Aydin 2003). According to Nietzsche, Christianity has effectively rendered one possible interpretation of the human being absolute by subjecting her[4] to an unchanging moral value system that is sanctioned by an omnipotent and eternal God. The seeds for this ideology he traces back to Plato's Idea of the Good. That is why Nietzsche can say: "Christianity is Platonism for the "people"" (BG Preface). Plato's Idea of the Good forms the bedrock of the belief that the human being has an invariable and ahistorical essence, which should be cultivated according to certain universal values and norms.

Here, I have to emphasize the crucial role that power relations and struggle play in Nietzsche's worldview. His concept of "will to power" is, as I have extensively elaborated in part I of this study, characterized by intrinsic relationality: power is only power in relation to another power. Moreover, power is a necessary striving to expand itself. Power is only power insofar as it can maintain itself against other powers and strives to predominate over them. Since the world is "'will to power' and nothing else" (BG 36), nothing has existence and meaning outside the "game" of power relations, and existence and meaning can only emerge within the "game" of power relations. Because of this, there is also no possibility to withdraw from this "game." Even rejecting the claim that reality is will to power is an expression of will to power. Also, setting a standard or criterion in order to establish or measure what is "normal" and what is "enhanced" is nothing else than an expression of will to power, which always can be questioned by other wills to power. Weakening or desisting struggle by absolutizing one criterion or standard that once and for all fixes what is good, enhanced or ideal (and, hence, not acknowledging other possible standards) is for Nietzsche "life threatening": it blocks the emergence of (new) life forms and (new) meanings (I will come back to this later on). This is exactly, Nietzsche believes, what the Platonic-Christian ideology is doing.

From this perspective, it is not very surprising that the features that Nietzsche attributes to the Overhuman stand in opposition to those that he identifies with the Platonic-Christian ideal. Against the Platonic-Christian desire for equality and a common value system, Nietzsche (SZ II, 78) endorses a world marked by inequalities and struggles. For him, the idea that all human beings are equal (in the sense of "similar") as *human beings* and have, therefore, equal rights, values and duties finds its source

in the belief that human beings have a common essence that is secured by some-thing or someone (for Nietzsche that is ultimately "God") that inscribes that essence in their nature. Human beings, according to Nietzsche, do not share a common essence. They should, therefore, not strive for similar worldviews and value systems but attempt to develop different life forms by continuously challenging the status quo.

Also, unlike the Platonic-Christian influenced masses who desire dependence and life within an inclusive herd, Nietzsche's Overhuman stands in splendid solitude (SZ I, 46). According to Nietzsche, the potential to establish radical new ways of living can only come from individuals who are not completely absorbed and exhausted by society and its *Sittlichkeit der Sitte* and have enough distance and power to challenge it (SZ I, 36; Aydin 2008).

In contrast to the Platonic-Christian desire to dwell on the eternal and dismiss the earthly body and way of life, Nietzsche's Overhuman loves the body and its senses.[5] In contrast to the soul, the body is not a unity but a multiplicity, and does not exhibit eternal forms but change and fluidity (SZ I, 23f). Nietzsche diagnoses views that want to escape the here and now as symptoms of not being able to cope with the fact that life has no *a priori* universal meaning and goal. The body, he believes, could make us see and appreciate the multiplicity of life forms and the necessity to create meaning and establish new goals. In Nietzsche's words: "Along the guiding thread of the body…we learn that our life is possible through an interplay of many intelligences that are very unequal in value, and thus only through a constant, thousand-fold obeying and commanding" (WN, 30, 77).

At first sight, the process of secularization seems to pave the road for the Overhuman: "God died: now *we* want – the overman to live" (SZ IV, 232). In different passages, Nietzsche welcomes this development as liberation. Equality he believes, for example, can no longer be justified: "'we are all equal, human is human, before God – we are all equal!' Before God! – Now, however, this God has died…this god was your greatest danger" (SZ IV, 232). Since God has functioned as an orientation point for how human beings understood themselves and lived their lives, with the death of God also the "human being" "dies": human beings become a problem for themselves, that is, how to define and understand "human nature" becomes a philosophical challenge. For Nietzsche, this "problem" or "challenge" brings new opportunities. The death of God renders an absolute criterion obsolete, enticing the human being to create her own values and, hence, be her own God (SZ I, 47). The death of God could liberate human beings from a uniform value system and open up a sea of interpretations and experiments, which could instigate radically new world(view)s and new identities.

Unfortunately, the modern, secularized human being, Nietzsche believes, has not overcome the Platonic-Christian ideal. The religious

"surface" has been peeled off, but the Platonic-Christian value system is preserved. The same normalizing and homogenizing tendencies can be detected in our modern culture. In democratic societies, for example, qualitative differences are considered to be bridgeable and are reduced to quantitative differences. The law has taken the place of God: "the democratic instincts of the modern soul! Everywhere, equality before the law" (BG 22); "the *democratic* movement is the heir to Christianity" (BG 202). The herd mentality is also reflected in socialist movements (BG 203). In modern philosophy and science, the multiplicity that we find in the world is disregarded by assuming that reason can capture pure objectivity and find absolute truth, which is for Nietzsche nothing else than yet another expression of the Platonic-Christian disposition (GS Preface 3).

The process of modernization, which appears to have the potential to subvert the idea of an absolute criterion for determining what a human being is or ought to be, has not, in Nietzsche's opinion, been potentiated. For Nietzsche, democracy, liberalism and also *humanism* are, among many other modern ideologies, all guises of the tendency of the modern human being to canonize and eternalize her present state after the death of God. The modern, subject-centered rational individual finds universal criteria to determine what is true and right in an "inner-world," which increasingly makes "The Great Lawmaker" redundant. Secularization, for Nietzsche, even becomes an additional threat. There is nothing anymore that can relativize the belief system of the modern individual, not even a God. The human being finds the measure for understanding and determining her identity and her values as well as the world in herself. He becomes "the measure of all things." The human being has no other objective and destiny outside herself. The modern, Western individual imprisons herself in the self-glorification and self-deification of her present state. Other possible life forms are, hence, excluded. Nietzsche detects this tendency even in evolutionist perspectives, where the leading idea is not self-overcoming but self-preservation (SZ IV, 232; WN, 134f, 257, 264).

This discovery is dramatically portrayed in *Also sprach Zarathustra*. After failing to convey his optimistic message of the Overhuman to the masses, Zarathustra attempts a different strategy: he warns them of what will happen if they are not willing to make the necessary preparations for the advent of the Overhuman. He tries to scare them by speaking to them of the "most contemptible person," the "last human," who is the Overhuman's antonym. The "last human" symbolizes not the will to overcome oneself, but rather the will to preserve oneself. It symbolizes human beings who, in Nietzsche's words, "no longer launch the arrow of their longing beyond the human" (SZ I, 9). To his surprise, Zarathustra discovers that his audience is not repulsed by the presented opposite scenario, but rather recognizes itself in the last human (SZ I, 10). The

modern, secularized individual is completely content with herself. The last thing that this individual wants is self-overcoming: "Each wants the same, each is the same, and who-ever feels differently goes voluntarily into the insane asylum" (SZ: 10). Redemption from the acquiescence of invariability ("nihilism" is another word Nietzsche uses for this state of mind) seems to have become impossible.

The "Over" as an Index of Transcendence

The Overhuman is a name for something that should liberate us from both the Platonic-Christian heritage and the modernist path on which we have embarked. It promises to overcome the idea that an external measure determines our essence and values as well as the view that we are the ones who determine this measure and, hence, our essence and values. The opposition between the Overhuman and the last human clearly shows that Nietzsche's concern is not only an episode in our history (Christianity). The "last human" rather expresses the persistent ("human, all too human") tendency to resign to a presumed invariable identity, whereas the Overhuman promises to undermine and overcome every possible invariability and sameness. This explains why Nietzsche can lump Platonism, Christianity and humanism, which at first sight are very different ideologies, together. We will use Nietzsche's notion of the Overhuman to assess whether this tendency to assume univocal criteria to determine what is an "enhanced" or an "ideal human" can also be detected in transhumanist views. The "humanist" dimension in trans-humanism relates, therefore, to this essentialist tendency. Of course, this Nietzschean assessment might not apply to other possible types of "humanism."

Sometimes Nietzsche seems to give concrete examples of Overhumans: Napoleon, for example, is characterized as a "synthesis of *Unmensch* [brute] and *Übermensch*" (GM I, 16). Cesare Borgia, another example, is called "a type of overman" (TI, 211). These identifications are, how-ever, always relative and oppositional. Napoleon and Borgia represent "a type that is an overman in comparison" (EH, 147): they are Over-humans "in contrast to 'modern' people, to 'good' people, to Christians and other nihilists" (EH, 101). In *Also sprach Zarathustra*, the ques-tion of whether there have been Overhumans is raised a couple of times. Zarathustra leaves no room for speculation: "Never yet has there been an overman" (SZ II, 71).

In *Also sprach Zarathustra*, this use of the opposite as a means to signify an Overhuman takes on many forms, as humans fail again and again to meet Zarathustra's expectations, and the gap between all the types of humans that we can imagine and the Overhuman becomes ulti-mately unbridgeable. The Overhuman is portrayed as anti-ideal par ex-cellence because it departs furthest from everything that we can imagine

and comes nearest to the radically different. Even the higher human beings who become Zarathustra's "disciples" and demonstrate characteristics that stand in opposition to Christianity and the democratic, secularized ideal are not fully identifiable as Overhuman. The last pope, for example, who was the incarnation of Christ himself, broke with God and claims that he is even more Godless than Zarathustra (SZ IV, 209f). Although these figures are Overhumans by comparison and contrast, they cannot be positively identified as Overhumans. If that were possible, the Overhuman would be yet another ideal that attempts to fix the human in a particular life form and value system, which is, as Nietzsche indicates, characteristic of *all worldviews so far*: "But have you ever asked yourselves properly how costly the setting up of *every* ideal on earth has been? How much reality always had to be vilified and misunderstood in the process" (GM II, 24). Also, the higher human beings ultimately do not and cannot live up to Zarathustra's standards (SZ IV, 264f).

The Overhuman not only has never existed, but also will never exist, as something particular. It also can never be a human being with extraordinary characteristics, a kind of Superman, and certainly not an ideal human as might be identified according to modern ideology. The Overhuman is portrayed in different ways as something that goes beyond every possible idea of a particular human. Every characterization of the Overhuman remains a negative indication. It lies, therefore, beyond every imaginable measure, which makes speaking of the Overhuman a highly paradoxical enterprise. By its very nature, the Overhuman cannot be conceptualized or realized. In fact, everything that is said about it is too much.

The Overhuman is a name for Nietzsche's conviction that the human being is not an end point but an in-between-being: "What is great about human beings is that they are a bridge and not a purpose: what is lovable about human beings is that they are a crossing over and a going under" (SZ I, 7). The prefix *über* in *Übermensch* indicates that the human being is an orientation without a beginning in a primary cause and an end in a final goal. Again, what is crucial is the *über*. It transcends every characterization. The Overhuman is a "transcending indication."

Nietzsche's notion of transcendence should not, however, be confused with traditional religious notions. It does not refer to an afterlife but tries to depict a fundamental dimension of our very earthly and bodily existence. This view is also expressed in Nietzsche's definition of the human being as the "not yet determined animal." This definition indicates that the human being "essentially" does not coincide with its actual state. By overcoming their present state, human beings do not alienate themselves from their nature but do justice to what makes them human: becoming (Aydin 2007b). The human would cease to be human (and entirely reconcile itself with its animal part) if it were to coincide completely with its actual state. For Nietzsche, new technological developments do not

represent the first time that humans have been faced with the prospect of overcoming their actual state; ideally, this potential is the human being's permanent condition.

From Nietzsche's perspective, human life can only recognize and justify itself by pursuing something that in no way can be reduced to but rather goes beyond its present state, even after the death of God. In Nietzsche's words: "It is not enough that you reconcile yourself with the one you kill [namely, yourself]. Let your sadness be love for the overman – thus you justify that you still live!" (SZ I, 26). The human being *is* a longing for something beyond herself, an "arrow and longing for the overman" (SZ I, 41). In something that she is not yet, the human being finds a justification for what she is and must become, which is also the only way to give meaning to her life.

7.3 Sounding Out Hollow Idols

The figure of the Overhuman challenges the essentialist idea of one possible interpretation of the human, which denies and destroys the potential richness of human life (EH, 155). In Nietzsche's idiom, this assessment can be characterized as "sounding out of hollow idols," which is done with a "tuning fork." An idol is an ideal that "rob[s] reality of its meaning, value, and truthfulness," which is, according to Nietzsche, something that all essentialist worldviews do (EH, 71). Although at first sight the transhumanists' concept of posthuman seems to discard an essentialist view, Nietzsche's notion of the Overhuman reveals, as I will show in this section, that the transhumanists' ideal human being is at its core *invariable, uniform* and *independent*, which are, as I have discussed in the first part of this study, throughout the history of philosophy the three main aspects of an "essentialist" or "substantialist" approach (Stegmaier 1977; Aydin 2003). Transhumanists' account of the posthuman being is essentialist because they misunderstand and misrepresent, as I will show on the basis of Nietzsche's account of the human and Overhuman, the very "nature" of the human.

The Invariable Posthuman

Predictions and anticipations of radical change in the constitution of the human being, as expressed by terms like "the singularity," suggest that transhumanists adopt an anti-essentialist anthropological framework. But, on reflection, their view is rather conservative: transhumanists want a very specific and restricted "radical" change.[6] They want, as we have seen, humans to improve radically their functions and capacities using new and emerging technologies, without changing or jeopardizing their humanist values and goals. Those values and goals secure unequivocal criteria for what can be considered "enhancement."

Within that unequivocal humanist framework, causal relations are then drawn between preferred values, capacities that can contribute to realizing those values and human enhancement technologies that could improve those capacities. This is done in the following way: (1) value x is identified as an intrinsic or at least as an indisputable preferred value; (2) capacity y is identified as an important capacity that affects value x, in the form: value x can be better realized by improving capacity y; technology z is identified as a technology that affects capacity y, in the form: technology z can improve capacity y. Technology z can then be qualified as a "human enhancement technology" because (1) it can improve capacity y, which in its turn (2) can significantly contribute to realizing value x, which (3) is identified as a preferred value.

To illustrate how particular technologies are in this way qualified as "human enhancement technologies," I give three examples: inserting a brain implant for extra information storage can be considered as a human enhancement technology, because (1) inserting a brain implant for extra information storage can improve (*inter alia*) memory, (2) improving memory can improve (*inter alia*) rational thinking and (3) rational thinking is identified as a preferred value. Genetic manipulation can be considered as a human enhancement technology, because (1) genetic modification can improve (*inter alia*) deductive/logical competence, (2) improving deductive/logical competence can enhance (*inter alia*) autonomy and (3) autonomy is identified as a preferred value (Schaefer et al. 2014. Nootropics can be considered as a human enhancement technology, because (1) nootropics increase (*inter alia*) IQ, (2) giving more people the opportunity to increase their IQ will lead to (*inter alia*) more equality and (3) equality is identified as a preferred value.

Assuming unequivocal humanist criteria for what can be considered "enhancement" and "enhancement technology" secures that a person can radically transform herself and essentially still remain the same person, as also indicated by the transhumanist presumption of continuity between the over-abled and differently-abled. Although transhumanists are in favor of altering and, therefore, "violating" human nature (which is not in line with certain bioconservative humanist views), they believe that this altering will not lead to a defilement, but rather to a more effective realization of humanist aspirations. This is why transhumanists' "humanism" is a "humanism plus." The differently-abled could have more and different functions and capacities – which are brought about by certain technological interventions – than the over-abled but will share and even better realize the same unequivocal humanist values, goals and ideals. The posthuman is in fact a superhumanist.

This view is highly problematic, because transhumanists covertly assume the possibility of *a priori* criteria for what the human is and ought to be, criteria falling outside of and, therefore, not affected by the empirical realm. However, Nietzsche's notion of the Overhuman and his

conception of the human being as the "not yet determined animal" disclose such conceptions as modern manifestations of yet another attempt to turn the human "into a mummy," capturing him with "mummified concepts" (TI, 167). Transhumanists reduce the human being to characteristics (faster, stronger, smarter) that they find ideal in their current era and scope. Their lack of historical and phenomenological sense prevents them from seeing that all criteria used to establish what the human is or should be are context-bound and time-bound. Also, the meaning of concepts such as "free," "rational" and "tolerant" that are attributed to the transhumanists' ideal human being are very much dependent on their (technological) framing of the human.

What is decisive here is not the question of whether technology z can really improve capacity y, which in its turn can contribute to realizing value x. Let us assume that there is or will be sufficient empirical support for those hypotheses. The crucial point here is that transhumanists fail to see how much their conception of the human is influenced by their scope of interest, namely, existing and emerging technologies, and their technological vocabularies. These technologies are not neutral but rather "norm," explicitly or implicitly, as I have also illustrated with Artifactual Mind Thesis (AMT) with regard to cognition, what is considered as a "human" and what is meant by "enhancing human capacities" and "optimizing values" (Aydin 2015).

This point sheds a completely different light on the causal relations that are drawn between values, capacities and technologies. Let us reflect again on the particular technologies mentioned earlier. A brain implant for extra information storage, which presumably would improve memory, is not a neutral technology, but rather expresses, feeds into or brings about a particular conception of "memory" and "rationality": memory is interpreted as "processing and retrieving information," which are characteristics that are normally attributed to computers. Improving rational thinking is then, to a great extent, interpreted in terms of faster processing and retrieving greater amounts of information. Rational thinking, hence, is believed to be computational in nature. Transhumanists (and many other people) do not seem to realize that this conception arose only in the 1940s and 1950s (Sternberg and Sternberg 2012).

Genetic modification (a.k.a. genetic engineering), which presumably would improve deductive/logical competence, is not a neutral technology but rather expresses, feeds into or brings about a particular conception of competence and autonomy: genetic engineering is compared to (re)programming DNA in cells. The concept of gene (re)organization then corresponds to the arrangement of "files" in a computer system, which are part of an information processing system. In a recent paper (Haraway 1991; Chen et al. 2013), the formalism of chemical reaction networks is used as a "programming language." Sometimes the genetic code is considered to be a (quaternary) digital code. Gene (re)organization, it is

believed, will then directly or indirectly affect mental rules, models and algorithms, which in their turn will improve computational capacities. Improving autonomy is also in this case interpreted in terms of increasing computational competence.

Nootropics, which presumably would improve IQ, are not neutral technologies, but rather express, feed into or bring about a particular conception of intelligence and equality: nootropics are believed to affect brain function (e.g., by strengthening synapses between neurons and underlying synaptic plasticity or by increasing blood circulation) and improve mental functions such as memory and concentration. These and related functions are operationalized in IQ tests in indices as "recognizing similarities," "picture completion" and "digital symbol coding." Intelligence is again framed in terms of "processing and retrieving information" and quantified as a score. Enabling people to increase this score, some transhumanists believe, could contribute to more equality. However, psychologists such as Stanovich (2009) have argued that IQ tests reduce "intelligence" to a particular selection of traits and identify it in terms of what the test can measure. These tests fail to assess skills such as judgment and decision-making that most people associate with "good thinking."

Recognizing how technologies "norm" and have always normed conceptions of the (ideal) human discloses copious variations of criteria, both diachronic and synchronic, for what can be considered "enhancement." In their characterization of the posthuman being, transhumanists generalize and substantialize a current self-understanding of what a human being is and ought to be and, implicitly or explicitly, disregard other possible conceptions of "normal," "healthy" and "enhanced," and disqualify other possible (ideal) life forms.

Here, Nietzsche's criticism of evolutionary perspectives might also apply to transhumanism: not self-overcoming but rather self-preservation seems to be the transhumanists' goal (recall also Bostrom's above portrayal of the posthuman). In Nietzsche's words: "They kill and stuff the things they worship, these lords of concept idolatry" (TI, 167). The standardization of the (ideal) human is nothing else than yet another echo of an essentialist metaphysics. This Nietzschean critique helps us to reveal how transhumanists wrongly assume that their criteria for determining what is an "enhanced being" are not affected by their particular understanding of the technologies on which they have pinned their hopes and their particular technological grammar, which allows them to believe that these criteria are *a priori*. What is considered "normal" and "healthy" is redefined in terms of what technologies are able to measure, diagnose and treat. It is possible that what is now considered "normal" or "healthy" will increasingly be thought of as defective or disvalued as what transhumanists consider "enhanced states" become more the norm (Hanson 1999).

From a psychological perspective, Nietzsche often interprets this essentialist tendency as a symptom of the inability to cope with the finitude and contingency of life and all our life projects. Transhumanists' goals such as self-preservation, longevity, equality and happiness fit, therefore, much better in Nietzsche's characterization of the "last human" than in his portrayal of the Overhuman (SZ I, 10).

The Disembodied and Disembedded Posthuman

Nietzsche argues that essentialist philosophers like Plato disqualify the body and its senses due to their exhibition of multiplicity, change and fluidity, and their inability to display eternal forms. In contrast to the ratio, the body and its senses are incapable of generalizing. Nietzsche attempts to turn, one might say, Plato upside down (Stack 1983, 56). He repeatedly stresses the love of the Overhuman for the body and its senses (Hauskeller 2010). His thinking "along the guiding thread of the body" displays humans as embedded and embodied creatures who due to their particular social contexts and specific material constitutions have often very individual, nongeneralizable standards for what is "normal," "healthy," "good" or "ideal."

Here, we again encounter the problematic presumption of an unequivocal criterion for enhancement. This presumption, however, is now explained by transhumanists' "rational" (generalizing) Platonic approach. Transhumanists identify humans who lack certain functions or abilities, which are from their perspective essential, as "disabled" and presume that adding a function or capacity by virtue of technologies will make those humans able again. Give a well-functioning prosthesis to a person who has a leg missing, and he will be abled. Meanwhile, if abled and over-abled people are given additional functions or abilities, then, transhumanists believe, they become differently abled. So, if an abled person is given extra information storage that can be implanted into her brain, then he will become differently-abled. Transhumanists, being liberalists, often indicate that deciding to use particular technologies is up to individuals, but at the same time assume that their prospected technologies will affect humans in a similar, uniform way, and in so doing disregard the highly particular social and material circumstances of real – flesh and bones (and metal and code) – people. In contrast, Nietzsche's focus on the body and its embeddedness in a particular context indicates that people could respond in various and sometimes even opposing ways to certain technologies and could have very idiosyncratic standards for what they find "normal," "healthy" or "enhancing."

In her dissertation on "somatechnologies," Dalibert (2014) focuses on the way people embody technologies and shows how big the discrepancy is between the transhumanist qualification of certain aids as "enhancement technologies" and the way real persons of flesh and blood

experience those aids. A prosthesis or a spinal cord stimulation device can indeed make people mobile again or make them function more efficiently. Whether those changes are experienced as "enhancement," however, is also dependent on how people bodily relate to those technologies. A woman interviewed by Dalibert confirms that with spinal cord stimulation, she has much less pain and can move much more easily. At the same time, she indicates that she does not experience the device unambiguously as improvement. For example, she finds it very awkward that her husband can feel the pulse generator under her skin when they are in bed (Dalibert 2014, 212).

How important the social context is can be further illustrated by divergent responses to a technology such as the cochlear implant. Although many people might consider a cochlear implant as "enhancement," there are members of the "deaf community" who do not embrace this technology. They believe sign language is their cultural cornerstone. Deaf children who receive cochlear implants at a young age will be less likely to be educated in sign language during their early years, which is seen by the deaf community as a loss of culture – one that, in some cases, has been passed down through generations (Ringo 2013). There are also deaf parents that argue that they are entitled to have deaf children. One way of doing this would be to select a "deaf embryo" by using in vitro fertilization and pre-implantation genetic diagnosis (Fahmy 2011). In their social context cochlear implants are seen neither as enhancement nor as normal or healthy, but rather as deterioration.

Nietzsche's critique of unequivocal criteria for what can be considered "normal" and "ideal" prompts us to go even further: it is even possible that in one context or one period in history, x is considered as enhancement, whereas in another context or other period in history the opposite of x is seen as enhancement. "Faster, stronger smarter" might be ideal in our era. In other eras or other contexts, "slower" (e.g., the wise man who reads and "processes information" slowly), "weaker" (e.g., the passivist who prevents a war), "simpler" (e.g., the holy "idiot" who knows no resentment) could qualify as enhancement, depending on how those notions are framed. The same criticism could be applied to the main values of transhumanists: rationality can also be interpreted as cold calculation and as an antonym of affection. Autonomy can be interpreted as egoism and as a danger to communal bonds. Equality can be interpreted as leveling everybody down to the same common denominator and as a destroyer of all competition and motivation.

This critique can be further refined from the perspective of Nietzsche's "will to power": it is in the struggle of different standards that the "realities" that are set by these standards emerge and gain meaning and significance. Let me clarify by an example: probably there was no other time in which there was so much struggle about how to understand and apply the Bible than the Middle Ages. However, it was by virtue of this

struggle that the Bible in the Middle Ages was "real" and had meaning and significance. The Bible loses its significance in secularized times not because it has been repudiated in a struggle of interpretations, but because there is no struggle anymore about how to interpret it. In other words: the Bible has not been repudiated, but it has become irrelevant.

These and similar cases make clear that understanding "enhancement" from one (trans)humanist perspective and envisaging a posthuman being with certain extraordinary features is only possible if the multiplicity that characterizes the lives of real human beings is disregarded, separating people's (rational) "souls" from their material and social contexts. However, historical and phenomenological inquiry shows that people are, in Nietzsche's words, no "thinking frogs, no objectifying and registering devices with frozen innards," but reliant on and moved by "blood, heart, fire, pleasure, passion, agony, conscience, fate, and disaster" (GS Preface 3). For some people and in certain circumstances, realizing certain goals in the most efficient way might be the decisive criterion for "enhanced." Other people in other circumstances, however, might find social, emotional or aesthetic criteria much more important.

From a more psychological point of view, Nietzsche often interprets this "human, all too human" tendency to disembody and disembed the (post)human being again in terms of an inability to cope with and take responsibility for life as it is. With no realm of "objective judgments," "goods in themselves" or any "natural right," (weak) human beings experience a finite, temporally mutable, contextual and contingent life as an unbearable burden or curse that requires redemption (SZ II, 76–79). Transhumanists want life, but they cannot cope with life as it is, with all its trouble, mess, banality and limitations. They want a "body," but they cannot cope with a body that might suffer, get ill or die (Babich 2011, 35). Nietzsche's ultimate response to this tendency to deny and escape "real life" is a radical, all-encompassing Yes-saying to life, which finds its completion in the willing of the eternal recurrence of the same (Müller-Lauter 1999b, 248f).

The Autotheist Posthuman

Besides "invariability" and "uniformity," a third aspect characterizes essentialist approaches, namely "independence." In modern Western philosophy, especially that influenced by Descartes and Kant, "independence" has been strongly associated with "autonomy." The capacity for autonomy, according to Kant, is "the basis of the dignity of human and of every rational nature" (Kant 2003, 53). The ability to determine and govern ourselves is not only seen as an essential characteristic of the human being, but also as the highest intrinsic value to which we should aspire.

Transhumanists not only adopt this Enlightenment ideal of taking control of our own lives, but also redefine it from the perspective of

emerging technologies: we could truly liberate ourselves from the restrictions of our body and mind if we were able radically to modify our biology according to our wishes and, hence, completely control it. In Kosko's words: "Biology is not destiny. It was never more than tendency. It was just nature's first quick and dirty way to compute with meat. Chips are destiny" (Kosko 1999, 256). Kevin Warwick declares: "I was born human. But this was an accident of fate – a condition merely of time and place. I believe it's something we have the power to change" (Warwick 2000, 145).

In addition, many transhumanists believe this notion of an autonomous self fits in the liberal or even the libertarian view that each individual is the final arbiter of what is right and appropriate for his life or body. Transhumanists not only claim that emerging technologies will enable us to optimize and expand our mental and bodily capacities, but they will also allow us to emancipate ourselves from (governmental) authorities that limit our freedom. We will be able to design our lives as we please. The posthuman is master and commander of his own fate.

At first sight, this notion of the posthuman seems to resemble Nietzsche's idea of the Overhuman. The figure of the Overhuman also urges people to detach themselves from oppressing forces (especially from mediocrity and the status quo!) and to become sovereign beings. On reflection, however, Nietzsche's Overhuman poses very different conditions for overcoming limitations from the transhumanists' view. From a Nietzschean point of view, "overcoming limitations" does not mean finding new (technological) ways to improve capacities that could contribute to realizing certain values set by a particular (in this case, humanist) value system. The Overhuman rather challenges the inevitably "limiting scope" of every particular worldview and value system and their *criteria* for establishing what is an (ideal) human being. That explains, as indicated before, why Nietzsche is unable to provide any description of the Overhuman and why envisaging an ideal human being as a "current human being + extraordinary characteristics" is not sufficient to radically transform the human being.

The elaborated view of the Overhuman as an index of transcendence is an attempt to express this paradoxical challenge. Recognizing a dimension in human existence that in no way can be controlled, appropriated or domesticated is a necessary condition for radical self-transformation. By virtue of this transcendent dimension, the human can never completely coincide with her current state, which is a conception that is also expressed by Nietzsche's idea that the human being can never be completely determined. Radical self-transformation is only possible if the anticipated ideal in no way can be reduced to the current (and past) self-understanding of the human being. Transhumanists who claim that the human being will be able completely to design her life

and fate deny this transcendent dimension and necessarily reduce, in their projections of an ideal human being, the human to a contemporary (humanist) perfect image, to an idol. The human being has from their view no other goal beyond herself.

In contrast to the transhumanists' autotheist posthuman, the Nietzschean "ideal" is a much less convenient and reassuring idea. The metaphor of the tightrope walker in *Also sprach Zarathustra*, who "lost his head and the rope" and "threw away his pole and plunged into the depths even faster than his pole" (SZ I, 11), shows that success is not guaranteed. At least, it is naïve to assume that adopting new and emerging technologies will not affect our standards for determining what is "normal," "healthy" and "enhanced." Transhumanists should take into account that in following unknown paths, we even could end up throwing overboard the very (humanist) ideals that impel us to enhance our lives. Overcoming and, therefore, "violating" human nature could indeed come with "violating" the humanist value system.

Some technologies that are available today already affect our current standards. Cochlear implants, for example, challenge common ideas about what is "normal hearing" and instigate intense debates and serious disagreements about what is "enhancement." Moreover, imagine cochlear implants that would allow people to hear relevant "sounds" that are completely beyond the reach of "normal" hearing. Would our traditional standards for "bad" and "good hearing" then still suffice? I think not.

Cochlear implants also challenge common views about agency. If somebody "wears" a cochlear implant, then who/what hears and, hence, who/what can be considered the "agent of hearing": the human being, the cochlear implant or (and that would fit within the proposed framework) the "synthesis of human being and implant"? Transhumanists consider technological developments as transparent and as always under control of humans without taking into account the possibility of technologies becoming "actants" themselves and co-shaping both human subjectivity and the world.

Radical self-transformation implies accepting the possibility that in adopting new technologies the very notion of what is an (enhanced) human being will be revised. Being susceptible to this possibility means accepting that we are not autonomous masters and commanders of our fate. In their depiction of the posthuman being, transhumanists illustrate, yet again, how comfortably essentialist their view of the human being is, a "comfort" that in reality disguises and suppresses fear for radical change and an unknown future. Paradoxically, the Nietzschean conception of self-transformation seems a better basis for the singularity idea. That we are "a bridge and not a purpose" (SZ I, 7) comes with the uncanny possibility that we will not be the ones who will cross that bridge.

7.4 Reframing "Enhancement"

Toward a Non-Essentialist View of "Enhancement"

Transhumanists envisage a development from abled to over-abled human beings, evolving in their turn into differently-abled posthumans. Nietzsche's notion of the Overhuman has been used as a "tuning fork" to "hear" whether transhumanists escape an essentialist conception in which one particular view of the enhanced or ideal human being is rendered absolute, disregarding how people in different contexts experience a "normal," "healthy" or "ideal" life form.

At first sight, the transhumanists' visualizations of radical self-transformation, as expressed by terms like "the singularity," seem to fit into an anti-essentialist anthropology. On reflection, however, their view of the posthuman only allows for a very specific and restricted "radical" change and variety. It displays the three main aspects that, throughout the history of philosophy, have been characteristic of an essentialist anthropology. Their posthuman is essentially *invariable*, because they assume that their criteria for establishing what is enhancement are pre-given and, therefore, not susceptible to change, and disregard the fact that their ideal human being is "normed" by the technologies on which they pin their hopes. Their posthuman individual is essentially *uniform* because they believe that they can presume a general notion of enhancement, overlooking that, due to very different material and social conditions, human beings respond in various and sometimes even opposing ways to technologies and have very particular standards for what they consider as "enhancement." Their posthuman being is essentially *independent* because they think that, irrespective of the profound influence of emerging technologies, human beings are and can be in total control of their fate, disregarding the fact that radical transformation might be impossible without the demise of the present human being and her values and norms.

This Nietzschean scrutiny of the posthuman being discloses the transhumanists' ideal as a hollow idol that "rob[s] reality of its meaning, value, and truthfulness." Transhumanists' univocal criteria deny the boundless (present and future) potential multiplicity that can be found in the real world. The more practical implication of this observation is that, first, transhumanists are unable to do justice to how people in different contexts and settings experience and conceptualize "normal," "healthy" and "enhanced." This way, they (implicitly or explicitly) disqualify standards that people have in everyday life. Second, in portraying the ideal human being as a "current human being + certain extraordinary features," transhumanists do not anticipate radical transformation, but rather strive to maintain the human being in her present state. Transhumanists are, therefore, not able to account for radically new or

alternative life forms. Third, in assuming that our current basic standards will also apply to the posthuman, transhumanists portray a too cozy picture of the future, disregarding the possibility that in embracing new and emerging technologies, we might be forced to challenge and maybe even give up our most cherished values and norms.

In this chapter, I have only focused on the question of whether Nietzsche's conception of the Overhuman can reveal certain unfounded, unsound or tenuous aspects in the transhumanists' ideal, disregarding the dispute between transhumanists and bioconservatives. However, this assessment has far-reaching implications for how to evaluate the positions of both camps. On the one hand, the transhumanists' ideal seems to be much more conservative than they realize, which brings them closer to bioconservatives than they think. Transhumanists assume that their standards for "normal," "healthy" and "enhanced" are generalizable and, hence, basically independent of how different people in different contexts experience and value "enhancement" technologies. On the other hand, the transhumanists' depiction of the posthuman being is much too reassuring, which challenges us to reflect more seriously on the incentives and implications of the possible impact of emerging technologies and radical self-transformation.

Enhancement and Self-Formation

What are the bearings of this reflection on the concept of "enhancement" on the proposed notion of self-formation? The answer to this question is multilayered. First, my reflection on "enhancement" confirms the view that our (technological) environment determines to a great extent how we evaluate our selves and what we consider "good self-formation." This is, as we have seen, also confirmed empirically: particular "flesh and blood (and metal and code)" individuals have diverse standards for what is "normal" and "enhanced" and, therefore, experience, appreciate and appropriate technologies in different ways. What is "normal," "healthy" and "enhanced" seems to be very much dependent on particular contexts and lifestyles. In addition, the introduction of new technologies often challenges, disrupts and "overwrites" existing standards and imposes new ways of framing our understanding of "enhancement." We have seen that technologies are not neutral instruments but "norm," what we consider a "disabled," "normal" or "enhanced" self. Moreover, dominant technologies come with the risk of imposing one notion of self and self-formation on us, damaging the potential to form ourselves in different ways and reducing us to one form, which is one of the most important concerns in Nietzsche's work. A proper assessment of how people respond to "enhancement technologies" like prosthetics, implants and pharmaceutical drugs requires taking into account other criteria besides efficiency and functionality.

Moreover, this view is consistent with the proposed interactionist perspective, radical externalism and AMT as well as with my elaboration of the influence of brain imaging technologies on our conception of mind, brain and self. Recognizing the normative character of technologies in a dynamic environment renders the idea of an invariable, independent and uniform (humanist) criterion to determine what is an "enhanced" or "ideal human" untenable.

Second, this interactionist and externalist view of the self as something "outside" itself is, as I will elaborate in the next chapter, structural: because the self is "another for itself" and has always been open for technologies, it does not completely possess itself. This explains why in attempts to form ourselves (through education or technologically), we come across something that resists complete appropriation, something uncanny, which is not recognized in transhumanist views. This view is an implication of going beyond the inside-outside distinction: if we recognize that our "inside" is to a great extent shaped by our "outside," then also our "inside" becomes to a great extent an "outside" for us.

Third, that our standards are greatly determined by technologies and that we cannot completely possess ourselves does not have to mean that, as argued in previous chapters, we are necessarily fixed by foreign forces. Although self-formation does not occur in a realm that is independent of social, physical and technological determinants and, hence, we have to recognize that we cannot simply form ourselves as we please, it seems to be possible to critically relate to those determinants and identify ourselves with certain determinants in a certain way and not identify with others. The question that now arises against this background is if and how it is possible to recognize that we are shaped by different social, material and, increasingly, technological factors and, at the same time, appropriate and organize those forces in such a way that we can *critically* impart a form to ourselves that we can consider "good." My answer to this question is "through sublimation." In the next chapter, I will work out the notion of the "extimate self" on the basis of an investigation of the technological uncanny. In the last two chapters of this part of the book, I will explicitly elaborate the idea of self-formation as sublimation.

Notes

1 See, among other publications, the 2010 special issue "Nietzsche and European Posthumanisms" in the *Journal of Evolution & Technology*. In a special issue of the journal *The Agonist* in 2011, Keith Ansell-Pearson and three other Nietzsche scholars renewed this debate.

2 I will also disregard Nietzsche's own views of technology, which consist of scattered remarks in predominantly the works of his middle period (1876–1882). Although Nietzsche sometimes mentions particular technologies

(such as the steam engine, the railway, the telegraph), his main focus is on how modern technology organizes society and, especially, affects human creativity. It is of note that Nietzsche's attitude toward technology is ambivalent: technology could be a source for creative growth but also cause inertia and decay (McGinn 1980). These views are interesting, but of no significant relevance for the theme of this study.

3 In most transhumanists' writings, the meanings of the terms "normal" and "enhanced" are unclear. Often, transhumanists (implicitly or explicitly) understand "normal" in the sense of statistical distribution within a population: the distribution of a trait is "normal" if most individuals are clumped in the middle of the distribution curve with some outliers. They then almost always assume that the categories "normal," "healthy" or "enhanced" are descriptive and do not require further analysis (Marsh 2006). The aim of this investigation is not to challenge (bio)statistical framings of "normal" or "healthy," but to show how transhumanists' conceptions of these terms are influenced ("normed") by their particular understanding of the technologies on which they have pinned their hopes and to make clear why that is problematic.

4 Modern theologians increasingly hold to the view that the human being is an indissoluble unity of mind/soul and body (Milne 1982, 120–122). From their point of view, Nietzsche's critique of soul/mind-body dualism might be more applicable to Platonist and Cartesian thought than to Christian.

5 Sparrow (2007) develops a similar argument in relation to nanotechnology.

8   The Technological Uncanny as a Permanent Structure of Selfhood

Transhumanists and other techno-optimists often assume, as I have elaborated in the previous chapter, that it is possible to refer to invariable, independent and uniform standards for measuring what is "disabled," "normal" and "enhanced," which are the features of an essentialist view of the (post)human. The essentialist element of "independence" also comes with a dualist notion of the human, because it apparently secures an autonomous and pure (real) self that is not fundamentally affected by external factors and can form itself on the basis of general guidelines and laws that it derives from itself; in one word, a *causa sui*. Overcoming the inside-outside distinction has a price: the guidelines, laws and standards on the basis of which we establish what is "good self-formation" cannot be autonomously established but are influenced by our social, material and technological environment and, hence, variable, both synchronically and diachronically.

Recognizing that technologies are normative and, hence, "norm," what we consider "successful or good self-formation" or an "enhanced self" has also a far-reaching *existential* implication. Going beyond the inside-outside dualism and recognizing that what we consider our "inside" self is to a great extent shaped by our "outside" world implies that our "inside" is to a great extent also *for us* an "outside," which we cannot completely possess. Therefore, we cannot completely master and constrain our own process of self-formation. Or put differently, we do not completely possess the self that we attempt to form. It is not merely a "patient" that we can mold as we please.

This sense of otherness within can be experienced, as I will discuss in this chapter, as uncanny. Our very selfhood seems to contain an otherness that cannot be simply externalized, but is a constructive and structural part of what makes up who we are, which can elicit an eerie feeling. The question that I will address is how this otherness within that goes beyond the inside-outside distinction should be comprehended, whether there are more "voices," more types of "otherness" within the self – which is already suggested by the idea of a self that forms itself – and how these types of otherness relate to one another. The notion of the "uncanny" will be used to unravel these relations of alterity within, and to shed light on our existential condition in the light of a world saturated with technologies.

The concept of the "uncanny" has a history. In his seminal 1906 essay, *On the Psychology of the Uncanny* (*Zur Psychologie des Unheimlichen*), Ernst Jentsch takes as a starting point for his investigation of the uncanny the etymological meaning of the German word *unheimlich* (literally, "un-home-ly"), indicating that someone who experiences something uncanny is not quite "at home" or "at ease" in the situation concerned. The impression of the uncanniness of a thing or incident involves a "dark feeling of uncertainty," which is related to a "lack of orientation" (Jentsch 1906/2008, 217, 224). Jentsch indicates that there is one exemplary experience that illustrates this uncanny feeling most clearly, namely the "doubt as to whether an apparently living being really is animate and, conversely, doubt as to whether a lifeless object may not in fact be animate" (Jentsch 1906/2008, 221). For Jentsch, this is portrayed particularly in fiction and more specifically in storytelling. The lifelike doll Olympia, which features in E.T.A. Hoffmann's story "The Sandman" (*Der Sandmann*), is for Jentsch the prototypical example of an artifact that instigates a gloomy feeling of uncanniness (Jentsch 1906/2008, 224).

The feeling of the uncanny that is brought about by automata was taken up in 1970 by the Japanese roboticist Masahiro Mori and designated as the "uncanny valley." Reviewing the different explanations of this "uncanny valley" will allow me to put forward an alternative interpretation of why encounters with humanlike automata elicit an eerie feeling. Hooking into how Jacques Lacan, via Sigmund Freud, takes up Jentsch's view of the uncanny, I will propose that uncanny feelings not only say something about our psychological responses to humanlike robots, but also echo an ontological structure at the ground of human existence. Inspired by Lacan's notion of "extimacy," I will depict uncanniness as a fundamental dimension of our self-relation, as a permanent structure of subjectivity.

Lacan's notion of "extimacy" (Lacan 1997, 139; 2006, 224, 249) contributes to explaining why our capacity to form ourselves is restricted. This concept displays how the self is to a great extent a product of external influences and, therefore, cannot simply mold itself into whatever shape it pleases. However, Lacan's analysis primarily focuses on the symbolic order (language, laws, customs), not sufficiently taking into consideration the increasing impact of technologies on our self. Taking up Jean-Luc Nancy's concepts of "intrusion" (Nancy 2008, 161, 163, 167, 168, 169) and "being closed open" for technology (Nancy 2008, 168), I will illustrate how in our current era a technological order is evermore strongly shaping our selfhood. This technological order is "other" and "own" at the same time, which explains why technology can be experienced as uncanny.

Acknowledging that the technological uncanny is increasingly becoming a permanent structure of selfhood indicates that technology cannot simply be externalized and seen as an outside factor that can determine

or liberate us, nor as something that can destroy or strengthen our autonomy. Both transhumanists, who put their hopes on technologies that could enhance our physical and mental capacities, and bioconservatives, who warn us for the dangers of technologically tinkering with our biological and psychological make-up, do not sufficiently consider the implications of technology becoming "extimate." The proposed view calls for a more sophisticated account of how technology is shaping us as well as how we would like to be shaped by it.

8.1 The Uncanny Valley

Ernst Jentsch (1906/2008, 223) already indicated that people confronted with clever automata are likely to grow more uneasy as the automata become more lifelike and refined. The more sophisticated the machine, the less confidence a spectator would have in drawing a line separating the animate from the inanimate. In his 1970 article entitled "The uncanny valley," Japanese roboticist Mashihiro Mori depicted more precisely the relationship of familiarity and similarity in human likenesses and the positive or negative feelings that automata and other humanlike artifacts provoke. As a robot or other human duplicate becomes more humanlike there is an increase in its acceptability, but as it approaches a nearly human state responses quickly turn to strong revulsion; as the robot's appearance continues to become less distinguishable from that of a human being, the emotional response becomes favorable once again (see Figure 8.1).

It should be noted that as the graph of the uncanny valley (Mori 1970/2012, 99) flattens toward its peak, there is very little distance between the last instance where we still appreciate the robot's clever
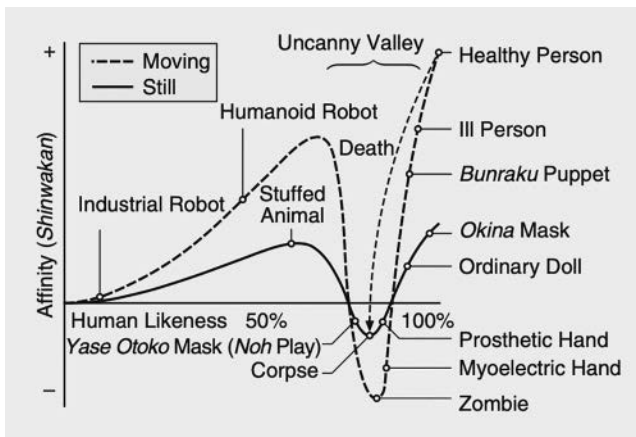
Figure 8.1 Mori's uncanny valley graph (1970/2012, 99).

resemblance and the first disorienting moment that we feel repelled by its appearance. This "little distance" indicates that it is "minor differences" that instigate an uncanny feeling, an observation that can also be found in a different context in Freud, which I will take up later in this chapter.

In designing humanoid robots, Mori advised to escape the uncanny valley by keeping a safe distance from complete human likeness (Mori 1970/2012, 100). Instead of realistic eyes or hands that prompt uncanny feelings, designers, Mori recommends, should attempt to manufacture stylish devices that are sufficiently different from human faculties and, at the same time, could be easily and comfortably incorporated or related to (Mori 1970/2012, 100). His advice has been taken up by different engineers and filmmakers who, also for commercial reasons, try to avoid having their designs fall into the uncanny valley (Geller 2008). However, at the end of his paper Mori indicates – without further explanation – that his graph could also fulfill another function:

> We should begin to build an accurate map of the uncanny valley so that through robotics research we can begin to understand what makes us human. This map is also necessary to create – using non-human designs – devices to which people can relate comfortably (Mori 1970/2012, 100)

The order suggests that understanding what makes us human through an analysis of the uncanny valley is of even greater importance than creating "homey" robots. I will return to this later.

In later years, multiple studies sought to establish whether the uncanny valley is a real phenomenon, and if it is, to explain why it exists. Participants' ratings on familiarity or eeriness have been plotted against the human likeness of human replicas using humanoid robots, androids and computer-generated characters; also morphing techniques have been employed to morph doll faces into human faces. Some of these studies show nonlinear relations that resemble the uncanny valley (MacDorman and Ishiguro 2006; Seyama and Nagayama 2007). A more recent study (Mathur and Reichling 2016), in which participants' ratings of 80 real-world android faces were observed and examined, also detected a curve resembling the uncanny valley. However, other empirical studies did not detect nonlinear relations and, hence, did not confirm the uncanny valley hypothesis (Hanson 2005; MacDorman 2006; Bartneck et al. 2007; Poliakoff et al. 2013). There are no rigorous, controlled studies that unequivocally support the existence of the uncanny valley. However, there is support for its existence from a large number of more anecdotal studies and observations. Hence, whether the uncanniness of humanlike artifacts is a function of their human likeness remains debatable (Wang et al. 2015, 394).

Multiple hypotheses have been proposed to explain the uncanny valley. Among these are a number of so-called perceptual theories. The *Pathogen Avoidance* hypothesis (MacDorman and Ishiguro 2006; MacDorman et al. 2009) was suggested by Mori himself, claiming that the uncanny valley must be related to "our instinct for self-preservation" (Mori 1970/2012, 100). From this perspective, visual anomalies in human replica, which are *perceived* as genetically very close to humans, elicit disgust because an evolved mechanism for pathogen avoidance detects these deficits as indicative of a heightened risk for transmissible diseases.

Alternatively, the *Mortality Salience* hypothesis (MacDorman and Ishiguro 2006) suggests that some humanlike robots remind human observers of their own inevitable mortality, thereby eliciting the uncanny feeling driven by a fear of death. Resembling death, people who move jerkily, humanoid automata would elicit the fear of being replaced by an android doppelgänger, being soulless machines or losing bodily control (see also Ho et al. 2008). Eerie feelings are explained in terms of defense systems that then are triggered to cope with that unpleasant prospect.

The *Evolutionary Aesthetics* hypothesis posits that humans are highly sensitive to visual aesthetics. This hypothesis suggests that selection pressures have shaped human preference for certain physical appearances signaling fitness, fertility and health. From this perspective, low attractiveness rather than lack of realism would explain the uncanniness of a human replica (Ferrey et al. 2015; see also Hanson 2005).

In addition to perceptual theories, theories have been proposed that focus more on cognitive aspects to explain the uncanny phenomenon. The *Violation of Expectation hypothesis* was also suggested by Mori himself (1970/2012), using the example of a prosthetic hand that appeared real at first sight, but elicited eerie sensations as people realized that it was artificial. From this perspective, human replicas elicit an uncanny feeling because they create expectations, but fail to match them (Mitchell et al. 2011). Here, uncanniness is elicited not so much by how humanoids look, but rather by how one thinks or assumes they will or should look like. Saygin et al. (2012) suggested that a humanoid stuck inside the uncanny valley elicits repulsion because it is no longer judged by the standards of a robot doing a passable job of pretending to be human, but is instead judged by the standards of a human doing a terrible job of acting like a normal person.

The *Categorical Uncertainty hypothesis* goes back to Jentsch, who argued that uncanniness is associated with uncertainty and mistrust which generates disorientation. From this perspective, the uncanny phenomenon concerns the process whereby cognitive uncertainty emerges at any category boundary; negative affective responses are seen as a result of

categorically ambiguous images, for example, morphed images of a real, a stuffed and a cartoon human face (Yamada et al. 2013).

The *Mind Perception hypothesis* addresses the question: "On what bases do people perceive each other as humans?" From this perspective, the uncanny feeling is linked to violating the cognitive expectation that robots lack certain capacities that characterize humans, especially subjective experience, that is, the ability to feel and sense things (Gray et al. 2007).

A theory that also focuses on robots coming too close to humans instead of not close enough is the *Threat to Distinctiveness hypothesis*, which suggests that humanlike robots, blurring category boundaries, undermine human uniqueness (Kaplan 2004; Ferrari et al. 2016). From this perspective, the fear of being replaced by a robot might not instigate fear of death but poses a threat to human identity, which elicits repulsion.

Wang et al. (2015, 395f) have evaluated the validity of different perceptual theories and indicated that they suffer from limitations attributable to the methodologies used to test their hypotheses. Another problem they raise is the usage of morphed images or computer-generated characters instead of existing human replicas, which forfeits, according to them, a certain degree of ecological validity. They have also evaluated cognitive theories that attempt to explain the uncanny feeling (2015, 397f) and pointed out that some theories of this kind neglect to explain what the cognitive expectations for humans and those for robots are, and why violating such expectations could elicit the uncanny feeling. They also note that cognitive theories fail to explain why attributing human feeling and sense experience to nonhuman and nonliving things, which belongs to a broader phenomenon known as anthropomorphism, does not seem to elicit negative effects in various other domains. In addition, Wang et al. (2015, 398f) discuss conceptual difficulties in the translations and definitions of "uncanny" (*shinwakan* in Japanese) and "human likeness," and problems in measuring the dependent variable in the uncanny valley hypothesis. They suggest that unclear interpretations and conceptualization of the variables in the uncanny valley hypothesis may have contributed to inconsistent findings.

Wang et al. (2015) stress the importance of studying the cognitive underpinnings of the uncanny phenomenon. They argue that many of the mentioned hypotheses provide plausible accounts of the uncanny phenomenon from different perspectives, while "they have neglected to verify the underlying assumption that observers would spontaneously perceive a human replica that closely resembles humans as a person" (2015, 401). Wang et al. (2015) believe that this assumption is plausible, given the proclivity we have to anthropomorphize inanimate or nonhuman entities in literature, the arts, sciences and in perception (Guthrie 1993).

Recognizing the cognitive process of anthropomorphism allows Wang et al. (2015) to propose their own *Dehumanization Hypothesis*. They argue that attributing humanlike characteristics to robots does not by itself explain the uncanny feeling; instead, the uncanny feeling, they believe, must be understood as a response to a lack of humanness. An anthropomorphized human replica is not perceived to be a typical robot, but is rather seen as a "robotlike" human. If the "robotlike" human then reveals its mechanistic nature, its humanness (above all, the capacity for emotions and warmth) is questioned, which leads to dehumanization, thereby diminishing its likability and eliciting the uncanny feeling. This hypothesis is not necessarily in conflict with other hypotheses but interprets their findings from a different perspective: "The more human observers attribute humanlike characteristics to (anthropomorphize) a human replica, the more likely detecting its mechanistic features triggers the dehumanization process that would lead to the uncanny feeling" (2015, 402).

8.2 An Alternative Explanation of the Uncanny Valley, or the Importance of "Minor Differences"

The various hypotheses that I have listed above undoubtedly explain relevant aspects of the negative responses of certain humans to certain humanlike robots. Moreover, Wang, et al. rightly show the plausibility of the assumption that in many studies, observers tend to spontaneously perceive a human replica that closely resembles humans as a person. What is also noteworthy in relation to this assumption is that it is not the big but rather the *little* differences that evoke feelings of repulsion: observers spontaneously take humanlike robots as persons, but are then repelled if they do not come close enough to humans, if small disparities reveal their lack of "humanness." The difference between having this "humanness" or not having it seems to manifest itself in very subtle and elusive features: a small delay, an unexpected acceleration, an unfamiliar gesture. One moment the humanoid is human and the other he is not.

From a psychological perspective, the nonhuman, mechanistic traits of humanoids are primarily revealed in a lack of emotions and warmth, which, from this perspective, might be a sufficient explanation. However, from a more philosophical-existential perspective, the looming "little big" question is: "what makes this humanness"? What makes the ability to feel and sense "human?" Would we consider an android that perfectly possesses these capacities human? Or are these capacities mere surface markers of a deeper layer that designates a human? What is required to bridge the gap between a humanoid and a human? Often, these questions lead to a kind of philosophical embarrassment: what makes

us human seems to escape us. The psychological accounts of feelings of uncanniness seem to allow us to see something that may have otherwise remained hidden, something strange about our own identity and existence.

I am not the first to make the move from a psychological to a more existential-philosophical account of the uncanny. Katherine Withy (2015, 48) argues – building on Heidegger – that the psychological accounts of the feeling that may accompany uncanniness refer to an "originary angst" that grounds falling (*Verfallen*), an "angst" expressing that the human cannot get a full hold of its own ground. From this perspective, "humanness" is not characterized in terms of certain capacities that can be observed and measured, but is rather rendered virtually *inaccessible*; as our mode of existence it is "too close to see." The feelings of uncanniness are interpreted as a fundamental mood that discloses a deeper ontological structure at the ground of human existence.

Yet, instead of building on Heidegger, I would like to remain closer to the originators of the analysis of the uncanny. In his 1919 essay entitled "The 'Uncanny,'" Sigmund Freud discusses and criticizes Jentsch's concept of the uncanny. He also draws on the work of Ernst Hoffmann and, like Jentsch, considers him the "unrivalled master of the uncanny in literature" (Freud 1981, 3686). In contrast to Jentsch, Freud did not regard almost-real objects *as such* as disturbing and dissonant, but rather believed that such feelings reveal deeper turmoil and psychopathology (Freud 1981, 3683). When Hoffmann's protagonist Nathaniel sees his object of love (the doll Olympia) partly dismantled with her eyes popped out of their sockets, thinks Freud, a repressed feeling resurfaced, namely the submerged fear of castration that survived from early childhood. Freud describes the uncanny as a "class of the frightening that leads back to what is known of old and long familiar" (Freud 1981, 3676) and, citing Schelling, as "the name for everything that ought to have remained … secret and hidden but has come to light" (Freud 1981, 3678). For him, Jentsch's conception of the uncanny is incomplete, since the recurrence of something repressed is required in order for a situation to be experienced as uncanny: without such resemblance, it can merely be frightening, which is different from uncanny. Freud stresses that this explains why the uncanny does not simply refer to something foreign but to an instance where something is foreign, yet disturbingly familiar at the same time. It is the "minor differences" that instigate a sense of uncanniness.

It is impossible and unnecessary to go here into questions regarding the validity of Freud's theory of repression. What I would like to take from Freud's approach is the idea that uncanniness revolves around the *tension* between unfamiliar and familiar, and hidden and revealed. Allowing us some freedom of interpretation and going outside of Freud's psychoanalytic framework, we might say that the humanlike robot elicits

a feeling of uncanniness because it reveals something that *ought to have remained hidden*, namely the unfathomability of that which makes us human. The "minor difference" between the robot and the observer of the robot disorients not only because the robot is slightly different but also because what makes the observer different appears to be incomprehensible. From this perspective, the uncanny feeling is interpreted not only as a response to a lack of humanness in the robot, but also as a response to the viewer's own inability to fathom and appropriate this "humanness" that the viewer herself possesses.

In line with this view, I propose that the uncanny valley might say at least as much about us as it says about humanlike robots. The robot might confront us with something uncanny in *us*. It is because a humanlike robot resembles me without being completely identical ("minor differences") that I am confronted with my own unfoundedness, which is constitutively strange to me. I not only become aware of what makes me different from the robot, but also of the impossibility for me to appropriate this difference. I do not suggest (nor exclude) that this explanation or interpretation could be validated by empirical research. Rather, I propose it as an explanatory, theoretical framework that could provide more insight into how technology is increasingly invasive and how our self has always been open for this technological intrusion. Following Lacan, who, via Freud, takes up the idea of the uncanny, will enable me to further elaborate the idea of these alterity relations within.

8.3 The Otherness of the Self as "Extimacy"

Freud uses the term "narcissism of minor differences" to show how it is the little differences "in people who are otherwise alike that form the basis of feelings of strangeness and hostility between them" (Freud 1981, 2355; see also 2553, 4506). Rudi Visker (2005, 433) explains that "narcissism" for Freud refers to an initially completely self-contained Ego that gradually opens up to reality. There is a movement from the inside to the outside: initially the Ego is a narcissistic entity exclusively focused on its libidinal drives, but then can gradually learn to redirect part of its energy and invest it in things outside itself. From this perspective, the self is originally a closed entity that can and should learn to gradually lose its protective shell and open up to the outside world and other people that, on first sight, seem strange and foreign. At this point, Visker (2005, 433) turns to Lacan. He notes that Lacan starts from the inverse hypothesis: the movement is not from the inside to the outside but from the outside to the inside. There is no closed original Ego, but rather the Ego is discovered and developed through the other.

Visker argues that connecting the notion of the "uncanny" to the concept of "narcissism of minor differences," which Freud himself did not

explicitly do, and, via Lacan, reversing Freud's hypothesis, can foster insight into "alterity-relations within." This type of relations indicates that not only the otherness of other people needs to be recognized, as Levinas relentlessly stressed in *Totality and Infinity* (1961/1969) and other works, but also the otherness that somebody finds in herself.

In his famous essay "The Mirror Stage" (1949/2005), Lacan argues that the child discovers itself as a unified entity in and through something else, such as its own mirror image, the body of another child and the responses of its parents. It would be inaccurate to say that the child recognizes itself in the Other, since it is only by virtue of that other and the discourses, goals, ideals and desires that others impart on it that the child develops and discovers a self. From this perspective, identity is the result of identification, though without assuming that there is a subject prior to that process of identification (see also Julien 1990, 43–51).

Lacan's view of the relation between self and other is paradoxical and uncomfortable: the other is both the necessary condition for forming a self and, at the same time, an obstacle that prevents the self from reaching the unity that it seeks. In Visker's words:

> identity will always bear the trace of an exteriority that it cannot fully interiorize. I am another (*je est un autre*) means: I cannot do without that other through whom I get an I. That other becomes someone that I cannot expel. In other words, my alienation is original, for it is implied in my self-constitution. There is no "self-hood" without "foreignhood." The self is not something I possess, my "self" is irremediably infected with an otherness that prevents me from being fully at one with myself (Visker 2005, 433)

Instead of understanding the alterity within in terms of introducing another "in" the self, the self is revealed as something that is from the beginning contaminated with another. Lacan calls this otherness of the self *extimacy*: the "own"-ness of the self is both strange and familiar, both inside and outside, neither inside nor outside. The self is always outside its center; the self is, one could say, referring to Helmuth Plessner's view that the human never completely coincides with herself, "excentric" (Plessner 1975).

Besides developmental psychological accounts of the self (such as the mirror stage), Lacan uses surrealistic and Escher-ish figures to visualize the dizzying structure of extimacy, for example, in the topology of the Möbius strip: the Möbius strip's half-twist results in an "odd" object (Lacan 2014, 120) because the single surface of the strip passes seamlessly from the "inside" to the "outside." Not only is it impossible to distinguish the inside surface from the outside one, but it is also impossible to tell left from right. It is disorienting and confusing: "You literally can't make heads or tails of it" (Robertson 2015, 18). For Lacan,

self-relations are characterized by this perplexing strain to distinguish "inside" from "outside."

From this Lacanian perspective, not only the other or otherness outside escapes definition – as Levinas (1961/1969) attempted to illustrate – but also the self that is confronted with that otherness. The self is not something I completely possess, but is rather irremediably infected with an otherness within that prevents it from being fully at one with itself. The alienation is original, for it is implied in its self-formation. The self finds itself attached to something within, which is experienced as its selfhood, without being able to sufficiently understand and explain this attachment. It was *already there* before the self discovered itself as a self-reflecting agent. It cannot be fully objectified because it is always too close to the self.

This otherness within has for Lacan different dimensions. For one, the self's alterity within entails the influences of the external world that we have gradually incorporated. As we have indicated above, a child discovers and develops a unified self through embodying different external instances. The image that the self projects on itself through others is, according to Lacan, also an *imago*: a unified, stable and ideal totality ("that's you Helena, yes you are a wonderful girl, you are a princess, you're going to grow up to be beautiful and smart, just like your mommy"). The self attempts to realize this ideal image through identification, and subsequently enters a lifelong quest to correspond wholly with this Ideal-I, a quest that, Lacan stresses continuously, can never be completely fulfilled (Lacan 2005, 12, 15, 18). The *imago* also refers to the *imago Dei*, the image of God in which human beings were created and with which they should strive to conform but can never completely achieve. It is important to stress that the *imago* is not an emanation of the individual but the result of an encounter with larger Others and their desires, goals and ideals. Lacan sometimes designates this dimension, which also corresponds to a phase in the development of a child, as the "Imaginary Order" (Lacan 2005, 158, 161).

The images that others project on the self, by virtue of which it develops a sense of an unified Ego, also gradually enable the self to enter into what Lacan sometimes calls a Symbolic Order (and sometimes the "Big Other"): the pre-existing order of customs, institutions, laws, mores, norms, practices, rituals, rules, traditions and so on of cultures and societies, which are entwined in various ways with language (Lacan 1997, 20, 81). The Imaginary and Symbolic do not coincide: the Imaginary is central to Lacan's account(s) of ego formation and manifests itself in dyadic relations (such as in the self and its mirror image), whereas the Symbolic constitutes triadic relations by introducing, besides dyadic and intersubjective relation, a trans-subjective symbolic order that normatively regulates the relations between particular beings and society (Lacan 1997, 81, 234). In short, the self is what it is in and through

mediations of the endorsed image that others project on it as well as through subjecting it to socio-linguistic arrangements and constellations.

There is, besides the Imaginary and Symbolic Order, also something else that constitutes the self, a dimension that Lacan designates with different (related) names: the Thing (*La Chose*), the Real Other or the Real Order. In the *Mirror Stage* (1949/2005), Lacan stated that in the image of the child reflected in the mirror there is one element, like the eyes of a creepy, living doll that fails to integrate into a functional totality and necessarily appears fixed and immobile: the gaze. It has an uncanny way of detaching itself from me, said Lacan (Lacan 2014, 97; Robertson 2015, 25). It refuses to integrate into a functional totality. The reflection in the mirror serves to organize the child's movements and body parts in a unified whole. At the same time, this framing seems to leave behind a "residue" that escapes the subject's sense of complete mastery over her body (Lacan 2005, 3). Visker stresses that this drive within, unlike *angst* in Heidegger or shame in Levinas, is not something that is liberating or beneficial but an uncanny guest, a Thing that the self needs to be protected from. In all my attempts to control and "domesticate" it, I recognize that it escapes me, might cross the borders that I and society have set, disorient me, and potentially might destroy me (Lacan 1997, 43–56).

However, the Thing is also not sheer negativity, as Lacan's depictions of the Thing might suggest. It is something that does not fit and cannot fit into an encompassing frame of meaning. By virtue of this aspect, the self can never be completely captured and domesticated by the ideal images that others project on it, nor by the symbolic order in which it is immersed. It is this dimension that gives the self particularity and singularity.

This characterization of the self renders Lacan's psychological anthropology completely at variance with Anglo-American ego psychology and the Enlightenment spirit, which seek to strengthen people's ego and liberate them from restrictions. Despite having consciousness, the self is not a locus of autonomous agency; it is not the seat of a free "I" determining its own fate. The self is thoroughly *compromised*. The other (in its three manifestations as Imaginary, Symbolic and Real) is both the necessary condition for forming a self and an obstacle that prevents the self from reaching the unity, autonomy and singularity that it seeks, not only because it cannot meet certain demands of others or because it has been shaped by a world that it was thrown into (to borrow a Heideggerian term), but also because it can never fully appropriate what it desires. Without the other, it is impossible to discover and develop subjectivity or selfhood and, at the same time, the other prevents it from becoming an autonomous being, unaffected by its traces, inscriptions and whims; or put yet differently, in all my attempts to become an independent and unique self, I remain to a great extent a repository for the projected desires and fantasies of larger others and a plaything of the idiosyncratic and disruptive vagaries of an unruly force within.

This makes, as indicated earlier, the "otherness" in the self more disturbing, since the self is unable to externalize it, detach itself from it and localize it, which explains its uncanniness. Since the self becomes what it is by virtue of its encounters with manifestations of this other, it remains always a stranger or other for itself. The self is never completely "at home." Its "own"-ness is, as we have seen, both strange and familiar, which explains why uncanniness is a permanent dimension of its subjectivity.

It is crucial to understand this other within from a radical anti-essentialist view that goes beyond inside-outside dualisms, since the other that is beyond our control is, at the same time, responsible for forming our self; our self-relation is inherently an alterity relation. The self is, contra Freud, not something that has to learn to open itself for others, but rather has to find a way to live and not to be crushed by that other that, from the beginning, is already inside: the stranger outside me can make me aware of and awakens the otherness inside me, which can fill me with incongruity, confusion and sometimes rage (see also Visker 2005, 435).

8.4 Being "Closed Open" for Technology

Lacan illustrates how otherness structurally constitutes the self. In his depiction of how society shapes the self, he predominantly focuses on the world of language, laws and customs. However, in our present culture, we are witnessing, besides or in addition to a Symbolic Order, the ever-stronger ubiquity of a Technological Order. Today, virtually all facets of our lives are saturated with technology. It must be said that the material world is not absent in Lacan's account. It is notable that his prime example of Otherness involves an artifact, namely the "*mirror image.*" In fact, in the 1949 text, Lacan seems to think of artifacts as equally relevant props as humans (parents, peers etc.) within the context of subjectivation, but in subsequent reinterpretations of the mirror stage during the 1960s, he increasingly highlights the supporting role of fellow human beings, caregivers' narratives and socio-linguistic factors. If Lacan would have lived and written in our era, where technologies are becoming more intimate and intrusive than ever before, he probably would have emphasized more the role of technologies such as screens, tablets, mobile phones, social networking services, brain imaging and other medical technologies, and algorithms and other digital grammars. In order to explain how we find ourselves in an "extimate" relation not only with a symbolic but also with a technological order, and how this relation is increasingly shaping our selfhood, I will complement Lacan's notion of "extimacy" with Nancy's view of being "closed open" for technologies. This technological order that is "other" and "own" at the same time, might further explain in what sense technology is experienced as uncanny in our current era.

In 1990, French philosopher Jean-Luc Nancy got severely ill and needed to undergo a heart transplant. In an autobiographical essay entitled *L'intrus* (*The Intruder*), he documented this experience. Nancy notes that his heart has always seamlessly kept him alive, supplying oxygen and nutrients to the tissues in his body. Before his illness, his heart was, as Nancy describes, the most private and intimate part of himself and, at the same time, not more than a piece of meat, invisible and without meaning. After he got ill, his relation to his heart radically altered: in order to survive, he had to get rid of it. Nancy says: "My heart became my stranger" (2008, 163). Nancy was still his heart, but, at the same time, his heart became something foreign. Instead of an ally, suddenly his heart became a dangerous enemy. His heart became an intruder, not one that enters from outside but one that enters from *inside* (Nancy 2008, 162f). We see here that the idea of an "intruder from inside" renders the apparently clear-cut distinction between "inside" and "outside" opaque.

Besides his sick heart, Nancy describes many other forms of strangeness that he experienced. The donor heart that he got was seen as a stranger. As Nancy states: "my heart can be a black woman's heart" (Nancy 2008, 166). Also his own immune system – normally his most important protector and ally – became a threat, since it needed to be suppressed in order to accept the donor heart. Furthermore, his age became a stranger, since the donor heart could be 20 years younger than he is (Nancy 2008, 169). And this was not the end of Nancy's strange encounters: after his heart transplant, Nancy got sick again and developed cancer; now the cancer cells, which prior to his illness were not identified as different, became a dangerous stranger.

The long list of strange entities that he came across in his body led to Nancy's observation that not only parts of his body, but also his body *as such* is a stranger to him. Moreover, while reading Nancy's essay, it gradually becomes clear that its main theme is not his heart transplant, nor his cancer cells, nor his illness. His line of thought culminates in a reflection about how the "intruders" from within and without reframe his view of his "self." He writes:

> I am the illness and the medicine, I am the cancerous cell and the grafted organ, I am these immuno-depressive agents and their palliatives, I am these ends of steel wire that brace my sternum and this injection site permanently sewn under my clavicle, altogether as if, already and besides, I were these screws in my thigh and this plate inside my groin (Nancy 2008, 170)

In addition, Nancy's focus shifts from observations on his body and the way he relates to it to a reflection on the technologies that are inserted in his body, the technological manipulation of his body and how his relation to these technologies sheds a different light on his body and self.

In relation to the notion of the "intruder," Nancy employs the idea of the self being "closed open," which together signify how the technologies that are used to treat and keep his body alive are evermore interwoven with his very self: "'I' always find itself tightly squeezed in a wedge of technical possibilities" (Nancy 2008, 162). The idea of being "closed open" indicates that the technologies used to treat Nancy should not be seen as strangers from an outside realm that infringe the self; rather, the self is exposed as always having been part of that "outside." As Nancy explains: "What a strange me! Not because they [the surgeon, the technologies] opened me up, gaping, to change the heart. But because this gaping cannot be sealed back up. (…) I am closed open" (Nancy 2008, 167f).

Nancy stresses that current technologies highlight the alterity in selfhood, though, at the same time, he makes clear that they did not cause or generate it: "never has the strangeness of my own identity, which for me has always been nonetheless so vivid, touched me with such acuity" (2008, 168; see also Slatman 2007). Nancy attempts to illustrate – very much in line with Lacan – that alterity is a constant dimension of our self and self-experience. It can also be experienced if the body is not ill. The heart transplant and other technological intruders make this experience only more acute, but have not generated this being "closed open." We have always been strangers to ourselves. In Nancy's words:

> The intruder is nothing but myself and man himself. None other than the same, never done with being altered, at once sharpened and exhausted, denuded and overequipped, an intruder in the world as well as in himself, a disturbing thrust of the strange, the *conatus* of an on-growing infinity (Nancy 2008, 170)

The self has always been outside itself and, hence, can never be completely *closed* in order to entirely possess itself.

It is clear that for Nancy (and for Lacan), the self has always been "closed open," but that does not imply that with the advent of new technologies there is nothing new under the sun. New and emerging technologies have expanded the possibilities to "intrude" in the "closed open self," which is also confirmed by Nancy: "I am turning into something like a science-fiction android, or else, as my youngest son said to me one day, one of the living-dead" (Nancy 2008, 170). Besides tradition, education and culture, now technology has become a dominant force in self-formation processes, as Nancy very intimately has experienced. The human has always been "closed open," but now she can immediately intervene in her own bodily constitution. The potential to be "closed open" has always existed, but technologies today take increasing advantage of this potentiality:

> Man becomes what he is: the most terrifying and the most troubling technician, as Sophocles called him twenty-five centuries ago, who

denatures and remakes nature, who recreates creation, who brings it out of nothing and, perhaps, leads it back to nothing. One capable of origin and end (Nancy 2008, 170)

Lacan's idea of "extimacy" highlights that the self is not a closed "inside" that then learns to open up to the outside world, but that the self is rather discovered and developed in and through a pre-existing symbolic order, an order that, on the one hand, is constitutive for its subjectivity and agency and, on the other hand, is an obstacle that prevents it from reaching the autonomous unity that it desires. What Nancy's elaboration of his experience of being "closed open" for technologies adds to this framework is that the self, as an epistemic object, is evermore deeply immersed in a pre-existing realm of biomedical knowledge and technology. This technological realm is increasingly shaping the self. The technology that potentially always can intrude in the self also affects and transforms how the self experiences itself, and in which direction the self is formed. For Nancy, technology does not extend the mind or the self, but the self has always been open and exposed and now technology is excessively *confiscating* it (see also Aydin 2015): "the subject's truth is its exteriority and its excessiveness: its infinite exposition. The intruder [in this case technology] exposes me to excess. It extrudes me, exports me, expropriates me" (Nancy 2008, 170).

Complementing Lacan's notion of "extimacy" with Nancy's view of being "closed open" for technologies makes it possible to reinterpret the other within in terms of technology within. The technology within is not completely strange or foreign, since it is a constitutive part of our subjectivity and selfhood. At the same time, technology prevents one from becoming an autonomous and singular being, unaffected by its engravings. Technology is strange and familiar, at the same time. That "at the same time" explains why it can be experienced as uncanny.

8.5 Alterity in Selfhood and the Question of Technological Self-Formation

The idea of the uncanny has been used to designate an alterity within that cannot be simply explained in terms of something external that challenges or influences our internal convictions, preferences, values or goals. From Freud, I have taken the view that the uncanny cannot be simply be opposed to the canny: *heimlich* and *unheimlich* are not simply opposites, since *unheimlich* signifies the concealed and the hidden and, at the same time, the familiar and domestic. The uncanny within is strange and familiar, at the same time.

Lacan's notion of "extimacy" has been employed to further illustrate how "ownness" does not exclude but rather includes "otherness." This notion expresses on the one hand that even our most personal goals,

aspirations and ideals that we attempt to realize in order to become an ideal-I are derived from significant others in our lives. The sense of being a unified Ego is derived from images that others project on me. My desires are ultimately desires of others, such as my parents, educators, role models, superstars, Party, God, Nature and Science. In confrontation with significant others, we gradually enter a symbolic order that enable us to become part of a community and define ourselves from a third person perspective as subjects with certain roles, duties and responsibilities.

On the other hand, Lacan stresses that there is also some-Thing in us that prevents being completely absorbed by societal aspirations, values and ideals, including ethical, political and (we can add) *technological* rules, regulations and grammars. Although by virtue of this drive humans are singular beings, Lacan points out that this dimension should not be romanticized; in its purest form, it is an unfathomable and disorienting abyss of withdrawn-yet-proximate alterity. In order to regulate its drives and impose a form to them, the self needs help in the form of a symbolic and technological frame or narrative. Lacan stresses the importance of the Symbolic and Imaginary and their protective, orienting and stabilizing workings. At the same time, he points out that the process of subjectivation and socialization always hold the chance of excessively repressing and fixing the self through a particular "sheltered" system that ultimately becomes a straitjacket and prevents developing a singular identity.

Through a reading of Nancy's *Intruder*, I have tried to complement Lacan's social order of language, laws and regulations with a technological order that is increasingly shaping the very nature of our selfhood. In our current era, technologies and technological systems can be added to the Lacanian Imaginary Other that projects its desires on us and the "Big Other" that regulates our conduct: an IPhone is not only a handy device for making calls, texting and surfing the web, but promises us to upgrade our identity and lifestyle. Brain imaging technologies are increasingly used not only to diagnose diseases and lesions, but also to correlate brain activation with psychological states and traits up to a level that, some predict, will enable us to correct the mental states that someone ascribes to herself or even establish whether someone really possesses free will (Aydin 2018a). Upcoming persuasive technologies will influence our wishes and desires more seamlessly, making it even harder to recognize them as being projected to us (Frischmann and Selinger 2018).

The idea of a socio-technological order influencing and regulating our conduct and interactions as well as generating social stability is not a completely novel view. Philosophers like Hegel and Gehlen have argued that institutions and institutionally conveyed mental habits have the formal and informal function to unburden and give coherence and continuity, to compensate for the human's lack of instinctual determination. Also, "alienation theories," often strongly influenced by

Hegel and Marx, thematize the estrangement of the self in our modern, technological culture. However, Lacan and Nancy illustrate that this order is a constitutive dimension of the self, and cannot be simply externalized and objectified. Instrumentalist and determinist approaches to technology as well as techno-optimist and techno-pessimist approaches (including transhumanist and bioconservative approaches) often overlook that technology cannot be simply situated outside humans and their condition. The "technological other" limits our capacity to form ourselves, not because it constrains an original capacity to make autonomous decisions, but because this "technological other" has engraved – and is evermore deeply engraving – its structures in our very origin. Technology increasingly enables us to form ourselves into stable and socially dependable beings and, at the same time, prevents us from reaching the autonomy and uniqueness that we seek, which could account for the uncanniness that some technology seems to elicit.

Reflection on the uncanny feeling triggered by a humanlike robot prompts the question not only of what makes robots different from humans, but also what makes humans different from robots: the lack of humanness that would elicit the uncanny feeling instigates the question of what makes up this "minor difference." In confrontation with the humanlike robot, I not only become aware of what makes me different from it, but also of the impossibility for me to appropriate that difference. The elaboration of the "extimate structure" of the self has led to the finding that the self is formed in the image that the "outside world" projects on it. Since the "outside world" is increasingly a world of technology, the self, being "closed open," is increasingly being shaped in the *image* of technology; technology is increasingly becoming the "Big Other," the dominant "intruder within." Thus, perhaps in the confrontation with the strange and, at the same time, familiar robot, the self not only uncannily senses the human in the robot, but also *the robot in the human*. From this view, it is inaccurate and inadequate to frame the self as something that could or should close itself off from, or alternatively learn to open itself for, technology. In line with what Lacan and Nancy say about the other within, I propose instead that the self should find a way not to be restricted and crushed by, but rather live in a *critical and deliberate* way with, the technology which from the beginning is already inside.

However, there is a complicating factor which I have ignored so far. For Lacan, the social order, on the one hand, protects the self from arbitrariness and excess, and, on the other hand, always comes with the chance that protection keels over to repression. What Lacan does not seem to have envisaged, besides the view that that the symbolic other could be toppled by a technological other, is that this technological other, instead of securing *order*, could also become a source of disruption itself. We have seen that transhumanists and other techno-optimists

who propose enhancing human capacities by means of existing and emerging technologies often do not take into account that these technologies are influencing, challenging and disrupting our very standards for establishing what *are* "enhanced capacities." They wrongly assume that it is possible to refer to univocal standards for measuring what is "disabled" and "normal" as well as what is an enhanced self or "successful or good self-formation" (Aydin 2017). In addition, the authority of traditional "Big Others" is more easily questioned and challenged in our current global society, which harbors different and sometimes opposing views, values and ideals, different and opposing views that are accessible to ever greater parts of the world population through the Internet and other media. Not only do we need to deal with the human being as a "monster and an abyss," a being that escapes every possible uniform categorization and, therefore, continuously is able to challenge and disrupt current standards, we now also seem to witness a "technological other" becoming an additional disruptive force. The technological other is becoming an additional disorienting dimension that also could further intensify the uncanny within.

In the wake of univocal standards being challenged and undermined from different others without and within, the question of how to form ourselves becomes evermore acute. How is it against this background still possible to sustain the ideal of "good self-formation"? How can one develop a coherent, social and a singular self in the light of our intrinsic technological condition? Recognizing that technology "conditions" our humanness, could we also consciously employ it to "condition" our humanness in a certain desired direction and form ourselves in a "good" way? I believe that the notion of "sublimation" might prove itself fruitful in this respect, which I will discuss in the next chapters.

9 Self-Formation as Sublimation and the Question Concerning Technology

The developed idea of "self-formation" is presented as an alternative to essentialist and dualist notions of the self. Instead of understanding it as an *a priori* given entity that is independent of its environment, the self rather shapes itself by imparting a particular form to the interactions it has with its world. Since technologies play an ever-stronger role in this shaping, self-formation is increasingly captured as technological self-formation, which for some techno-optimists gives rise to boundless aspirations for self-enhancement. New and emerging technologies could, they believe, improve our neural organization and capacities as well as reshape our motivational patterns and emotional responses. These technologies promise us mastery over our biochemistry and ultimately total control over our lives.

In the previous chapters, I have shown that this enhancement rhetoric does not go beyond essentialist and dualist accounts of the human and its relation to technology but only rehabilitates such ontological frameworks. Transhumanists propose overcoming essentialist views and anticipate radical transformation of humans through technology. Nonetheless, they assume that present humans will not cease to exist, nor radically change in this radical transformation, and that the criteria to determine what is "normal" and what is "enhanced" are both in our present time and in the future, to a great extent univocal, and that humans are and will remain in control of their self-formation.

These assumptions bring bioconservatives, despite their opposition to human enhancement technologies, closer to transhumanists than on the first sight is noticeable. In both approaches only – technologically enabled – humans are seen as agents, without taking into account that technologies could also function as "actants," shaping both human subjectivity and the world. This way technological development is seen as transparent and as always under control of human subjectivity; technologies are considered as value-neutral tools that are employed by autonomous humans in order to realize the goals that they have set for themselves. This view of human-technology relations we came across in various forms in, among other, instrumentalist, determinist, techno-optimist and techno-pessimist approaches.

Throughout this study, I have argued that the environment, which is increasingly a technological environment, cannot be simply situated outside humans and their condition. Overcoming an essentialist and dualist view requires recognizing that technology has become an intrinsic part of our mind, body and self, and both "norms" and disrupts the standards that we use to establish what we consider "normal," "healthy" or "enhanced" features and capacities.

This picture of a self beyond the inside-outside distinction appears to be far from harmonious and reassuring. Technologies increasingly seem to shape the very fabric of our selfhood, including the desires, goals and ideals that we consider as most personal. In addition, these desires, goals and ideals seem to be continuously subject to the disruptive influence of new and emerging technologies as well as to the idiosyncratic whims of an unruly force "within." The self appears, as I have elaborated in the previous chapter, to be compromised and seem to have virtually no space to critically and deliberately form itself.

How to recognize this unsettled structure of the self and prevent adopting the view that the self is not more than a plaything of contingent influences and forces? How is it against the displayed background still possible to sustain the ideal of "good self-formation"? In the wake of univocal standards being challenged and undermined by an external and internal technological "other" as well as from uncontrollable drives, the question how to form ourselves becomes evermore acute. In this chapter, I will argue that the notion of "sublimation" might prove itself fruitful in this respect. Instead of understanding self-formation as enhancement, I will propose that it can be better captured as sublimation, while also reflecting on the role of technologies in that process. This notion of sublimation will be developed through revisiting some of the authors that have functioned as guides throughout this study. Most importantly, the idea of critical self-formation by virtue of longer-term goals and ideals will be taken up again and further integrated in the proposed self-formation as sublimation framework.

9.1 Sublimation as Substitution of Raw Drives for Social Valuations

Sublimation is a key notion in Freud's psychoanalysis. It is a necessary condition for full psychic health and the desired result of successful psychoanalytic treatment. In general, the aim of psychoanalytic treatment is bringing hitherto repressed drives, desires and wishes to consciousness through language and dialogue, and canalizing and redirecting energy that has previously displayed itself in unpleasurable symptoms to more useful and appropriate ends. For Freud, sublimation involves the redirecting of a repressed sexual drive toward a nonsexual aim. In his words: "the diversion of sexual forces from sexual aims and their direction to

new ones – a process that deserves the name of 'sublimation'" (Freud 1981, 1500; see also Freud 14:94).

Freud's definition of "sublimation" might appear straightforward, but has prompted a great deal of discussion and has been a source of much confusion. As Laplanche and Pontalis stress: "The lack of a coherent theory of sublimation remains one of the lacunae in psychoanalytic thought" (1973, 433). A major problem with Freud's notion of sublimation is that it seems to fail to distinguish sublimations from neurotic symptom formation (Gemes 2009). Freud's "repression theory" indicates that drives that have been repressed can manifest themselves as "substitute formations" in various types of behavior. However, his account of sublimation does not seem to enable separating those substitute formations, those instances of the "return of the repressed" (Freud 1981, 2215, 2984), which are symptoms, from those that are genuine sublimations. Relatedly, illness cannot be separated from health. An obsessive-compulsive who washes her hands 100 times a day may have redirected a sexual aim to a nonsexual one, but that could hardly be considered a case of successful sublimation.

This difficulty might be partly explained by Freud's famous recognition that it is impossible to sharply and unambiguously distinguish the normal from the pathological: "We no longer think that health and illness, normal and neurotic people, are to be sharply distinguished from each other, and that neurotic traits must necessarily be taken as proofs of a general inferiority" (Freud 1981, 2300). Nevertheless, Freud's psychoanalytic practice could not do without making a distinction between health and illness.

Freud makes the distinction between health and illness throughout his work, and stresses that sublimation results in behavior or activity that can be considered socially or culturally valuable. Sublimation is explained by referring to social valuations: "A certain kind of modification of the aim and a change of object, in which our social valuation is taken into account, is described by us as 'sublimation'" (Freud 1981, 4702). He repeatedly relates sublimation to "deflecting the sexual instinctual forces away from their sexual aim to higher cultural aims" (Freud 1981, 1956).

Some critics of this view of sublimation have attempted to make the distinction between pathological and normal without having to introduce nonclinical criteria like social valuation. For example, Fenichel (1945) tried to provide a strict psychoanalytic account on the basis of the distinction between sublimation as a successful defense and repression as an unsuccessful defense. Although the distinction between repression and sublimation is not always clear in Freud, there are passages that strictly distinguish the two, which Fenichel seems to take as lead: "Premature repression makes the sublimation of the repressed instinct impossible; when the repression is lifted, the path to sublimation becomes free once more" (Freud 1981, 2238). Fenichel discusses three

differences between repression and sublimation. First, he indicates that in sublimation, "the original impulse vanishes because its energy is withdrawn in favour of the cathexis of its substitute" (1945, 141). Second, he argues that sublimated impulses "find their outlet," whereas repressed impulses do not (1945, 141). Third, he points out that in sublimation, as opposed to neurotic substitute gratifications, there is a desexualization (1945, 142).

Gemes (2009, 45f) tries to illustrate on the basis Freud's discussion of the Leonardo case that Fenichel's distinction between unsuccessful repression and successful sublimation cannot be sustained. He argues that one could hardly say that Leonardo's original homosexual impulse had vanished, since they are expressed in many of his activities, such as his drawings of perfect idealized male bodies. This observation also challenges Fenichel's idea of desexualization as a marker that separates sublimations from pathological symptoms. We have seen that also the example of the obsessive-compulsive displays the unsustainability of this distinction, which makes Fenichel's as well as Freud's distinction between pathology and sublimation, in terms of the distinction between sexualized and nonsexualized expression, untenable. Both sublimations and pathological symptoms seem to appear in sexualized and nonsexualized forms.

It seems to be difficult to understand sublimation in strictly psychoanalytical terms independent of social valuations. At the same time, making sublimation dependent on what a particular society happens to find acceptable is, as Laplanche and Pontalis indicate, also not without problems, since it could lead to repression and arbitrariness: "Should the fact that activities described as sublimated in a given culture are accorded particularly high social esteem be taken as a defining characteristic of sublimation?" (1973, 433).

It is not a coincidence that reflection on the meaning of "sublimation" in Freud culminates in a tension between drives and social valuation. The tension between "unregenerate instincts and overbearing culture" lies at the heart of his psychoanalysis and his notion of the self: "Since the individual can neither extirpate his instincts nor wholly reject the demands of society, his character expresses the way in which he organizes and appeases the conflict between the two" (Rieff 1965, 28). From this perspective, a clear-cut distinction between framing sublimation in terms of social valuation and understanding it in strictly psychophysiological terms is difficult to sustain. Freud seems to consider a compromise between these two forces attainable but not a final resolution; culture remains always potentially repressive and drives always lie in wait below the surface, ready to disrupt socially accepted behavior. There is always the hazard of a renewal of conflict. While favoring sublimation, Freud stresses the limits of our capacity to redirect the drives and substitute raw gratification for culturally valued representations.

9.2 Sublimation as Organizing Drives Around
     a Disruptive Void

Partly continuing Freud's line of thought, Lacan has taken up, as I have indicated in the previous chapter, the tension between drive and culture. This tension is expressed more or less consistently throughout his work in his account of the three categories or dimensions of otherness: the Real, the Imaginary and the Symbolic. In order to explain his view of sublimation, I have to review and map out these three categories and make clear how they constitute the self in its psychological development in its early years and, more importantly, how they form structural dimensions throughout its life. Lacan explains, as we will see, the tension between drives and social valuation eventually in terms of a tension between the singular and the general, which I will use to more precisely identify what is at stake in the idea of self-formation as sublimation.

The Real, the Imaginary and the Symbolic

From Lacan's developmental psychological perspective, the Real is associated to the state from 0 to 6 months of age in which the infant is completely entwined with nature, primarily with its mother. There is no sense of separation between itself and the world that exists beyond it. The Real refers to a stage prior to individuation, language and culture. Although in this neo-natal stage, the self is closest to the pure materiality of existence, it marks already a break with wholeness and fullness, which makes the Real in its purest form incognizable and impossible. The Real is a border category that can only be denoted as a loss and a void. The primal drive that determines the child's conduct in this stage is, according to Lacan, "need" (Lacan 2006, 580). Lacan sometimes compares this primal need to the primordial animal need for copulation (Lacan 1997, 208). A baby tries to satisfy this need with no sense for any separation between itself and the external world or the world of others. For this reason, Lacan sometimes also depicts this state of nature as a realm of fullness or completeness: there is a kind of primordial "need" but no ego that experiences itself as a separate agent in need of something outside itself.

   At the age between 6 and 18 months, the infant gradually steps out of the order of the Real and starts to experience itself as a detached being. This awareness of separation is both a gain and a loss; the infant develops a sense of self that is accompanied with a sense of loss of fullness and non-separation that it had in the Real. The separation of the infant from the mother as a unity is initially not accompanied with the infant experiencing itself as a unity. Having no control over its body, it experiences itself as fragmented and in pieces. It may see its own hand without knowing the hand belongs to it. Instead of need, demand is the

dominant drive in this phase (Lacan 2006, 580). The infant demands to return to the sense of original unity; it wants the idea of "other," and it being another, to disappear. Demand is thus the demand to return to a primordial state of fullness and non-separation that will also end the lack and vulnerability the infant is experiencing. But of course that is impossible, because that lack or absence is the condition for individuation, for becoming a separated Ego and, eventually, a functioning cultural being. Whereas needs can be fulfilled, demands are unsatisfiable.

The outside world offers a kind of rescue from this unpleasant situation. At some point in this period, the baby will see its image in the mirror and recognize it as its own self, recognize its own body in the bodies of other children that resemble her or be encouraged by responses of its parents to recognize itself as a separate self (Lacan 2006, 75f). Lacan characterizes this recognition simultaneously as misrecognition: the image that is seen by the infant or is echoed by her parents is not the real thing, but a reflection or a fantasy as well as an idealization ("yes, that is you Helena, good girl Helena, you are so great, you are a real princess"). The misrecognition has as function idealization as compensation for the loss or alienation from the order of the Real. The infant moves from insufficiency to anticipation: still unable to be whole, it begins to anticipate being whole, moving from a "fragmented body" to an "orthopedic vision of its totality," to a vision of itself as whole and integrated, which is "orthopedic" because it serves as a crutch, a corrective instrument, an aid to help the child achieve the status of wholeness (Lacan 2006, 78).

This process, of misrecognizing one's self in the mirror image and image of significant "big" others, creates the ego, the thing that says "I," which is an "ideal I," a perfect whole self who has no insufficiency (Lacan 2006, 76). The fiction of the stable, whole, unified self that we see in the mirror becomes a compensation for having lost the original oneness with the mother's body. For the rest of its life, the infant will for Lacan misrecognize its self as the image in the mirror that provides an illusion of self and of mastery. The idea of a unified Ego and a sense of autonomy is ultimately a fiction created to cope with an original trauma caused by loss of a primal sense of unity as well as of safety and security. The leap from "nature" into "culture" comes with a price: in order to enter culture, the unity with mother's body, the state of nature, is lost; in order to protect ourselves from the knowledge of the loss of being one with mother's body, that is, the state of nature, we misperceive ourselves as autonomous, stable and whole, and not lacking anything (Lacan 2006, 80).

After this self-identification through its own mirror image, the child will gradually understand the place of that image within a larger social and eventually symbolic order. The self enters this order and becomes a speaking subject that can designate itself by "I." The symbolic order, through language, is in essence a contract or tie between an individual and society. Submission to socio-linguistic instances outside ourselves,

such as customs, institutions, laws, mores, norms, practices, rituals, rules, traditions, all intertwined with language, is required in order to enter into the Symbolic order. To become a speaking subject, one have to be *subject-ed* to, one has to obey the laws and rules of language, which Lacan sometimes designates with "the name of the Father": "It is in the name of the father that we must recognize the basis of the symbolic function which, since the dawn of historical time, has identified his person with the figure of the law" (Lacan 2006, 230). Whereas the Real concerns need and the Imaginary concerns demand, the Symbolic is all about desire, according to Lacan (Lacan 2006, 580). At the heart of desire is, as I have indicated earlier, a misrecognition of fullness where there is really nothing but a screen for our own projections, a fantasy that does not correspond to anything in the real – in short, a lack. It is that lack at the heart of desire that ensures we continue to desire. To come too close to our object of desire threatens to uncover the lack that is, in fact, necessary for our desire to persist, so that ultimately desire is most interested not in fully attaining the object of desire but in keeping a distance, thus allowing desire to persist (Lacan 2006, 465, 583).

The Real, the Imaginary and the Symbolic are presented in a chronological order, marking three stages in the psychological development of an infant. Although for Lacan they reflect subsequent dominant dimensions in the early development of a human being, one stage is not simply superseded by another. The Real, Imaginary and Symbolic not only chronologically but also structurally constitute the very fabric of the self, which means that all three continue to exert their influence throughout the life of an adult (see Moyaert 2011).

By subjecting itself to laws and restrictions that control both its desire and the rules of communication, the self enters the Symbolic Order and becomes a subject. However, the ideal, fantasy image that the self has of itself – in its core, a misrecognition as compensation for the loss of a primordial unity in the transition from the Real to the Imaginary order – continues to exert its influence: although other Big Others, like role models, superstars, Party, God, Nature, and Science, might have overwritten the dreams and wishes parents initially projected on the self, the self remains a repository of the conscious and unconscious wants and machination of other speaking beings (Lacan 2006, 230f). Also the Real is never completely absent (Lacan 2006, 255). Indeed, Lacan often says that as far as humans are concerned, "the real is impossible," impossible because the very entrance into language marks our irrevocable separation from the Real (Lacan 1997, 70, 125). Although it cannot be expressed in language, it nevertheless continues to exert its influence, since it is the rock against which all the self's idealizations and societies linguistic structures ultimately fail. The fact that all fantasies and ideal constructions always fail ensures that desire is continuously fueled; desire in the symbolic order could, in fact, be said to be a way to avoid

coming into full contact with the Real, so that desire is ultimately most interested not in obtaining the object of desire, but rather in reproducing itself (Lacan 1997, 14, 68, 76). Lacan often distinguishes between desire and drives. An essential characteristic of desire is its restlessness and insatiability, whereas drives, according to Zizek (1996), derive a perverse enjoyment from this desire-fueled libidinal circling around an unattainable point. Insatiable desire is fueled by ever-wanting drives; there where desire is frustrated, drive is gratified. This way the Thing as object-cause of desire is identified with the realm of drives.

It must be clear that the Imaginary and Symbolic do not coincide: the Imaginary is central to Lacan's account(s) of ego formation and manifests itself in dyadic relations (such as in self and mirror image), whereas the Symbolic constitutes triadic relations by introducing, besides dyadic and intersubjective relation, a trans-subjective symbolic order that normatively regulates the relations between particular beings and society (Lacan 2006, 44, 365, 388). The Imaginary becomes dependent on the Symbolic, insofar as evermore sensory-perceptual phenomena, like images and experiences of one's body, affects as consciously lived emotions or envisionings of the thoughts and feelings of others are shaped, steered and (over-)determined by socio-linguistic structures and dynamics. Similar to Freud, one can recognize in Lacan's account of these three dimensions a tension between "culture" and "nature," since the Imaginary and the Symbolic, often inextricably intertwined, representing stabilizing culture, work in tension with the Real as wild nature.

How can the distinction between culture and nature, derived from the relation between the Imaginary and the Symbolic, on the one hand, and the Real, on the other, contribute to the proposed view of self-formation? And how could it shed light on the role of technologies within this framework? In order to clarify this, we first need to understand how for Lacan the distinction between culture and nature is on a par with the distinction between the general and the singular.

The General and the Singular

The Imaginary and the Symbolic constitute for Lacan the inter- and trans-subjective world that enables the individual to engage with other people and hook into a "general reason." Entering this world marks a separation from the Real, but does not completely abolish its workings. For Lacan, it is by virtue of the Real that an individual cannot be completely absorbed by the general, including ethical, political, social and (we can add) *technological* rules and regulations. Visker says: "it is *that* about my being which makes it *my* being and not someone else's, it is the difference that differentiates me" (Visker 2005, 438). What I cannot grasp, generalize and compare, what I cannot express in language and symbols, is also what makes me different, what makes me a singular

person. Especially in his later works, Lacan extensively discusses the role and significance of the Real in different contexts under different (related) names, such as the Real, the Real Order, the Real Other, the unknowable x, the Thing, the gaze and the little object a (Lacan 1997, 43–87; Lacan 2006, 671–703).

A person is characterized by some-Thing that is not generalizable and cannot be completely reduced to general structures and features. Besides significant others that project their desires and ideals on my desires, and the symbolic and technological order that imparts its rules and regulations on me, my selfhood is also characterized by a "difference" or "singularity" that can never be completely absorbed and nullified by those ideal images, projections and regulations. My mirror image or the functional Magnetic Resonance Imaging (fMRI) image of my brain always misses out on something – a "residue" – that cannot be articulated, nor appropriated; at the same time, what is "left out" makes me the different, unique, singular person that I am. This Thing, this residue remains, as Lacan has pointed out repeatedly, inaccessible, also for the self in question (Lacan 1997, 52, 54). My mirror or fMRI image, and every other possible image of me that I can envision, is incapable of capturing what makes me "me," not only for others but also for myself.

Let me further clarify this line of thought with an example: I am attached to a male sex that biologically does not appear to be very different from the biology of other members of the human species. This male sex is not neutral, but accompanied with images and ideals that others have projected on me ("a man is strong," "a man is rational," "a man is a provider," "a man is also sensitive" etc.), others that often are conveyors of the signs and signifiers of the Big Other, that is, the trans-subjective symbolic order of explicit and implicit laws, regulations and conventions. Significant is that I have not chosen this biological constitution as well as the signs and signifiers that are accompanied by it and translate "sex" in terms of "gender," and, hence, I cannot completely control or possess it. It rather possesses me; notwithstanding that it can "possess" people in different ways, since it is possible to attempt to alter it by, for example, a surgery, or challenge, deny or reinterpret its signs and signifiers.

However, even if I would completely accept my sex and its symbolic connotations, my biological features and their signs and signifiers do not exactly tell me or explain to me what it is to be or how to act as "this" man. On the one hand, I am called to account and take responsibility for the sex/gender that I have not or at least have not completely chosen, and, on the other hand, my sex/gender is not able to completely determine nor articulate what exactly makes me "me" and, hence, makes me different from other people with the same sex/gender. There seems to be always something that escapes the biological order and the symbols that we use to articulate it; in the words of Lacan: "some-thing that cannot or refuses to integrate into a functional totality" (Lacan 2006 78, 345f).

This analysis allows drawing an analogy with technology. We are not only born in a biological and cultural constitution that we have not chosen but that, nevertheless, to a certain extent defines us; we are also born, as Nancy's account of his illness illustrates, in a technological order that we have not chosen, but that nonetheless to a certain extent defines and shapes us. However, similar to the symbolic order, this technological order is not completely able to capture, articulate and determine what makes me "me." I might wear the same brand of clothes, own the same IPhone, attend the same technical university, make use of the same technological infrastructure, use the same technical phrases and think in the same technological grammar as other people, but there is also always, at least potentially, some-Thing in me or attached to me that refuses to completely integrate into a functional technological order. This otherness in or rather of the self, as Lacan indicates, can never be completely captured and domesticated by the symbolic and technological order and the (ideal) images that it projects on the self.

It is important not to interpret the relation between the Symbolic/Technological Other and Imaginary Other, on the one hand, and the Real Other, on the other hand, in terms of a simple opposition between the "symbolic/technological influences and projections of a superficial me" and a "pure, authentic and real me." The Symbolic/Technological/Imaginary Other is not "more other" or "less mine" than the Real Other. Identifying myself as male or an Apple-evangelist, and passionately defending these identifications is not less authentic or more "other" (contra Heidegger) than attempts to distance myself from them. At the same time, I consider myself not only a man or an Apple-evangelist, but also something that escapes every general identification, although I am unable to articulate what this "something else" comprises.

The relation between my three "others" within is and remains uncomfortable: the symbolic and material order and the ideal images that we pursue are – and I repeat what I have said earlier – both the necessary condition for forming a unified, coherent and social self and an obstacle that prevents it from reaching the singularity that it seeks, not only because it has been shaped by them or cannot meet the demands imparted on it, but also because it can never fully appropriate its own singularity.

The relation between my three "others" within is also uncomfortable because the Thing within can be a threat. It not only secures the possibility of liberation from the symbolic and technological order without and within and the desires of others projected on the self, which attempt to fix the self in a particular form, but it can also be a troublesome and dangerous force that unsettles existing configurations, orientations and life plans. On the one hand, the Thing provides individual singularity. On the other hand, this singularity can give way to an eccentricity that can disrupt and even destroy both society and the self (Moyaert 2011, 250).

This explains why society in all its manifestations has, according to Lacan, the secret task to defuse "the Thing," which always bears the dangers of tilting the balance: attempts to socialize humans always hold the chance of excessively repressing and fixing them through a particular "sheltered" system that ultimately becomes a straitjacket and make them ill. Although Lacan stresses the importance of the Symbolic and Imaginary and their protective, orienting and stabilizing workings, he seems to see recognizing the Real as the biggest challenge for every society. For Lacan, the Thing is the dimension that not only makes a human a particular, singular person and ensures plurality and novelty, but also the dimension that makes her irreplaceable and, hence, gives her dignity (Lacan 1997, 112, 118; Moyaert 2011, 250).

The social and technological order, being an inherent part of the self, does not necessarily completely determine the self and the way it forms itself. The self seems to encompass a power or drive that cannot be completely tamed and restrained, though this should not be taken as a comforting idea or a dimension that secures harmony in the self, let alone autonomy. On the contrary! The self is and remains compromised. Lacan proposes an ethics that does not deny and completely neutralize the Thing, but instead finds ways to make room for it and live with it. At the same time, signifying this Thing as a void, a gap, a lack, an abyss around which the symbolic order is structured and a cause that prevents ever fulfilling our desires, leaves us with a tragic conception of the self and self-formation. Against this background, he introduces his notion of sublimation, which I will discuss in the next section and relate it to self-formation.

Self-formation as Sublimation from a Lacanian Point of View

From 1936 onwards, Lacan mentions sublimation in different contexts and explains it in terms that show similarity with Freud's view. In 1954, for example, he characterizes sublimation as working to meet the requirements of the law (Lacan 1988). However, Lacan believes that the notion of sublimation is underdeveloped in the writings of Freud. Although he explicitly takes up ideas he finds in Freud's account of sublimation, his view, as I will show, significantly diverges from the Freudian notion of sublimation.

In *The Ethics of Psychoanalysis*, Lacan, using his own idiosyncratic idiom, defines "sublimation" as the process that "elevates an object to the dignity of the Thing" (Lacan 1959/1960, 112). He develops his view of sublimation in relation to the Symbolic, the Imaginary and the Real, which, as I discussed, are his three essential registers of the human condition. Lacan's definition of sublimation immediately confronts us with a paradox: an imaginary object should be elevated through a particular transformation to represent the Thing that lies outside the signified and,

hence, cannot be imagined. How to represent something that cannot be represented? Sublimation seems in one way or another to allow the symbolic to move closer to the real by means of the imaginary. If the Thing is a void, a gap, an emptiness around which the symbolic order is organized, then to raise an object to the dignity of the Thing in sublimation is to attempt to (symbolically) represent the Thing by something else (in the symbolic order) that will end up being nothing but an emptiness.

In order to make sense of this paradoxical and very abstract articulation of sublimation on the basis of the Symbolic, Imaginary and Real, we need a better understanding of the significance of these dimensions for ethics. Lacan often characterizes the Real as a "primordial field" that is related to the pleasure principle, the mother, and the foundation of ethics:

> Well now, the step taken by Freud at the level of the pleasure principle is to show us that there is no Sovereign Good – that the Sovereign Good, which is *das Ding*, which is the mother, is also the object of incest, is a forbidden good, and that there is no other good. Such is the foundation of the moral law as turned on its head by Freud (Lacan 1997, 70)

In his reading of Freud (and Lévy-Strauss), Lacan explains the prohibition of incest as the taboo that is at the root of becoming a social being, and, hence, ethics, which also marks the step from nature to culture. As Robertson (2015, 86) explains:

> The prohibition of incest forces the self to employ its drives for the benefit of society. Insofar as the incest prohibition casts the mother out of the subject's field of possible object-choices, it promotes her to the function of the object-cause of the subject's desire.

The prohibition of incest coincides with the introduction of the moral law but, at the same time, transforms the incestuous object into the primordial agent of all desire. The mother functions, Lacan says, as an object that desire organizes itself "around" (Lacan 2014, 106). The original loss of the mother remains the hidden drive behind desire as well as the lack that ensures the continuation of desire. A certain distance from the Thing is required and maintained by the prohibition of incest, which keeps the self from and keep it near the original void, and that way, keeps its desire going. Sublimation concerns organizing desire "around" the original void in such a way that we remain close enough to the void to continuously fuel the desire and not so close that we are crushed by it.

Lacan attempts to recognize both the value of a moral law and the desire to undermine it. In fact, both are strongly interrelated: order produces the perverse desire to undermine order. For Lacan, without the

symbolic order, desire cannot be channeled and, hence, exist. And at the same time, by channeling the desire, the desire to transgress that structure and order is initiated. It is not a coincidence that Lacan sometimes equates the Thing with sin, since the Thing refers to the possibility of transgression of the moral law:

> Is the Law the Thing? Certainly not. Yet I can only know of the Thing by means of the Law. In effect, I would not have had the idea to covet it if the Law hadn't said: "Thou shalt not covet it." But the Thing finds a way by producing in me all kinds of covetousness thanks to the commandment, for without the Law the Thing is dead. But even without the Law, I was once alive. (…) I believe that for a little while now some of you at least have begun to suspect that it is no longer I who have been speaking. In fact, with one small change, namely, "Thing" for "sin," this is the speech of Saint Paul on the subject of the relations between the law and sin in the Epistle to the Romans, Chapter 7, paragraph 7 (Lacan 1997, 83)

The Thing always marks a form of transgression. It is predicated upon prohibition. The law regulates but also offsets desire in such a way as to increase the sense of temptation and going beyond traditional morality. The Thing is beyond good and evil, although it is usually figured as evil by conventional morality – the parallels with Nietzsche's critique of morality are quite obvious, which, however, I will not discuss here.

Now the difference between Freud and Lacan's notion of sublimation becomes clearer. In line with Freud, Lacan conceives sublimation as a "form of satisfaction of the *Triebe*" (Lacan 1997, 110). However, he does not conceive sublimation as redirection of the libido toward a nonsexual object, as providing a socially accepted substitute for sexual gratification. Sublimation for Lacan does not relate to a sublime moral beauty that supports traditional ethics, but rather concerns symbolically organizing our drives in such a way that the Real, which is the emptiness that the original love object has left behind, is not evaded but enabled to exert its influence. In organizing and satisfying our drives, we need to attempt to revisit the original object of desire and, at the same time, to somehow affirm it as the empty center (and lack) that fuels desire. Lacan's credo "don't give up on your desire" is no libertarian slogan, but expresses the idea that desire can only be sustained and fueled if the Thing as cause of desire is recognized and affirmed, which always comes with the possibility and danger of transgression.

Lacan also sometimes relates sublimation to Freud's notion of the death drive (Lacan 1997, 2015–2017), but again reframes it in such a way that it cannot be simply reduced to Freud's view. Freud explains the death drive as a fundamental desire inherent in all organic life to restore an inorganic unity, which will end all suffering and resolve all

tension. Lacan takes up this Freudian idea of the death drive as restoring a primordial unity but relates it to the unity in the order of the Real. For Lacan, the death drive is desire carried to the extreme, to the absolute limit of desire, beyond pleasure (*jouissance*). It is the unleashing of absolute animal desire without any constraint or law, which leads ultimately to destruction and self-destruction, that is, to death. However, this drive to destroy everything also has a "positive" side. The death drive should be regarded as a "destruction drive" because "it challenges everything that exists. But it is also a will to create from zero, a will to begin again" (Lacan 1997, 212). In this way, Lacan connects sublimation to the death drive and a *creatio ex nihilo*:

> [T]he notion of death drive is a creationist sublimation, and it is linked to that structural element which implies that […] there is somewhere […] beyond [the signifying] chain, the ex nihilo on which [that chain] is founded and is articulated as such (Lacan 1997, 212)

In representing the Real, sublimation operates as creation from nothing, from emptiness, within the Symbolic. Sublimation "is precisely that which reveals the true nature of the Trieb [drive] insofar as it is not simply instinct, but has a relationship to das Ding as such, to the Thing insofar as it is distinct from the object" (Lacan 1997, 177). Sublimation is symbolic creation fueled by the Real and pushed to its limits, beyond traditional moral constrains and considerations of good and evil. In short, sublimation is the creation of the singular within the symbolic.

How can this Lacanian notion of sublimation contribute to the further development of the proposed notion of self-formation? First of all, we need to acknowledge that a self that attempts to form itself is already subject to the laws, mores, customs and restrictions of an Imaginary and Symbolic Order that controls and regulates its desires and conduct. Without the images imparted by others and the laws and morals imposed by the Big Other, the self would not be able to form an ego and become a subject in the first place. Its desires are to a great extent a repository for the projected desires and fantasies of significant others. However, the Symbolic Order and its laws and morals that channel the desires of the self and impart a form on it, by its very existence, initiates the possibility of transgressing that form. If the Real is in its purest form disorientation and disruption, then organizing our drives "around a void" implies regulating and imposing a form on them without destroying the possibility of going against the status quo and socially accepted laws and morals. As Thurston formulates in the context of sublimation within art: "sublimation is Janus-faced poised between the raw singularity of the drive and the polite decorum of the art gallery" (2003, 30). The inherently paradoxical challenge of sublimation involves recognizing that forming ourselves is impossible without subjecting ourselves to the laws, norms,

values and ideals imposed by society (the Imaginary and Symbolic) and, at the same time, affirming and fueling a force (the Real) that can never be completely appropriated by society and ourselves, a force that can challenge and disrupt every possible standard for self-formation, which is a requirement for the emergence of novelty and the formation of a singular self (see also Zizek 1989).

Lacan often turns to art to illustrate the importance of emptiness and the paradoxical challenge of sublimation: "in every form of sublimation emptiness is determinative. […] All art is characterized by a certain mode of organization around this emptiness" (Lacan 1997, 129–130). Various examples of "organization around a void" are provided: a building that revolves around empty spaces, painting on an empty canvas, the writing of poetry in the absence of love, the rituals of the Medieval France troubadour in his longing for the unattainable Lady (see Lacan 1997, 161–164). Maybe Lacan's most prominent example is the shaping of a vase on a potter's wheel:

> If [a vase] really is a signifier, […] it is in essence a signifier of nothing other than of signifying as such, in other words, of no particular signified. […] It creates the void and thereby introduces the possibility of filling it (Lacan 1997, 120)

The idea of creating a void that solicits filling it also articulates Lacan's view of sublimation as "creation ex nihilo" (Lacan 1997, 115–127; see also James 2009). The Thing represents a creation out of nothing because it is a signifier of signifying as such. The Thing is not the object (the vase) but the emptiness that is represented by the object in order for it to function in symbolic discourse. The vase is, in the words of Lacan, "an object made to represent the existence of the emptiness at the center of the Real that is called the Thing, this emptiness as represented in the representation presents itself as a *nihil*, as nothing" (Lacan 1997, 121). For Lacan, genuine art is able to symbolically represent an emptiness that allows filling it, at least potentially, with something that did not exist before and, hence, transgress the status quo and accepted values, norms and ideals. Genuine art exemplifies how remaining close to the emptiness of the Real allows stretching the limits of the symbolic and making room for the emergence of the terrifying beauty of the sublime.

9.3 Sublimation as Imposing Form to Great Diversity

We have seen that Freud attempts to separate sublimations from neurotic symptoms by taking social valuations as a demarcation criterion; sublimation is conceived as the capacity to substitute raw instinctive gratification for culturally valued representations. Freud's view makes, to a certain extent, sense: on the one hand, it does justice to the fact that we are born, develop and form ourselves in a society that imposes on us

certain values, norms and ideals as well as to the experience that what we consider normal and abnormal, healthy and ill and enhanced and abled is greatly dependent on those societal values, norms and ideals; on the other hand, it acknowledges the constant tension between what society expects and what uncontrollable libidinal drives that seek psychophysiological gratification demand. Problematic in Freud's view is that it makes sublimation very dependent on what a particular society happens to find acceptable, which not only could lead to repression and arbitrariness, but also seems to disregard the importance of enabling individuals to develop themselves into singular persons.

As we have seen, Lacan takes up the tension between the cultural and the natural on the basis of his three dimensions of the Real, the Imaginary and the Symbolic. He highlights that the Imaginary and the Symbolic cannot be simply externalized and viewed as sociocultural influences that repress the individual, since they are a constitutive condition for developing a unified ego as well as moral and communicative subjectivity. At the same time, he emphasizes the importance of a "singularity" that by virtue of the Real could (at least potentially) and should never be completely absorbed and nullified by ideal images, projections, laws, values and regulations imparted by society. In Lacanian sublimation the tension between culture and nature is strongly intensified and situated in the heart of self and self-formation: the pre-existing imaginary and symbolic order is both a necessary condition for forming a stable social self and an obstacle that prevents it from reaching the singularity that it seeks. Sublimation refers to the paradoxical challenge to (re)discover a void or emptiness and organize our drives around it in such a way that it becomes possible to stretch societal boundaries, without abolishing them. This way, a singular self can be formed within the social order.

We came across the tension between indeterminateness, chaos and struggle, on the one hand, and regulation and organization, on the other, earlier in this book, though without explicitly relating it to the notion of "sublimation," namely in my discussion of Nietzsche and Peirce, the two main guides in the process of developing a non-essentialist, non-dualist and non-relativist view of self-formation that takes into account the influence of technologies on the self. Freud and, more aptly, Lacan's notions enable framing the proposed notion of self-formation, which has been strongly inspired by Nietzsche and Peirce, explicitly in terms of sublimation. Moreover, these approaches provide complementary ingredients that enable finalizing the development of the proposed notion of self-formation in a highly technological world beyond the inside-outside dualism.

Sublimation as Organized Struggle

In my discussion of Nietzsche's ontology of the will to power, we have seen that, besides "struggle," Nietzsche extensively focuses on "organization" and offers what I have called an Organization-Struggle Theory (OST)

that enables analyzing and evaluating will to power unities in terms of strong and weak or healthy and sick. Because for Nietzsche "The world seen from inside (…) would be just this 'will to power' and nothing else. –" (BG 36; see also *Z* II: "Self-overcoming"; BGE 13), what he says about reality on an ontological level could be also applied to human self-hood. In fact, Nietzsche applies an early version of this theory to humans and explicitly relates it to the notion of "sublimation," though in that phase in his work (and also often later) instead of "wills to power," his main ingredients are "drives" and "instincts" (KSA 9.211; 9.273).

To what extent Nietzsche's scattered ideas regarding and related to sublimation have influenced Freud and Lacan is difficult to establish and is not very relevant for my investigation. What is relevant is that Nietzsche's OST can now be explicitly interpreted in terms of sublimation and contribute to the development of this notion in a manner that takes into account the tackled issues and challenges regarding self-formation. First of all, Nietzsche further radicalizes the problem as formulated by Freud and Lacan: not only is there a tension between, on the one hand, a stable social order that enables regulating and organizing human drives, and, on the other hand, wild libidinal drives that are continuously fueled by something that transcends the symbolic order and can never be completely neutralized, but Nietzsche also foresees that the social realm of laws, morals, customs and regulations is increasingly incapable of fulfilling its function of organizing that realm:

> In an age of disintegration where the races are mixed together, a person will have the legacy of multiple lineages in his body, which means conflicting (and often not merely conflicting) drives and value standards that fight with each other and rarely leave each other alone. A man like this, of late cultures and refracted lights, will typically be a weaker person (BG 200)

For Freud and Lacan, the social order offers order. It can function as a counterpart to disruptive libidinal drives. Conversely, one of the most important strands in Nietzsche's work is the erosion of stable value systems and grand narratives. The modern human is becoming a disorganized collection of competing drives and value systems, incapable of forming herself into a strong, healthy and fruitful composition.

For Nietzsche, Christian morality has been extremely successful in organizing conflicting drives and value systems, though by denying and splitting of (aggressive) drives that are found unacceptable to a degree that one does not even acknowledge that one has such drives. This repression and splitting off of certain drives have led to stabilization, socialization and cultivation, although at the cost of making the human

a weak and ill creature: in Nietzsche's words, a herd animal. Nietzsche foresees that modern humans, including those who have become indifferent toward the Christian God, lack any power and creativity to impart new and more healthy forms to themselves; due to this impotence, Nietzsche predicts, they will gradually disintegrate.

OST indicates that only the combination of strong organization and intense struggle is a trait of strength and health. From this theory, weakness can manifest itself in two types: through achieving a high degree of organization by excluding all struggle or through harboring intense struggle without great organizational force. Both types of weak "will to power"-organizations are unfruitful, since they are incapable of bringing about radical novel life forms. Although the type that contains little divergence and struggle can sustain itself for a long time, it is unproductive and in the long run liable to a process of disintegration; because it has no potential to generate, it is ultimately life negating and life threatening. For Nietzsche, this is what Western culture is facing in the aftermath of Christianity. The strong or healthy "will to power"-organization that Nietzsche is promoting is characterized by considerable divergence and struggle that is forced into a unity in a structured manner.

In contrast to Freud and Lacan who were primarily focused on common individuals and their development in a (often anonymous) culture, Nietzsche was much more concerned with very particular cultures and value systems and analyzed what made one culture weak and ill and another vibrant and healthy (e.g., the Christian versus the Greek) as well as with the role great individuals played and could play in the growth and decay of cultures. He did refer to individuals, but often only as opportunities or models that could elevate Western culture. At one point, he took, for example, Wagner to be such a case, or as explicitly displayed in *Ecce Homo*, he also took himself as an example of a strong type (EH, 96–98). However, Nietzsche's elitist approach and his focus on great exemplars or geniuses does not have to prevent deriving ideas that can contribute grasping what could be strong or healthy self-formation with regard to all humans in a democratic society. It is clear that the strong or healthy type for Nietzsche is characterized by redirecting drives in such a way that they contribute to the development of a more encompassing unity without amputating them:

> a cultural edifice in the single individual will have the greatest similarity to the cultural architecture of whole eras and, by analogy, provide continuous instruction about them. For wherever the great architecture of culture developed, it was its task to force opposing forces into harmony through an overwhelming aggregation of the remaining, less incompatible powers, yet without suppressing or shackling them (HH I, 276)

Nietzsche stresses the importance of redirecting and organizing weaker drives without shackling and annihilating them:

> Overcoming of the affects? No, if that means their weakening and annihilation. But instead employing them; which may mean a long tyrannizing of them… At last they are confidently given freedom again: they love us as good servants and happily go wherever our best interests lie (KSA 12:1[122])

The distinction between repressing drives and reorienting their direction and function enables contrasting what Nietzsche would call "ressentiment" with sublimation, which enables highlighting an important feature of the latter. From a Nietzschean point of view, sublimation refers, as Gemes has pointed out (2009, 48), to substituting a drive's primary aim for a secondary aim that is consonant with the orientation of a master drive without destroying or diminishing its energy. Richardson puts it in a logical terminology: "Drive A rules B insofar as it has turned B towards A's own end, so that B now participates in A's distinctive activity" (1996, 33). Conversely, repression refers to denying and blocking a drive's primary aim by splitting it off from other drives and not integrating its aims with the aims and orientation of other drives; as a consequence, this drive will be suppressed and shackled, and will not have the opportunity to express itself (Gemes 2009, 48). For Nietzsche, resentment is the result of forcing a drive to stifle and denying it, that is, repressing a drive. Gemes puts this distinction in more psychoanalytic terms: "sublimations involve integration or unification, while pathological symptoms involve splitting off or disintegration" (2009, 48). This example illustrates how successful sublimation should be understood – Leonardo, instead of stifling his homosexual drives, redirects the literal possession of male bodies toward the secondary aim of expressing this possession in idealized representations of male bodies. It must be clear that this example can only be appreciated if one disregards the problematic allusion of taking again particular social valuations as demarcation criterion and presupposing that there is something wrong with directly acting on homosexual urges by having sexual relations with other males.

Nietzsche's appraisal of integration, instead of repression, adds a significant element to the proposed notion of sublimation, namely "affirmation," which is also the antidote for resentment. The idea of affirming one's life, affirming both the harsh facts of one's life and one's never completely controllable drives, runs throughout Nietzsche's whole work. It must be clear that Nietzsche does not promote a type of *passive* affirmation: it is not about resigning in our fate or simply letting all of one's drives have free expression. That would not result in a self characterized by organized struggle, but would rather lead to unbound internal conflict, chaos and, inevitably, disintegration. Nietzsche defends

active affirmation: it is about actively shaping the self by integrating all, and sometimes conflicting facets of one's life in a preferred lifestyle and harnessing one's drives to allow them strengthening one another that leads to a form of concerted expression in a particular direction. In Nietzsche's words:

> The multitude and disaggregation of impulses and the lack of any systematic order among them results in a 'weak will;' their coordination under a single predominant impulse results in a 'strong' will: in the first case it is the oscillation and lack of gravity; in the later, the precision and clarity of direction (KSA 13:14[219])

In contrast to Freud, the Nietzschean notion of sublimation does not involve redirecting repressed sexual drives toward a nonsexual, socially valued aim. The erosion of traditional value systems makes employing social valuations as demarcation criterion for sublimation unattainable. Moreover, for Nietzsche, desexualization is not a dimension of sublimation, though redirecting and integrating aggressive drives, which otherwise would manifest themselves in violent sexual outburst, in such a way that they strengthen a particular constellation could be part of the process of sublimation.

Yet another significant difference between Freud and Lacan, on the one hand, and Nietzsche, on the other, is that for Nietzsche, the realm of drives is not surpassed by other (higher order) realms (an Ego and Superego or an Imaginary and Symbolic Order). In Nietzsche's words:

> While "we" believe we are complaining about the vehemence of a drive, at bottom it is one drive *which is complaining about the other*; that is to say: for us to become aware that we are suffering from the *vehemence* of a drive presupposes the existence of another equally vehement or even more vehement drive (D 109)

This implies that also so-called higher-order values and laws are manifestations or symptoms of a certain drive having dominated other drives. In his later work, this struggle between drives is elaborated, as we have seen earlier, in terms of will to power relations.

These differences should not conceal similarities between Nietzsche and, especially, Lacan's view of sublimation. Lacan's idea of "organization around a void" has resemblances with Nietzsche's view of organized struggle. Both stress the importance of imparting to drives a form and simultaneously enabling the possibility of change, transgression and emergence of something novel. Lacan understands the tension between organization and stability, on the one hand, and the emptiness that can give birth to novelty, on the other, as a tension between the law and desire that can never be completely fulfilled. What is at stake is the

formation of a singular self without completely disrupting and destroying both society and the individual self. For Lacan, sublimation is, as we have seen, the creation of the singular within the symbolic. Nietzsche does not characterize sublimation in terms of a conflict between the law and wild drives. For him, societal configurations, values and laws are also expressions of certain drives having surpassed and organized other drives. In contrast to Greek culture, Christian culture, in his view, does not combine strong organization with intense struggle (as expressed in Greek, *Wettkampf*), which does not foster the emergence of new life forms. Nevertheless, the mark of the healthy and strong, and, hence, successful sublimation, is, for both Nietzsche and Lacan, organizing drives in such a way that they possess the potential to bring about new life forms.

Nietzsche and Lacan provide different ingredients that enable further elaboration and operationalization of the idea of self-formation in terms of sublimation. On the one hand, Nietzsche specifies the conditions that enable the emergence of new life forms: the combination of strong organization and intense struggle. Hence, sublimation requires organizing intense internal struggle into a form that has the potential to bring about new life forms. A decadent or nihilistic type of self-formation can be seen as the antonym of self-formation as sublimation. The decadent is a type of will to power that is unable to organize the conflicting variance within itself. We have seen earlier that Nietzsche defines decadence as the "[i]ncapacity for organic formation," which is further explained as a physiological deficiency of organizing force, which manifests itself as a becoming independent of the part with respect to the whole (Müller-Lauter 1999a, 4). This incapacity of organizing itself is resolved by the weak by negating and repressing the parts that it is unable to integrate. Nietzsche's predominant example is the Christian, who is unable to integrate the harsh facts of life and aggressive drives, and "resolves" this incapacity by negating and repressing and resenting these factualities and drives with the help of a fabricated, "true" reality after death (Müller-Lauter 1971, Chapters 2–4). The decadent imposes a form on itself by a weakening or restraining internal struggle. That internal struggle is, however, at the same time, necessary for growth and bringing about novelty. Since a "will to power"-organization only exists and grows by virtue of struggle, excessive organization has disintegrating consequences in the long run. In contrast to decadence, self-formation as sublimation from a Nietzschean point of view highlights the importance of actively affirming all occurrences in one's life, bitter and sweet, ugly and pretty, allowing them to strengthen one another by integrating them in a particular lifestyle. This emphasis on *active affirmation and integration* cannot be found explicitly in Lacan and adds an important aspect to the notion of sublimation.

On the other hand, Lacan, first of all, helps to interpret Nietzsche's scattered ideas that I earlier tried to structure in OST in terms of

"sublimation." He also relates more strongly than Nietzsche sublimation to self and self-formation, and enables this way including ideas derived from Nietzsche to the proposed notion of self-formation as sublimation. More substantively and thematically, Lacan highlights the importance of the conflict between the general that is depicted as societal expectations, values and laws, on the one hand, and the singular that is made possible by transgressing the general by virtue of a void that can never be filled, on the other. Moreover, he stresses the compromised status of the self, since the self *is* constituted by an Imaginary and Symbolic order that secures individual and social stability as well as by the order of the Real that is responsible for bringing about novelty.

Although Nietzsche's diagnosis of traditional narratives, values and laws losing their dominance and their capacity to regulate forces within a society is very important, it does not explicitly relate the fact that we are still born in a certain society with an existing "Imaginary and Symbolic Order," which constitutes us, to the challenge of self-formation as sublimation. Maybe Nietzsche's idea of "active affirmation and integration" could also be related to societal narratives, values and laws: from that perspective, they would also need to be actively affirmed and integrated in the process of self-formation as sublimation. However, neither Nietzsche nor Lacan relate these elements to one another and explain how this possible affirmation and integration should be comprehended. In fact, their mistrust of the repressive potential of society often brings them very close to one another and possibly reveals the influence of Nietzsche on (Freud and) Lacan. This mistrust also explains why both Nietzsche and Lacan primarily focus on making room for novelty and singularity through social transgression, since they both believe that attempts to socialize humans always hold the chance of excessively repressing and fixing them through a particular "sheltered" system that ultimately becomes a straitjacket and makes them ill. Although Lacan stresses the importance of the Symbolic and Imaginary and their protective, orienting and stabilizing workings, he remains a Nietzschean by stressing that recognizing the Real is the biggest challenge for every society, since the self ultimately derives its dignity from the workings of the Real.

The emphasis on the importance of struggle, transgression and novelty does not imply that Nietzsche and Lacan completely oversee the importance of social organization. In fact, there is also another type of decadence in Nietzsche that could be considered an antonym of sublimation. Internal struggle is, as I have discussed earlier, a condition for a "will to power"-organization to become strong and remain healthy. When, however, the internal struggle cannot be organized in such a way that tension is generated that can be discharged in a certain direction, decline occurs. The ruling drive is then incapable of organizing the internal struggle. Although struggle is necessary for gathering more power,

it contains, at the same time, the danger of decadence. The greater the struggle, the more difficult it is to organize it, the greater the chance of decay. Lacan, as I have discussed, stresses that she self also needs to protect itself from the abysmal and uncanny Thing /Real and can only create something new within the Symbolic Order if it keeps a distance from and engage in "organization around a void." Both Nietzsche and Lacan depict the paradox inherent in the process of self-formation as sublimation in different ways, but they do not sufficiently value and elaborate the constitutive contribution of society. Although they both recognize that society provides goals and values that regulate the individual's drives and desires, they do not seem to consider the possibility that these goals and values do not have to be necessarily restrictive but could also foster, and even be a necessary condition for change, novelty and singularity. Next, I will show how Peirce's notion of the ethical-esthetic ideal could fill this gap and further nuance and enrich the notion of sublimation.

Sublimation as Organized Heterogeneity by Virtue of Social Goals and Ideals

Peirce attempts to illustrate (as I discussed in Chapter 2) that in our encounters with the world, whether we are awake, sleeping or hallucinating, we always and necessarily have to adopt or presuppose three phenomenological categories which he simply calls the categories of Firstness, Secondness and Thirdness. These categories subsequently correspond with the elements of novelty, brute interactions and regularity. Compared to Nietzsche and Lacan, Peirce provides a more congealed and specified account of especially the aspect of regularity on the basis of his category of Thirdness as well as how regularity relates to the two other dimensions. We have seen that, according to Peirce, self-formation is only possible if we impart a form to ourselves by virtue of an orientation toward certain goals and ideals. Self-formation is, as I explained earlier, depicted as a challenge, as something that is not simply given but should be realized. For Peirce, goals and ideals are imperative in the process of self-formation. Without goals and ideals, which we derive from the society that we happen to live in, imparting a form to ourselves is impossible.

Since the orientation toward goals and ideals is a normative challenge, the question is not only how self-formation is possible, but also how we ought to form ourselves. Peirce acknowledges that people, depending on the society in which they live, devote themselves to different and possibly opposing goals and ideals, which nevertheless enable them to organize their interactions and to develop themselves in a certain direction. In addition, these goals and ideals are established and continuously assessed and adjusted in concrete interactions with the world: new scientific findings, unpredictable events or unexpected experiences (Firstness) could destabilize certain conceptions and challenge established habits, which

prompts the adjustment of pursued goals and ideals. For this reason, Peirce stresses that goals and ideals are virtual, since they are always situated in the future and always subject to adjustment and correction. This approach indicates that Peirce does not ignore the Nietzschean view that societal values and norms are not fixed but subject to change and disruption, both synchronically and diachronically. However, he, in contrast to Nietzsche, believes society, under certain conditions, can evolve and gradually dissolve aggressively contesting goals, values, ambitions and aspirations as well as enable critical self-formation. Instead of posing the individual *against* society, he considers individuals as "cells of a social organism" (CP 673), who discover and develop themselves in an interaction with their environment.

Peirce conceives self-formation not as an isolated, individual challenge, but situates it in a social environment that can constructively contribute to evermore apt interactions with and conceptions of what we encounter. In the process of forming ourselves by virtue of certain ideas, goals and ideals, we attempt to persuade "a future, critical self," which is not another individual person, but a personified community, that is, a virtual interlocutor who can question my ideas, goals and ideals continuously from different points of view; this implies that the social is always present in the individual.

Although Lacan and, indirectly, also Nietzsche, take into account the influence of society on self-formation, both of them consider this influence primarily as restricting the potential imbedded in the (ever-wanting, wild drives of the) individual and, hence, a danger for self-formation. In contrast, Peirce does not evaluate society as such as a danger, although some societies could be perverse, constraining and suffocating; he rather considers personified community as a "critical self" that is a necessary condition for enriching and improving self-formation. Without goals and ideals that transcend immediate individual feelings, drives and actions, self-criticism could not be genuine, since real criticism requires standards that cannot be reduced to the individual's preferences; otherwise, it would be a bit like having a butcher inspecting the quality of his own meat. Those ideals are also subject to criticism, but again that is only possible in the light of yet a higher ideal that is pursued and so on and so forth. It must be clear that in order to be able to challenge societal goals and ideals, favorable conditions need to be created. From a Peircean perspective, a healthy, open society is, as we saw earlier, characterized by built-in mechanisms and tools (such as democratic regulations, informed consent, the possibility to opt-out etc.) that enable people to deliberately and critically identify themselves or not-identify themselves with certain needs and desires, including the goals and ideals that society finds preferable. Crucial in this line of reasoning is that criticism does not exclude but rather requires anticipating standards and ideals that transcend individual, particular preferences. Ultimately, we

have to appeal, Peirce says, to an ultimate ideal that is in itself admirable, "a habit of feeling which has grown up under the influence of a course of self-criticisms and of heterocriticisms" (CP 1.573f). For Peirce, this ultimate ideal that gradually manifests itself under the influence of a course of criticisms from different directions is mirrored in a perfectly cultivated emotional life form.

As far as I know, Peirce never uses the term "sublimation" to characterize this perfected type of self-formation. However, his characterization of this ultimate ideal shows striking similarities with Nietzsche's and Lacan's depictions of sublimation. Let me briefly recall those passages found in Peirce's writings:

> I should say that an object, to be esthetically good, must have a multitude of parts so related to one another as to impart a positive simple immediate quality to their totality; and whatever does this is, in so far, esthetically good, no matter what the particular quality of the total may be (EP II, 201)

Developing a sense for the esthetic ideal entails forming, one could say, a sublime practical and emotional disposition. This general disposition enables the orientation toward goals by virtue of which evermore meaningless, useless and inefficient parts of the social organism are related to one another in such a way that they become meaningful, useful and efficient. Instead of denying, repressing or cutting off parts of the organism that do not seem to fit, the challenge is to gradually integrate them and use them to strengthen and further develop the organism. We can recognize here the idea that sublimation involves, instead of repression, integration, which we have found through Freud and Lacan in Nietzsche. Peirce also stresses the idea that this integration of parts within a whole, which from a Nietzschean perspective was elaborated in terms of organizing intense struggle, should not be achieved at the expense of multitude and diversity: "See that self-government is exercised; but be careful not to do violence to any part of the anatomy" (M 675, 15–16; see also Colapietro 1989, 111f).

From a Peircean perspective, diversity and regulation do not exclude one another, but diversity can only be preserved and intensified *by* submitting it to a form or idea. He sometimes uses the terms "evolutionary love" or "creative love" to indicate how diversity and regulation could go together. I recall a significant passage in his works that attempts to explain how this is possible:

> The movement of love is circular, at one and the same impulse *projecting creations into independency* and *drawing them into harmony.* This seems complicated when stated so; but it is fully summed up in the simple formula we call the Golden Rule. (…) It

is not by dealing out cold justice to the circle of my ideas that I can make them grow, but by cherishing and tending them *as I would the flowers in my garden.* (…) Love, recognizing germs of loveliness in the hateful, gradually warms it into life, and makes it lovely [my italics] (CP 6.288f)

Instead of exercising brute force (Secondness) that would restrict and confine, the "movement of love" evokes, arouses, infects (Thirdness) a thing to realize its potential, to develop itself from something homogenous and undifferentiated to a well-organized heterogenous and differentiated form. We can recognize a Nietzsche-like concept of "amor fati" in Peirce's idea of not only accepting the "hateful," but also transfiguring it to constructive and meaningful life. Applied to the self, we can say that sublimation is a type of self-formation that involves not restricting or cutting off parts of the self, but rather accepting and transfiguring even the meaningless, useless and hateful parts, and stimulating the realization of their constructive potential by imparting on them a form by virtue of a certain ideal. This explains why ideals that are more inclusive are more admirable than those that are less inclusive, and why the ultimate ideal that can include all parts of the organism and prompts them to constructive, maximum self-realization is the most admirable ideal.

The idea of sublimation as imposing form on multitude and diversity and as actualizing potentialities, we also can find in Nietzsche and Lacan. Peirce can add to this idea that it is the orientation toward goals and ideals and eventually toward an ultimate ideal that can bring about maximum inclusion, regulation and self-realization. Moreover, for Peirce it is crucial that the orientation toward the ultimate ideal, which is a regulative ideal projected in the future, is not an individual but a social enterprise that takes place within an endless community of critical minds, an ideal that gradually brings about "a habit of feeling which has grown up under the influence of a course of self-criticisms and of heterocriticisms." Self-formation as sublimation, one could say, necessarily entails seeking beyond our actual, individual preferences and becoming part of a bigger, even a cosmic process. For this reason, I have indicated earlier that Peirce could be considered an anti- or posthumanist philosopher, since he does not take the human being to be the final product of evolution. The same could be said about Nietzsche, as have been illustrated by his notion of the *Übermensch*.

9.4 Technological Sublimation Theory

Instead of conceiving self-formation as "enhancement" in terms of "faster, stronger, smarter," I have proposed that it is more adequate and feasible to understand it in terms of "sublimation." Or alternatively, "enhancement" could be reinterpreted in terms of "sublimation." If we bring

together the different and most important aspects of "sublimation," derived from ideas in the works of Nietzsche, Peirce, Freud and Lacan, and take into account the role that technology plays in self-formation, we can sketch the contours of the view of self-formation as sublimation in a technological environment. This configuration results in a theory that I will call "Technological Sublimation Theory" (TST).

Intrinsically bound to the proposed view of self-formation as sublimation, there are two fundamental caveats: first of all, technology has increasingly become "extimate," that is, it has become an "other within" that we cannot simply externalize and use as a neutral means to form ourselves. Being outside of ourselves implies that we are to a certain extent outsiders for ourselves, strangers to ourselves. Not only do we possess technology, but technology also possesses us, since it increasingly has, besides language, laws, rules and regulations, become a permanent structure of our very selfhood. Self-formation as sublimation entails that the self is and always remains compromised: technology both enables and constrains self-formation. Technology as "other within" has become both the necessary condition for forming a socially dependable, consistent and stable self and, at the same time, an obstacle that prevents it from reaching the singularity that it seeks.

Second, sublimation entails that there are no univocal standards for self-formation. The anti-essentialist self beyond inside-outside dualism is formed in interaction with and derives the goals and ideals that enable imposing a form to itself from its environment. Since environments can be very different and are subject to change, the standards for what can be considered "good self-formation" or "enhancement" also show copious variations, both diachronic and synchronic. Since our environments are increasingly technological environments and developments in technology are gaining serious momentum, self-formation may result, as I have discussed, in self-transformation. However, radical transformation through technology does not simply enable us to become "stronger," "smarter" or "healthier," but it can and often will also influence norm and disrupt the very standards for establishing what is "good self-formation" or "self-enhancement." In other words, new and emerging technologies are not neutral means but often bring about different and, from our current perspective, foreign standards for determining what are "normal" and "enhanced" capacities. Moreover, technology not only influences the immediate goals and ideals that are used to regulate our thoughts, actions and habits, but also higher-level goals and ideals that are used to criticize, adjust and validate lower-level goals and ideals. This is an implication of technologies becoming extimate; they are no longer external factors that influence an inner, autonomous realm but are able to affect the very standards that guide our conduct, shaping our most intimate and cherished convictions and beliefs. Therefore, establishing universal or univocal standards for what is considered "good

self-formation" or "enhancement," independent of our particular technological surroundings, is impossible.

These caveats do not render "good self-formation" impossible. In fact, understanding self-formation as sublimation is motivated by the aim to indicate what is "good self-formation" without endorsing essentialist and dualist views of the self that consider the self as an independent and invariable unity, disconnected from its dynamic environment. Against this background, self-formation as sublimation can be characterized under three strands: *what, why* and *how*. Before summarizing the most important aspects of the *what, why* and *how* of self-formation as sublimation, I need to stress that these distinctions are artificial because sublimation is as such a process without a clear-cut beginning and end, and cannot be simply framed in terms of isolated steps directed at realizing a static aim. This categorization attempts to make the complex notion of sublimation graspable by structurally highlighting its most important dimensions and phases, which in reality are strongly interwoven.

The *what* of self-formation as sublimation attempts to characterize in general what is self-formation as sublimation, what are its general characteristics. Being "extimate" and "closed open" entails that, as we have seen, I find in myself an "other within" that I have not chosen and, at the same time, cannot simply detach myself from but rather have to relate myself to. In fact, this "other within" involves a double alterity within, since it contains both our symbolic and material order, which provides regularity and stability and some-Thing, corresponding to the realm of drives that are feeding off the dissatisfaction of desire, which is a source of instability but can also soak off rigid constraints and enable singularity. Among this symbolic and material order, we find, among other dimensions, "our" name, skin color, ethnicity, body, customs and laws, but also, and increasingly, "intrusive technologies" that are ever-more-strongly shaping our very selfhood. Instead of denying and rejecting this double alterity within, sublimation entails *accepting* the "other within," since denying it would result in a kind of contradiction: the self would reject what made it what it is. With regard to technologies, this means that we cannot simply externalize the technologies that are shaping our self and evaluate them as detached entities but that sublimation entails accepting in a particular way these intrusive and extimate technologies.

If this would be the whole story, sublimation would be nothing else than passively resigning to our fate. However, sublimation does not simply call for accepting the "other within," but also *integrating* it in a life form or lifestyle that we could endorse and find admirable.[1] Sublimation entails making ours what is not completely ours and taking responsibility for what we are not completely responsible for. We could say that we need to *actively* accept the "other within," which is only possible if we

deliberately and critically integrate it, that is, make it part of, incorporate it in a fitting life form or lifestyle. With regard to technologies, this means that sublimation entails critically integrating "intrusive technologies" in an appropriate-for-us life form or lifestyle. Actively "accepting" and "integrating" the "other within" could be captured under the heading "appropriating." Self-formation as sublimation, hence, entails appropriating our symbolic and material-technological order as well as our ever-wanting drives and insatiable desire.

The *what* provides a general characterization of self-formation as sublimation, but does not yet make clear *how* sublimation could be realized. The *how* of sublimation highlights relevant means and measures to realize sublimation but also specifies and operationalizes the structure and aim of sublimation. Appropriating the "other within," that is, our symbolic and material order, as well as ever-wanting drives and insatiable desire require that we do not completely coincide with our present self. Although the "insatiability" and "instability" that drives bring about enable soaking off rigid constraints that the existing symbolic and material order convey on us, it is by virtue of particular goals and ideals that we can actively accept and integrate both this order and our drives. Goals and ideals enable the "other within" being incorporated in a more encompassing project that we aspire to and in a more coherent narrative that reflects who we want to be. Hence, sublimation entails, first of all, actively accepting and integrating (appropriating) our symbolic and material order as well as our ever-wanting drives and insatiable desire (the "other within") by virtue of making visible the goals and ideals embedded in or promoted by that order. Acknowledging that technologies are increasingly part of the order that shapes the self, it is by virtue of goals and ideals that we can actively appropriate them. With regard to technologies, sublimation entails, therefore, appropriating "intrusive technologies" through making visible the goals and ideals embedded in or promoted by them.

Second, crucial is not whether the goals and ideals by virtue of which we appropriate our symbolic and material-technological order and drives determine or do not determine our conduct but whether we deliberately and critically identify with them. Real critical identification with goals and ideals, which could lead to their adjustment or even rejection, is only possible in the light of other or longer-term goals and ideals; these could also be subjected to criticism but only in the light of yet other or higher goals and ideals and so on. This requirement also applies to technologies: it must be always possible to subject "intrusive technologies" and congealed technological infrastructures to fierce criticism through making visible and enable challenging the goals and ideals embedded in or promoted by them by virtue of other or longer-term goals and ideals. The filter bubble is a good example of how deliberate and critical self-identification is rendered impossible, since, in the words of Eli

Pariser, "personalized filters usually have no Zoom Out function, it's easy to lose your bearings, to believe the world is a narrow island when in fact it's an immense, varied continent" (Pariser 2011, 61).

Third, deliberately and critically appropriating the "other within" in general and technologies in particular, which is a mark of sublimation, is a challenge that cannot be only realized on an individual level but has to be fostered by and embedded in society as such. The possibility of individuals deliberately and critically identifying themselves with certain goals and ideals needs to be secured by an open society that has incorporated certain mechanisms and regulations, such as informed consent, the possibility to opt-out and the acceptance of minority views and life forms. Since the organization and design of societies is evermore delegated to technologies, it is crucial to secure possibilities to deliberately and critically engage with the technologies and technological infrastructure that is evermore shaping not only our interactions with our surroundings, but also our sense of self. Critically engaging with that infrastructure requires political and societal measures that make visible the goals and ideals embedded, endorsed or promoted by it, and enable challenging, adjusting or even dismissing those goals and ideals on the basis of other, higher or longer-term goals and ideals.

Fourth, good self-formation as sublimation entails the continuous incorporation of and devotion to evermore inclusive ideals. The ultimate ideal as the most inclusive ideal refers to a phase that could be reached after an endless process of "self-criticisms and of heterocriticisms." We see here how criticism, society and evermore inclusive ideals go together: that an ultimate ideal can be "consistently pursued in any and all circumstances" (Potter 1997, 49) means that it does not encounter any resistance within an endless community of critical minds. Being a regulative ideal and always projected in the future, the ultimate ideal entails "nothing more" than hope. However, this hope is a necessary condition for the persistence of critical self-formation, since it prevents absolutist claims bringing further criticism to a halt. The more we devote ourselves to the pursuit of inclusive goals and ideals, the more we will be able to form and cultivate a disposition that is responsive to an infinite (internal and external) community of critical minds, which in its turn will again increase sensitivity to the attractivity of even more inclusive ideals. The ultimate ideal secures that there is always, one could say in Lacanian terminology, a "permanent void" that prevents self-formation reaching a final state, which structurally and permanently enables the possibility to overcome the status quo.

With regard to technologies, this aspect of sublimation calls for devotion to an ultimate ideal that fosters the ongoing cultivation of a critical disposition toward technologies and technological environments. This requires, at the very least, making visible which values, goals and ideals are embedded in or promoted by the technologies and technological

environments in question, which could be done in public (physical and virtual) spaces, debates, discourses and performances. By making these goals and ideals visible in public, an (infinite) community of critical minds is invited to evaluate, challenge and adjust them.

The *why* of self-formation as sublimation attempts to characterize *why* we should appropriate the "other within" by virtue of a critical identification with ever-inclusive goals and ideals, secured by sociopolitical measures. Although the *why* of sublimation is motivated by realizing a kind of ideal, it intends, at the same time, to recognize the limitations and challenges that modern reality poses as well as the possibilities that it offers. It wants to recognize that self-formation lacks univocal standards and cannot be completely controlled without giving up the possibility to impose on ourselves a form that, although necessarily reflecting the imprints of our environment, still can be considered "ours." Moreover, the absence of univocal and definitive standards for establishing what is suitable or good self-formation implies that the ultimate ideal cannot be ascribed a definite particular content that is derived from a moral theory, since every moral theory presupposes standards that could also be subjected to criticism. This "absence" is a necessary condition for developing a singular form, which also explains why good self-formation as sublimation ultimately refers to an *esthetic* ideal that is always projected in the future, and, hence, is regulative, tentative and virtual. Its most important working is to secure a "permanent void" that enables individuals to critically engage with and appropriate the "other within" (their symbolic and material order as well as their drives) in order to form a singular self.

With regard to technologies, this means that self-formation as sublimation entails appropriating intrusive, extimate technologies (technologies that are increasingly shaping the very fabric of the self) in order to give ourselves a stable and socially valued as well as a singular form. Securing the "void" in a technological context entails the possibility of making visible the values, goals and ideals that are embedded in or promoted by technologies and subjecting them to criticism from an infinite community. Not only is it important to cultivate a critical disposition toward technology but, conversely, the aim is also to design technologies that foster a critical disposition. These technologies we might call "sublime technologies." Sublime technologies and technological environments not only prompt certain actions and decisions, but also foster the cultivation of a critical disposition that enables *feeling* toward which ideals we have to orient ourselves and which particular habits we have to develop in order to actively and critically appropriate the "other within" in such a way that we give ourselves a socially valued and singular form, a form that does justice to being both a dependable social being and a being that never can be captured in general characteristics. Art and artist might be able to provide examples and methods that demonstrate or illustrate

Table 9.1 Technological Sublimation Theory

		Self-formation as sublimation	With regard to technology
Caveats	1	The self is compromised, since it contains an "other within" and, hence, cannot completely control its course of self-formation	Technology as "other within" has become both the necessary condition for forming a stable social self *and* an obstacle that prevents it from reaching the singularity that it seeks
	2	There are no univocal standards for "good self-formation"	Technologies bring about different and new standards for determining what is "good self-formation"
What		Sublimation entails actively accepting and integrating (appropriating) our symbolic and material order as well as our ever-wanting drives and insatiable desire (the "other within")	Sublimation entails actively accepting and integrating (appropriating) technologies that are increasingly shaping the very fabric of the self ("intrusive technologies")
How	1	Sublimation entails appropriating the "other within" by virtue of making visible the goals and ideals embedded in or promoted by it	Sublimation entails appropriating "intrusive technologies" through making visible the goals and ideals embedded in or promoted by them
	2	Sublimation entails appropriating the "other within" through deliberate and critical identification with and, if needed, adjustment or rejection of relevant goals and ideals by virtue of other or longer-term goals and ideals	Sublimation entails appropriating "intrusive technologies" through making visible and enabling challenging the goals and ideals embedded in or promoted by technologies and technological infrastructures by virtue of other or longer-term goals and ideals
	3	Sublimation entails sociopolitical measures that make visible dominant goals and ideals and enable critical self-identification and self-formation	Sublimation entails sociopolitical measures that make visible and enable challenging dominant goals and ideals embedded in or promoted by technologies and technological infrastructures
	4	Sublimation entails devotion to an ultimate ideal that prevents absolutist claims bringing further criticism to a halt and self-formation reaching a final state	Sublimation entails devotion to an ultimate ideal that fosters the ongoing cultivation of a critical disposition toward technologies and technological environments
Why		Sublimation entails appropriating the "other within" by virtue of an ultimate, esthetic ideal that secures a permanent "void" in order to give ourselves both a stable and socially valued and a singular form	Sublimation entails appropriating "intrusive technologies" by virtue of an ultimate, esthetic ideal that secures a permanent "void" in order to give ourselves both a stable and socially valued and a singular form

how a "void" in technologies and technological environments could be incorporated and how sublime technologies could be developed.

This sketch of how self-formation as sublimation in a technological environment should be grasped is summarized in the Table 9.1. It must be stressed that this characterization is programmatic and solicits further elaboration and illustration.

To sum up what self-formation as sublimation in a technological world entails:

> Acknowledging the impossibility of complete control and the lack of univocal standards, self-formation as sublimation entails appropriating intrusive technologies, which are increasingly shaping the very fabric of the self, by virtue of an ongoing deliberate and critical identification with goals and ideals embedded in or promoted by those technologies, through sociopolitical measures and devotion to ever higher and more inclusive ideals that culminate in an ultimate, esthetic ideal, that enables forming both a stable and socially valued and a singular form.

The culminated self-formation as sublimation view in the context of appropriating technologies can be named "Technological Sublimation Theory" and abbreviated as TST. It must be clear that this theory is normative. It is normative not because it entails making visible norms, values and goals embedded in technologies, but rather because it provides normative guidelines for appropriating technologies and forming a "good" self. In the next and last chapter, I will attempt to further operationalize TST by applying it to three technological fields.

Note

1 Here also insights in the social influence of technology acceptance in social psychology research might prove itself useful, especially the effect of self-identity on technology acceptance decision in relation to the Technology Acceptance Model (TAM) (see Lee et al. 2006).

10 Technological Sublimation Theory Applied to Three Technological Fields

Technological Sublimation Theory (TST) attempts to recognize that our technological environment greatly shapes who we are and influences who we ought to be and, at the same time, it resists the idea that the self is not more than a plaything of contingent forces that render the possibility to form ourselves in a good way an illusion. TST attempts to do justice to both the impact of the technological world on the self and the potential of the self to impose a unique, singular form to itself that can be considered suitable and admirable.

Increasingly penetrating our very selfhood, technologies are becoming evermore *existential*, that is, part of what makes us who we are. Moreover, the radical externalist, interactionist view of the impact of our environment, and particularly technologies, on the self entails that the self is "closed open" for technologies: technologies as the "other within" cannot be simply externalized but are increasingly becoming something that is and is not us, at the same time. They are increasingly becoming "extimate"; both familiar and strange, they can be experienced as uncanny. Since they increasingly *make* us who we are and, hence, cannot be completely controlled by us, the question of how we can and ought to form ourselves becomes increasingly urgent. TST offers guidelines in order to tackle this challenge.

In this chapter, I will apply TST to three different types or fields of technology: technological environments, brain imaging technologies and smart drugs. I will illustrate how this theory can help to reassess our relation to these technologies and provide guidelines for how we could appropriate them in order to impose a good or suitable form to ourselves. These fields of technology have been selected because they lucidly illustrate how technologies are increasingly becoming extimate: they no longer can be simply identified as external technologies that we can independently relate to, but are rather increasingly shaping who we are and how we experience ourselves. Applying TST to these technologies will also enable further clarifying its structure and logic, and assessing its practical value.

The set-up of this chapter is as follows: first, I will discuss technological environments that are becoming evermore "smart" and "nudge" us

in certain directions. Technological environments influence our behaviors from the outside but are evermore strongly penetrating in the inside. Second, I will discuss brain imaging technologies and the role they play in influencing and explaining psychological functions, including our capacity to form ourselves. Brain imaging technologies are situated, one could say, between the inside self and the outside world but are evermore commonly blurring that distinction. Third, I will discuss nootropics or so-called smart drugs and the way they increasingly influence what we consider good self-formation. Smart drugs affect the self from the inside but increasingly shape outside behavior. I need to stress that this illustration and application is programmatic and preliminary and is intended to set the stage for further research.

10.1 Self-formation as Sublimation in Smart Technological Environments

Smart Technological Environments

Human capacities have always depended on how humans engineer their environment to support their activities (Laland et al. 2000; Ingold 2000; Godfrey-Smith 2014). However, these processes have recently reached a new level as digital technologies are starting to drastically transform our environment. After several technological revolutions in which technologies became evermore present in our daily lives, the digital technologies that are currently being developed are actually fading away from sight (Clowes 2015). Although much of the network that constitutes the current computer era is invisible to us (wires, base stations, servers, antennas, satellites etc.), in the upcoming years the "computer as we know it" is expected to undergo a great transition: a new generation of technologies will move, and are already gradually moving, from our desktops and smartphones to the background, merging into objects and infrastructures (Weiser 1991; Floridi 2014; Bibri 2015). Information and communication technologies (ICTs) will not only be embedded in devices that we explicitly "use," but increasingly become an intrinsic part of the material environment in which we live. ICTs will, as numerous signs bear witness, be evermore seamlessly and unobtrusively integrated in our lives. Instead of one-directional commands or messages fed in our devices ("what time is it, Siri?"), we will find ourselves in technological environments that will sense, predict and adapt to our behavior. Not only voice and gestures but also cues like respiration and body temperature could function as vehicles for interaction (Aarts and de Ruyter 2009). Instead of new technologies being introduced in a natural environment (which is currently the leading paradigm), these new technologies are aimed to become the environment itself.

This vision of new technological environments is exemplified by the Ambient Intelligence (AmI) and the Internet of Things (IoT). These paradigms display smart environments equipped with sensors and communication networks that not only detect and perceive their inhabitants, but also work on them by virtue of increasingly complex sensors, system architectures and software that can gather, store, manage and analyze data with great sophistication. The full operability of these technological environments will depend on progress in different frameworks, including the accuracy of sensing technologies, the robustness of wireless communication and the advancement of the human-technology interaction approaches.

Characteristic of these technological environments is not only the connectedness of multiple devices embedded in an individual environment (shades being connected to lights in a home), but also of multiple devices embedded into multiple environments (cameras connected in a home and also connected to a fire department) as well as the connectedness of multiple technological environments (different smart homes in a city) and so on and so forth. Depending on the degree and type of connectedness, the different technological environments could be, at least potentially, considered part of one, ubiquitous and always on, technological environment. By connecting to and exchanging data with other devices and technologies, smart environments are able to increasingly act autonomously. The extent to which smart environments will shape the course of their own cognitive development will probably depend on their capacity and success to self-structure their learning (see Settles 2012 and Smart 2017).

How Smart Technological Environments Shape our Selves

Instead of being a neutral background for human existence, "active technological environments" are powerful, intrusive and, at the same time, increasingly concealed technologies that could strongly regulate our interactions, deeply influence our thoughts, habits and values and intimately shape our existence (see also Van den Eede 2011 and Kiran 2012). Studies on how the online self is constituted in cyberspace (Hongladarom 2016) and how social media constructs, maintains and remixes identities online (Papacharissi 2011; see also Poletti and Rak 2014) already demonstrated the major impact of digital environments and their characteristics on identity formation. However, now being immersed in a technological environment also increasingly is taking place in the outside world. These "active technological environments" give rise to the coming about of what Aydin et al. (2019) call "Technological Environmentality" (TE). This concept highlights that the environment in which human existence plays itself out has taken on a technological character and that this environment is actively *doing* something,

resulting in a new, technological condition for the life of *homo faber*. Imperceivable, user-sensitive, artificially intelligent electronics and software are aimed to merge seamlessly with our biological selves and, as Clark (2003, 34) indicates, "in so doing they will ultimately blur the boundary between the user and her knowledge-rich, responsive, unconsciously operating electronic environments." TE has not only great potential to further liberate us from heavy, physical labor and improve different aspects of our daily lives – ranging from logistics and security to health and intimacy – but also to strongly influence and regulate our actions and decisions as well as give rise to a new repertoire of thoughts, experiences, emotions and habits.

Aydin et al. (2019) have derived insights from Postphenomenology and Material Engagement Theory (MET) and combined them in order to more suitably and sufficiently conceptualize how active technological environments are increasingly influencing human beings. More specifically, they have reflected on how Postphenomenology and MET employ the concepts of "agency" and "intentionality." First, they have elaborated how the postphenomenological concept of "mediated agency" and the MET concept of the "dance of agency" can be combined in order to understand the emerging constellation of humans and technological environments. These two approaches highlight, the authors claim, a double dimension of agency: from a MET approach, humans and material objects interact and are engaged in a "dance" that shape an interwoven and hybrid agency. From a postphenomenological approach, the interaction between humans and material objects also results in a particular agency, though not only in relation to the objects that they interact with and that mediate their agency but also in relation to other humans, objects and environments.

In Aydin et al. (2019), the example of a smart elderly home that determines which doors will open and which will not, on the basis of radio-frequency identification (RFID) chips integrated in clothes, is given to illustrate this double agency of technological environments: first, the inhabitants with, for example, Alzheimer, and the smart house engage in a dance – and, hence, an agency – that results in specific patterns of the people moving through the house and experiencing the house. Second, the technological environment as an agent also mediates how the inhabitants relate to their world, including people that they can meet inside and outside the house and the activities that they can engage in.

The other basic concept that is utilized in order to understand how "active technological environments" influence humans is "intentionality." From a MET approach, intentionality emerges from the interplay between humans and the material environment. For example, the intention of a potter to make a particular pot is the result of the interplay between the potter's actions on the one hand, and the clay and its physical affordances on the other (Malafouris 2008a; see also Merleau-Ponty

1962, Bateson 1973, Gibson 1979 and Greeno 1994). Because the mediating role of environmental technologies does not easily fit the human-technology-world scheme that is at the heart of the postphenomenological approach, the concept of a mediated intentionality that is brought about by the environment needs to be complemented with the MET approach. Aydin et al. (2019) propose to expand the concept of "immersion" to the extent that it can indicate how technologies can merge with our world and, at the same time, can have a bidirectional intentional relation with humans. Schematically, this immersion relation is noted as "human ←→ technology/world." In contrast to background relations that have as structure "human (technology/world)" (Ihde 1990) and also depict technologies as part of the world, emersion relations display a high degree of interactivity between human beings and the environment.

The example of the smart elderly home can illustrate the added value of this configuration. The technologies immersed in the elderly home intentionally direct human beings through sensor networks that detect patterns of movement; these environmental technologies can be ascribed intentions because they intentionally can lock certain doors and open others for particular inhabitants. The responses of the elderly to these technological intentions will, in turn, influence the workings of their technological environment.

From this broadened perspective, it becomes possible to grasp how technologies immersed in the environment are not only interactive but mediating as well. They help to shape human practices and experiences, but in radically different ways than the mediations conceptualized in Postphenomenology so far. Underlying the postphenomenological approach is the insight that technologies are not elements of the world, but of the relation between humans and world. By complementing postphenomenology with MET, it is possible to see that in "active technological environments," technologies do indeed become part of the world and still play a role as mediators.

Aydin et al. (2019) propose that environmental technologies are not means used to establish a connection between user and environment, but rather a "milieu" that connects humans and world by interactively "encompassing" the ways in which humans are related to the world. They provide as analogy water being the milieu for fish: the water connects fish in specific ways to other fish and plants and objects and, more fundamentally, constitutes them as fish and the environment as their world. In their words:

> This, then, is what we would like to call Technological Environmentality. Technology is becoming a mediating milieu, merging with the world to the point of becoming invisible, but at the same time intentionally directed at humans and helping to shape how humans

act, perceive, and live their lives. This radically new environmental character of human existence marks a new stage in the history of homo faber and, therefore, of the human condition itself (Aydin et al. 2019, 337)

The idea of TE becoming a mediating milieu is in line with the notion of extimate technologies developed in this study: the technologies embedded in the environment cannot be simply identified as external technologies that we can independently relate to, but are increasingly part of the fabric of our selfhood and evermore determining the way in which we form ourselves.

Sublimation in Smart Technological Environments

How can we critically relate to "active technological environments" that, due to their relative autonomy and active processing nature, are becoming more and more opaque in their workings and evermore effective in influencing and directing our conduct? How can we deliberately and critically form ourselves in smart technological environments that are increasingly becoming "extimate"? I do not pretend that TST provides conclusive answers to this question. It can rather map out the most important conditions and provide some guidelines for forming a "good self" in smart technological environments.

In line with the first caveat that has been formulated, TST recognizes that smart environmental technologies become a necessary condition for forming a stable and socially dependable self and, at the same time, an obstacle that prevents reaching its desired singularity. A smart elderly home can enable an inhabitant to stay in her familiar surroundings, adequately interact with other people and institutes, fulfill certain social and individual responsibilities and remain relatively independent of caregivers who decide how she should spend and live her day; at the same time, it can intrude in different degrees in her life, selecting the TV programs that she could watch, choosing what food she can eat, evaluating how much medication she needs to take, regulating her sleeping pattern and determining where she can and cannot go.

Now the example of the smart elderly home can also illustrate more specifically what an "extimate technology" entails. The information collected by the different devices in a smart elderly home needs to be related, interpreted and evaluated. A so-called central decision-making platform can gather data from the sensors and actuators, process and analyze measured data and send feedback to the user or to the actuators. It may also store measured data and run prediction algorithms that can exploit features of artificial intelligence and make use of deep learning and machine learning techniques to learn and develop models for the home environment as well as for the behavioral and physiological patterns of

the occupants (Majumder et al. 2017). Although smart houses could this way enable the elderly to remain in their familiar home environment, improve safety, reduce social isolation, increase self-esteem and enjoy their normal lives with friends and family, at the same time, the on- and off-body sensors and actuators that monitor and detect anomalies, perform health analysis, influence and regulate conduct will establish an intimate alignment with the inhabitants and instigate a particular life style. Inhabitants will design and experience their life in accordance with the possibilities and impossibilities offered by a particular technological infrastructure, an infrastructure that can intrude to such a degree in their views and decision-making processes that they eventually will not be able to simply externalize it but will experience it as part of their very self, as an "other within."

This brings me to the second caveat that has been identified: TST recognizes that smart technological environments could bring about different and new standards for determining what is "good self-formation." How does that work out for smart elderly homes? From a more traditional perspective one could, for example, say that, comparable to the influence of a nurse or care provider, smart environmental technologies diminish the inhabitant's privacy and autonomy in order to secure her safety. Both are then considered as neutral means that in themselves have no normative implications. However, one needs also to recognize effects that are different from conventional nursing or care homes: whereas in a nursing home the infringement on privacy and decrease of autonomy often goes hand in hand, in a smart elderly home this does not have to be the case, at least not in the same way. It is possible that as smart environmental technologies become more intrusive and the infringement on privacy increases, sensors, actuators and even cameras are accepted if the loss in privacy is traded for autonomy. Moreover, compared to a care provider, who depending on temperament and mood may verbally or physically obstruct or direct an inhabitant's action or decision in a certain way, the intervention of an often "invisible," consistent and stable technological environment might be experienced as less of a privacy infringement. In addition, the actions and decisions delegated to a technological environment rather than to another human being might also be experienced as an increase of autonomy for the elderly, if sufficiently incorporated. Instead of another person doing certain things for her, an elderly person in a smart home might experience that she is doing those things herself.

One might notice here that this is not a very apt evaluation of the differences between conventional and smart homes from the mentioned caveat: instead of expressing them in terms of more or less autonomy or privacy, this caveat inherent in TST rather highlights that new smart elderly homes might not only quantitatively change, but also qualitatively reframe values like autonomy and privacy, attributing another or new

meaning to them. A smart elderly home is not a neutral environment but will gradually inscribe in the thoughts, actions and habits of its inhabitants its "script" of what "privacy" and "autonomy" are, of what a "humane" way of living is. "Privacy" and "autonomy," after living in a smart house, might no longer be the same thing as they were before living there. Not being able to choose one's own food might earlier have been experienced as a serious limitation of autonomy, whereas after having embodied these technologies it might be experienced as a smart and deliberately chosen healthy diet. The more strongly the elderly in the smart home embody the environmental technologies that regulate their behavior, the more the technologies will become transparent and experienced as intrinsic part of their constitution, which, as said, explains why certain technological interventions will not be felt as usual privacy infringement or autonomy decrease. Nevertheless, the technological environments, in taking part in the decision-making process and regulating and changing it, will introduce different and possibly even radically new thoughts, habits and experiences. The smart technologies will intrude in the inhabitant's "inner realm" and transform what she finds useful, meaningful and admirable. They will not only replace the means that an inhabitant employs to realize certain goals, values and ideals, but also might and often will affect the meanings ascribed to and that "feel" associated with them as well as introduce new goals, values and ideals that are worth pursuing and cherishing.

Self-formation as sublimation entails, as I elaborated, appropriating, that is, deliberately and critically accepting "intrusive technologies," technologies that are increasingly shaping the self and determining its decisions and actions. The aim is to integrate these "intrusive technologies" in such a way that a person can form herself in a socially acceptable as well as a singular way. In order to disclose the *how* of sublimation, the different aspects that indicate how this should be done have to be highlighted. The first step is making visible the goals and ideals embedded in or promoted by the technologies and technological infrastructures at hand. With regard to smart elderly homes, various goals and ideals are mentioned in different stake holder analyses (Sponselee et al. 2008; Clark and McGee-Lennon 2011; Ehrenhard et al. 2014): economic goals and ideals with regard to reducing elderly healthcare and well-being costs; business goals and ideals that suppliers of AmI and IoT technologies want to achieve; goals and ideals of the inhabitants and their families with regard to safety and well-being; and, importantly, the idiosyncratic preferences of elders that a palette of options provided by standardized technologies does not automatically cover. TST especially encourages foregrounding and revealing how in smart homes important goals and ideals are operationalized from the embedded technologies perspective, which are also attributed a particular type of agency; for example, "safety" could be framed and operationalized by the technologies

embedded in the smart house in terms of detecting smoke and gas leakage, or unusual movement and falls.

Second, appropriating "intrusive technologies" requires not only making visible but also enabling challenging goals and ideals embedded in or promoted by technologies and technological infrastructures by virtue of other or longer-term goals and ideals. The question here is whether the mapped-out goals and ideals sufficiently coincide with the various goals and ideals that can be identified from different perspectives after careful deliberation. Sensors and cameras might be able to detect smoke and gas leakage as well as unusual activity and falls, but the inhabitant of the smart house might associate safety more with her familiar neighborhood or the nice couple that lives at the other end of the hall and helps her with her groceries. Or she might consider activity detection a decrease of safety because she worries that burglars could hack into the system and use it to find out when she is sleeping. Think also of the possibility of an elderly lady having a lover but being afraid that the sensors will detect erotic interactions or interpret them as unusual behavior, which might be a reason for not inviting her boyfriend to her home. The biggest challenge is to be as inclusive as possible in identifying goals and ideals that are of crucial importance for the well-being of the inhabitant. Often, this boils down to including, besides goals and ideals that concern functional and technical aspects, goals and ideals that are especially specific for the inhabitant in question. However, now the biggest challenge is realizing communication between all the sensors and actuators as well as integration of different technologies, systems and services from different manufacturers through establishing standard protocols and enabling interoperability. This might be critical for the consistent operation of the system but is often only a necessary and not sufficient condition for a "good" smart home, just like having electricity and heat is not the end but the beginning of designing an appropriate home; a house is not yet a home.

The third aspect of self-formation as sublimation that has been highlighted entails sociopolitical measures that make visible and enable challenging the goals and ideals embedded in or promoted by technologies and technological infrastructures. The smart elderly home example immediately makes clear that realizing a "good" smart house cannot be accomplished by efforts of individuals but only through collaboration of all involved stakeholders: engineers, suppliers, programmers, ICT support, but also family of the elderly, insurance companies, law makers, politicians and of course the elderly person herself. This aspect underlines the importance of creating buffers in society that make room for debate and criticism on different levels. Governments and insurance companies need to safeguard the rights of elderly people to choose between different smart systems. Lawmakers need to take measures that protect the data that these systems collect and store. Public discussions

need to reveal and map out the different interests in order to make sure that all the ingredients required to design smart homes (and not just houses) that go beyond resolving operability issues are there. The main aim here is to discuss and develop frameworks that articulate which goals and ideals should be used to evaluate what is a "good" elderly home on the level of a particular society and culture.

The fourth and "final" step in the process of self-formation as sublimation is to ensure that debate about and reflection on how technological environments are shaping and determining the thoughts, habits and experiences of the people involved never reaches an end. This permanent room or void can only be maintained if goals and ideals that are established in social interaction and conversation always remain provisional. By projecting an ultimate, virtual and maximally inclusive ideal in the future, it becomes possible, at least potentially, to challenge every particular goal and ideal, as beautiful as it seems. Applied to technological environments, the ultimate, esthetic ideal prevents smart systems from becoming closed black boxes that cannot be adjusted by users. The ultimate, esthetic ideal ensures taking into account that inhabitants of smart elderly homes have different and divergent worldviews and that what is good for others does not have to be good for them. It creates a "void" that enables forming a singular life style, without losing sight of the different relevant practical, legal and social conditions that need to be met in the design of the house. Elderly individuals have their own particular histories, views, habits and feelings, which cannot be completely generalized and captured in a standardized technological environment. Taking this into account and building rich modularity in the design of a smart elderly home allows elderly people to deliberately and critically shape their selves in interaction with different stakeholders as well as the devices that promise to make their lives better.

This example also makes clear how from TST, if certain conditions are met, there does not have to be a discrepancy between the preferences and desires of the inhabitants of smart homes and the influences that other stakeholders impose on the design of the house. Enabling idiosyncratic preferences of the inhabitants are, as such, not a sufficient condition for singular self-formation. While they could make room for new configurations, they can also limit the range of possibilities and block actualizing potential, potential that remains out of an inhabitant's view. Put differently, singularity is not the sheer result of fulfilling individual preferences and desires by decreasing the influence of "evil" others (e.g., producers and suppliers of the devices) in the design of the house, although giving producers and suppliers too much and unwarranted power should be prevented. TST rather promotes collectively finding out what would be a "good home" for the particular inhabitant in question through mapping out which goals the smart house seeks to realize, what are the interests, aims and limitations of relevant stakeholders, including

those of the inhabitant, in the light of an ultimate ideal, that is, an ideal that secures a void that solicits unforeseen sublime designs.

10.2 Self-formation as Sublimation through Brain Imaging Technologies

Brain Imaging Technologies

In the 1990s, brain research witnessed such a gigantic increase in interest and popularity that in July 1990, with the Library of Congress and the National Institute of Mental Health, former US President George H. W. Bush declared that period "the decade of the brain." Expectations were materialized in neuroscientific funding, research, publications, public debates and emerging applications. In the following years, advances in brain research, in combination with the dominant scientific (and philosophical) consensus that mind in one way or another emerges from the brain, have led to the ever-stronger view that brain research has the enormous potential not only to clarify and solve biomedical issues, but also to tackle behavioral and social problems. That explains why various scholars and policy makers have coined the current era as the "decade of the mind" (Kutas and Federmeier 1998; Moriarty 2008, 29).

Brain imaging technology has commanded attention that goes beyond its usage in the laboratory and the medical field; it has increasingly featured as a "holy grail" in, among other fields, public policy making (Snead 2007; Littlefield 2009), education (Mobbs et al. 2007), law (Sinnott-Armstrong et al. 2008), popular media (Frank 2009) and private enterprises (e.g., NoLie MRI, Cephos). It also plays an increasingly prominent role in social sciences: brain imaging technologies such as positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) are, as I have discussed earlier, increasingly used not only to diagnose diseases and lesions but also to correlate brain activation with psychological states and traits. Many philosophers have become interested in brain imaging technologies because the images that they display seem to have bearing on psychological and philosophical issues such as consciousness, free will, weakness of will, responsibility, moral deliberation, rational choice, agency and identity. Data and images provided by brain scanners are used to draw (causal) relations between brain processes and cognitive functions, with the aim to potentially resolve problems concerning thinking and behavior. Although technology to visualize brain activity is fairly new and at a very incipient level of development, it is precisely the visual quality of these productions that make them particularly appealing, stirring both scientific and public imagination (Beaulieu 2002; Dumit 2004; Eastman and Campbell 2006; Weisberg et al. 2008). The idea that brain imaging allows "reading" what is going on in the brain and provides "brain facts" that

enable uncovering the secrets of the human body, mind and behavior is a prospect that causes tempers to flare (Choudhury et al. 2009).

Of all brain imaging technologies, fMRI stands out: being noninvasive and depicting different parts of the brain in moving colors and shapes, it has an appeal and perceived accessibility that makes it particularly popular. The neuroscientist Nikos Logothetis comments: "Its popular fascination is reflected in countless articles in the press speculating on potential applications, and seeming to indicate that with fMRI we can read minds" (Logothetis 2008, 869). For this reason and in order to apply TST on one particular brain imaging technology, I will focus my further discussion on fMRI.

In Chapter 5, I have discussed brain imaging technologies, and in particular fMRI, in the context of the attempt to reframe free will in terms of self-formation. In that context, different characteristics of fMRI have been briefly depicted. We have seen that the relation between fMRI images and brain activity is not immediate and clear-cut. In fact, the technology measures, as we have seen, the so-called Blood Oxygenation Level Dependent (BOLD) response, because oxygenated and deoxygenated hemoglobin in the blood have different magnetic properties, an MRI scanner can read changes in magnetization and translate them into a computerized image (Matthews and Jezzard 2004). The data and images provided by brain scanners are used to infer something about the role of particular brain areas in cognitive functions. In addition, brain imaging data are taken as evidence that a given brain region plays a particular causal role during the performance of a cognitive task (so-called "reverse inference").

I have discussed different methodological objections against these inferences: first of all, it is clear that brain images are not simple pictures of BOLD signal differences. In addition, the BOLD differences that are associated to brain activity are small, noisy and multifaceted. Instead of measuring quantitative signal magnitudes, brain images rather display *statistical* significant differences in BOLD signal. Brain images are in fact maps of places where we can assume with a certain probability that the presumed correlation between data and a stereotyped pattern of activation is unlikely to be the result of chance fluctuations from a true zero signal (Klein 2011). Moreover, interpreting brain imaging is belief-opaque: inferences are dependent upon the beliefs and experience of the researchers about how to analyze their results. We have also seen that reverse inference is no less controversial: a causal relation between activation of a certain brain area and a particular cognitive function cannot be drawn because brain regions can perform multiple functions and be involved in a variety of different processes (Klein 2010).

In an attempt to more radically reflect on the usage of fMRI images as functional evidence, I have argued that the discussed methodological objections show that brain images do not directly display neural activity,

but nevertheless, neither challenge nor sufficiently explicate the particular way brain imaging technologies frame the relation between brain, cognition and world. These methodological objections indicate that brain images might not be reliable and valid pictures of brain processes but, at the same time, they reinforce the view of the brain as an *agent* by framing the brain as an isolated realm that – if certain methodological problems are solved – could provide functional hypotheses.

How Brain Imaging Technologies Shape our Selves

Brain imaging technologies in general and fMRI in particular are increasingly shaping our image of ourselves. In Chapter 5, I illustrated how brain imaging technologies are rehabilitating the Cartesian framework, which was itself influenced by the camera obscura: fMRI mediates and transfers similar Cartesian notions of self and freedom. By focusing on the inside of the head and (interactions between) certain regions of brain activity, an fMRI image depicts the brain as a relatively isolated realm and allows approaching it as an independent inner agent. Instead of the mind, fMRI attributes a privileged status to the brain: it ascribes to the brain the capacity to process information collected by the senses, which enables the execution of certain decisions and actions as well as control and manipulation of the world. fMRI portrays the brain as the locus of the self and attributes to it a particular concept of freedom, namely freedom as "not being determined by something else but capable of determining behavior and utilizing the world," which is taken as a synonym of "control." One could say that fMRI increasingly shapes the self in its image.

Although this idea of freedom was also attributed to mind and brain before the discovery and usage of fMRI, fMRI (and other brain imaging technologies) has made it appear more credible because it allegedly is capable of clearly displaying the interior of the "inner dark room" and disclosing not only what really determines our mind and behavior but also how it is able to do it. The prospect that fMRI can provide objective knowledge of how the brain causes decisions and actions does not weaken the camera obscura framework of freedom, but rather reinforces it and makes it "scientific." This way, propositions, like that fMRI can grant "a glimpse into someone's mind" and enable us to "measure what someone sees with his mind's eye" (2010, 264; see also Lamme 2006), suddenly do not sound completely esoteric but can be expressed in scientific journals. fMRI potentially becomes a mirror of the soul, which could disclose all the secrets of human thought and behavior. Somebody who can adequately read that mirror, a "brain technician" (Armstrong 1968/1993, 109), would be able to know someone's thoughts and behavior even better than the person herself and explain why she acts in a certain way and is not able to act differently. Proposed applications

range from offering better understanding of human learning, shedding light on schizophrenia, providing new insights into addiction and violent behavior, tamper-proofing lie detection and screening employees for pedophilia. And if the brain is the central operating system of both the "hardware" and the "software" of humans and fMRI is the portal that provides access to their thoughts and can explain their behavior, then making adjustments in and through that operating system will also enable controlling and determining their conduct.

The field of law is one of the most prolific and growing spheres of emerging fMRI applications, primarily taking place in the US. This is, according to Wolf (2008, 21), not surprising because "law has everything to do with human intentions, states of mind, competence, culpability, and responsibility, the points of possible connection between law and neuroscience are legion." Criminal law in particular is integrally concerned with understanding the mind of the defendant. In order to prosecute, lawyers are required to "prove beyond a reasonable doubt" that a suspect had a particular mental state indicating an intent to kill before committing the murder (this is a defining feature for first degree murder), the intent to deceive (a defining feature of fraud charges) or the capacity to control his/her actions (a defining feature of the insanity defense). Jones and Shen clarify:

> Culpability of the accused thus depends, in part, on a determination of his/her mental state at the time of the offense. The phrase 'mens rea' ('guilty mind') derives from the Latin phrase 'Actus non facit reum nisi rea sit,' which means 'An act is not guilty unless the mind is guilty' (Jones and Shen 2012, 361)

All issues in criminal law depend, to some extent, on inquiring into the content of "the mind" of the defendant, which makes fMRI images extremely appealing.

It must be clear that, in contrast to smart technological environments, fMRI images are not shaping the self in the sense of immediately directing its behavior or manipulating its environment. However, the assumption that fMRI enables reading the mind not only expresses a representationalist view, but also mediates a particular conception with regard to the role that the brain plays in relations between self, behavior and world. By portraying the brain as the locus of the self, fMRI images are depicted as representations of the "real" determinants of a person's behavior, rendering the (conscious) self a powerless spectator who is very much dependent on the brain's condition and working order. If, in the context of law, fMRI images show that deviant behavior is the result of the brain being "impaired," "not functioning well" or "out of control," a perpetrator could be cleared from responsibility and liability.

This line of reasoning raises a disturbing question in this respect (that I will not further discuss here): why does only an "impaired brain" and not also a "healthy brain" relieve a person from responsibility and liability? Are they not both beyond the conscious self's control? We were not able to choose our brains, so why are we made responsible for their doings, bad or good? I am not saying that this conclusion must be necessarily drawn but only that strongly endorsing the brain as the central operating system of the self and its behavior raises this kind of questions.

In addition, fMRI images not only provide information to specialists that have to interpret and evaluate a person's decisions and actions, but they also impel the examined person to identify herself with the images of her brain, and form herself in light of what those images display. Perpetrators often adopt the language that clinicians use to understand and explain their behavior, as illustrated by Strakowsi: "the fMRI shows that I am not a bad person but have a disruption of the ventrolateral prefrontal-amygdala emotional pathway" (Strakowski et al. 2011). I am not being cynical here; I acknowledge that clinical diagnosis can be helpful in controlling and treating deviant behavior and recognize that clinicians are often very much aware of the limitations of brain imaging technologies. The point is that fMRI is not a neutral instrument but that it mediates a very particular conception of the self and its relation to its behavior and world. It informs and misinforms how the self is and can be shaped – and can be treated.

From the perspective developed in this study, brain images are not seen as representations of a "within"– the brain – that is not affected by a "without." The idea that brain images can provide direct access to the "inside self" can only be validated insofar as the technical conditions that have generated it remain invisible. This does not imply (as I elaborated earlier) that the information that fMRI images provide is completely irrelevant. fMRI images can display certain correlations between brain activity and behavior and, more importantly, enable a person to use that information to intervene in and adjust her behavior. From this perspective, which is crucial, the self is not understood as an isolated, *a priori* existing entity that is "immediately" determined by something "within"– consciousness or brain – but as something that deliberately discovers and forms itself by virtue of its identifications and non-identifications with its determinants on the basis of certain goals and ideals. A person can critically relate to the information provided by fMRI images and ask herself whether she wants to identify herself with her current habits. Although it is possible to recognize certain brain patterns or mechanisms as playing an important role in a person's decisions and actions, they are not understood as representants of the "real" determinants of a person's conduct. Instead of representing the "real" determinants (and rendering the self a powerless spectator), fMRI images allow for the practical-perceptual visualization of one's

neurophysiology. By foregrounding neurophysiological processes and revealing how our habits relate to our neurophysiological make-up, they can offer insight in how we more effectively can pursue our long-term goals and ideals.

It is clear that from this perspective, it is not suitable to demonstrate on the basis of fMRI who someone "really" is. fMRI images will not enable discovering whether someone is a "real killer," not only because fMRI images do not provide valid and reliable knowledge of a particular mental state or a "guilty mind" but also, and more importantly, because brain or mental states are not taken as loci of the self. From the proposed perspective, the self is not completely determined by something within or without but rather continuously discovers and forms itself by virtue of its identifications and non-identifications with its determinants on the basis of certain goals and ideals. fMRI images could contribute to discovering certain inclinations and offer insights on how it more effectively can become the self it aspires toward.

Sublimation through Brain Imaging Technologies

In Chapter 6, I briefly discussed "neurofeedback" as an illustration of how brain imaging technologies could be reinterpreted from the proposed perspective. Now that we have further developed that perspective in terms of self-formation as sublimation, it is possible to more structurally and comprehensibly elaborate how it could be applied to this field. I will take again "neurofeedback" as an example in order to illustrate how the proposed and now more advanced theory (TST) could be applied and operationalized. As we have seen, neurofeedback entails offering people dealing with certain afflictions, via real-time displays of electroencephalogram (EEG) or fMRI, feedback on the way they respond to certain situations. This self-reflection enables actively influencing through fMRI brain activity in order to bridge possible tensions between how somebody acts and how she would like to act. In order to further specify the usefulness of TST, I will again focus on the treatment of *food craving* through neurofeedback, a phenomenon closely associated with developing obesity. As a disclaimer, I highlight again that the clinical merit of neurofeedback on food craving is still in a trial phase (Weingarten and Elston 1990; Bowell and Kober 2016; Volkow et al. 2011).

In line with the first caveat that has been formulated, TST recognizes that brain imaging technologies in general and fMRI in particular disclose the brain and its processes and mechanisms as an essential determinant for forming a preferred stable self and, at the same time, as an obstacle that prevents reaching the self's desired singularity. On the one hand, it acknowledges that fMRI displays the brain as an active agent by virtue of which we can form ourselves in a certain direction. fMRI can make visible how the brain plays an important role in regulating and

determining our thinking, decision-making and behavior. By providing extrospective knowledge of neural pathways that give rise to, for example "food craving," fMRI makes it possible to intervene in and adjust the self's undesirable behavior, shaping it into the stable self that it aspires toward. On the other hand, fMRI displays the brain "as another within" that cannot be simply controlled by the self. The neurological proclivities that it discloses, which explain, for example, the intense desire to consume a particular food or food type, cannot be simply externalized and detached from the self. Those desires and tendencies are experienced as part of the self, explaining why the self can say "*I* am craving for a hamburger." And it is precisely because the self identifies itself with "craving for a hamburger" that it cannot easily control that tendency and form itself in another direction. Moreover, if neurofeedback would enable the self to use "extrospective knowledge" to intervene in the brain's proclivities and successfully prevent eating a hamburger, one could again say that it is the "plastic" brain that provides room and opportunity to intervene in certain tendencies and adjust the self's conduct. fMRI discloses the brain both as an opportunity to form a stable and socially dependable self and an obstacle for deliberately forming a singular self.

The second caveat TST highlights is how fMRI could bring about different and new standards for determining what is "good self-formation." One might say that the displayed fMRI images with regard to our example of food craving are not very different from a friend holding a (symbolic) mirror to confront a person with her behavior: "Stop eating hamburgers! Driving to the McDonald's during lunch break has become a dangerous habit!" However, on reflection, it becomes clear that fMRI images are far from neutral and interchangeable. Whereas a friend's analysis and advice could be reduced to an opinion ("Shut up, that is like your opinion dude!"), fMRI images have a more scientific status. They seem to immediately display what is going on with somebody who is suffering from food craving and why it is so difficult for her to resist the desire to drive to McDonald's. And because fMRI images could be seen as more objective than a friend's opinion, they might also be experienced as more compelling and persuasive. Different from a friend's advice, fMRI images used in neurofeedback seem to leave a person's autonomy intact, since that feedback does not come from another particular person with certain motives and intentions but from an impersonal, scientific device that seems to neutrally mirror what is going on in the autonomous person's biology.

Again, it would be insufficient to evaluate the difference between a friend's opinion and the fMRI image in terms of more or less scientific and objective or more or less securing autonomy. If fMRI images are presented and discussed in a clinical setting, using clinical terminology to describe what they indicate ("Ma'am, you have an eating disorder!" or "Ma'am, you have a form of hyperalimentation!"), moral valuations

are replaced by clinical valuations. Instead of a person being too weak to resist greasy food or too lazy to prepare a proper meal, she now has a brain condition preventing her from resisting hamburgers; "having a weak will" or "acting irresponsibly" are reinterpreted in terms of certain solidified brain mechanisms determining certain habits in the self's behavior. This could be used to evade responsibility: "It is not me; it is my brain that craves hamburgers!" More importantly for our theme, we see now that the way that fMRI images depict the brain and frame its relation to self and world has practical bearing. fMRI's particular framing enables ascribing to the brain a causal agency that strongly constrains or even annuls the self's ability to think or act against the brain's propensities. It replaces the freedom that was ascribed to the self with a control that should be attributed to the brain, a control that is further elaborated in terms of groups of neurons firing and wiring together in certain patterns. fMRI displays the brain as the locus of the self, dictating what and when the self craves but also how that craving could be resisted and regulated. The more fMRI images are seen as real representations of the brain as a central operating system of the self, the more the technologies and procedures that generate these images and instigate that particular framing become transparent and invisible.

These caveats are an intrinsic part of TST and do not render self-formation impossible. As we know, self-formation as sublimation entails appropriating technologies that determine the nature of the self while still allowing a person to create who she is in a way that is both socially stable and singular. Instead of taking brain images as representations of a brain that controls the self and manipulates the world, the proposed framework considers brain images as a particular and potentially useful mediator of the relations between brain, world and self. TST entails, first of all, making visible the goals and ideals embedded in or promoted by the fMRI images. We have seen that envisaged aspirations of fMRI images go beyond making visible diseases and range from tamper-proofing lie detection to providing insight in various psychological disorders. In the context of using neurofeedback to tackle food craving, the fMRI images intend to provide knowledge and methods in order to alter neural pathways correlated to food craving in favor of a healthy lifestyle. It is important to explicate here how a "healthy diet" and an "unhealthy diet" are framed and operationalized in the context of using fMRI images in neurofeedback. Food craving is correlated to particular hardened neural pathways in the brain that cause unhealthy habits; noninvasively stimulating specific regions of the brain with the help of fMRI imaging seems to significantly reduce food craving by modulating neuronal constraints, which lead to a healthy diet and a healthy lifestyle (Goldman et al. 2011; Ljubisavljevic et al. 2016).

Second, appropriating "intrusive technologies" requires not only making visible, but also enabling challenging the goals and ideals embedded

in or promoted by those technologies by virtue of other or longer-term goals and ideals. The concern here is whether the mapped-out goals that motivate overcoming food craving are similar enough to the goals that will be identified after careful reflection from other points of view. Although fMRI images mediate our self-identification, the goals and ideals that we identify with in the process of realizing a preferred self cannot be fully captured by or reduced to the goals and ideals embedded in or promoted by the fMRI images. In the context of neurofeedback, fMRI images might help to alter certain neural pathways that give rise to food craving and contribute to developing healthier habits, but the person suffering from food craving might have associations with the more healthy lifestyle as well, and maybe even more with the social aspect of hanging out with friends and occasionally giving in to a nice, greasy hamburger. Or after deliberate and critical reflection, she might discover that her wish to fight food craving is not motivated by health reasons at all, but rather by trying to match standards set by thin models in music and film. After this discovery, she still might proceed with the attempt to overcome food craving, but then with the aim of realizing an ideal that has grown up a bit, due to more thorough critical reflection. The challenge here is to integrate the fMRI images that mediate self-identification in such a way that they contribute to realizing maximally inclusive ideals that mirror not what somebody spontaneously prefers or what is fashionable or trendy in a particular subculture, but that prove admirable after thorough deliberation and critique through mapping out all involved stakeholders and their interests and motives.

The third ingredient of TST entails sociopolitical measures that make visible and enable challenging the goals and ideals embedded in or promoted by, in the case at hand, fMRI images in the context of neurofeedback. Although the example of treating food craving through neurofeedback appears to be a particularly individual endeavor (individuals set their goals in concrete practices and decide how to realize them), adequately interpreting and valuing brain images and the aspirations explicitly or implicitly embedded in them requires the involvement and collaboration of various stakeholders. First of all, clinicians are responsible for making clear to users what the advantages as well as the limitations will be in using fMRI images in order to treat food craving. They need to make clear that the information provided by fMRI images can make the discrepancy between short-term urges and long-term goals physiologically accessible, enabling them to relate and respond to it. However, they also need to make users aware of the fact that the displayed fMRI images are not pictures of mind states and that these images mediate in a particular way the brain and its relation to behavior; they need, in other words, to prevent reductionist accounts of what fMRI images display. By approaching brain imaging and neurofeedback pragmatically, it also becomes possible to include other ways to frame and tackle food craving:

instead of or in addition to neurofeedback, people who struggle with food craving could, for example, arrange with their spouses that their car or car keys are unavailable at certain times of the day, making it more difficult for them to get to the McDonald's. Extrapolating, it must be clear that raising awareness of the opportunities, limitations and relativity of the applications of brain imaging technologies is of crucial importance among the public and political actors at all levels.

Besides enabling insight in what fMRI images depict, sociopolitical measures could be taken to make visible and challenge the goals and ideals embedded in or promoted by fMRI images. These measures are aimed at creating room for considering other and more inclusive goals and ideals in the context of treating food addiction. Besides offering individuals therapies that enable them to control addictive impulses, one could also insert buffers that make visible and challenge the large-scale promotion of fast food in our society, which might, at least partly, prevent individuals from developing unhealthy addictions. It needs not much effort to show that legally forcing companies like McDonald's to disclose the ingredients in their hamburgers on little labels on their boxes is not sufficient for raising the required awareness. Another way to insert a buffer in society is to instigate debate about ads portraying the skinny role models that trigger some people to change their eating habits. This is not to discourage people from developing healthier diets, but to enable them to discover their motives for doing it and adjust their conduct using standards that they find admirable after thorough deliberation and reflection.

The fourth and "final" step in the process of self-formation as sublimation is securing a permanent void by projecting an ultimate and maximally inclusive ideal in the future, which enables challenging every particular goal and ideal set in the present. With regard to fMRI used in neurofeedback to treat food craving, projecting an ultimate, esthetic ideal enables developing a tailor-made healthy diet that fits within the particular lifestyle aspired to by a person, without losing sight of the social concerns, valuation and interests of a healthy life. The question of what is a healthy lifestyle is increasingly difficult, if not impossible, to answer in general. Of course, one could name habits that are proven unhealthy: today no one would deny that smoking, alcoholism and highly processed fatty foods are hazardous to health. By no means does the TST endorse an "anything goes" philosophy. However, one needs also to acknowledge great variations, both diachronic and synchronic, in appraisal of substances being good or bad for health. Whether treatment of food craving through neurofeedback will make a person healthy and experience herself as healthy will greatly depend on its particular integration in very particular practices and particular outlook and preferences of the person in question as well as her particular social context and circumstances. This does not exclude the possibility of many people

experiencing the neurofeedback treatment of food craving improving their health in a similar fashion. Its most important purpose is to secure room for inclusive approaches and valuations of the effects of this kind of technologies. Some people might be able to successfully integrate those technologies in their life projects and evaluate them as a contribution to living a healthy life. Others might value their effects very differently and even take an opposite stand, discovering that, for example, the neurofeedback has strengthened their obsession with a healthy diet and only increased their dissatisfaction with their bodily condition. TST recognizes the importance of a practical and socially effective appropriation of fMRI images (people functioning better), without losing sight of the fact that people could respond very differently, depending on how they value and integrate the extrospective knowledge about their neurophysiology provided by fMRI images into their life projects. Again, it must be clear that singular self-formation cannot be reduced to subjective preferences and desires with regard to living a healthy life, but rather is the mark of a collective effort to map out which goals, in this case, neuroimaging seeks to realize. It is about identifying what are the interests, aims and limitations of relevant stakeholders, including those of the subject that is undergoing treatment, in the light of an ultimate ideal that counterbalances every attempt to definitely fix what is meant by a healthy lifestyle.

10.3 Self-formation as Sublimation with Smart Drugs

Smart Drugs

The use of smart drugs, also called nootropics or cognitive enhancers, is on the rise. Smart drugs refer to natural and synthetic substances that are presumed to enhance cognitive function, particularly memory, alertness, creativity and motivation. Besides stimulants like caffeine, the best-known and probably most used smart drugs are methylphenidate (Ritalin) and Adderall (mixed amphetamine salts) and increasingly modafinil (provigil), which are medically prescribed for treating attention deficit hyperactivity disorder (ADHD) and narcolepsy, respectively. These are among a suite of pharmaceuticals now being used by healthy people, particularly university students, to enhance their capabilities for learning or working.

The effects of smart drugs are not well investigated, let alone determined, but their popularity is evident. In a large survey of tens of thousands of people (79,640 respondents in 2015 and 29,758 in 2017), 14% reported using stimulants at least once in the preceding 12 months in 2017, up from 5% in 2015. US respondents reported the highest rate of use: nearly 30% said they had used drugs for pharmacological cognitive enhancement at least once in 2017, up from 20% in 2015. However, the

largest increase in use is found in Europe: use in France rose from 3% in 2015 to 16% in 2017; from 5% to 23% in the UK; and from 10% to 25% in the Netherlands. Expectations are that this trend will intensify in the future (Frood 2018).

It is known that smart drugs modulate important neurotransmitter systems such as dopamine and noradrenaline. Laboratory studies have shown that modafinil enhances aspects of executive function in rested healthy adults, particularly inhibitory control (Turner et al. 2003). ADHD medications show a modest degree of memory enhancement (Grön et al. 2005). Other compounds with different pharmacological actions are in early clinical trials, having shown some positive effects on memory in healthy research subjects. However, it is still very uncertain whether any of these new drugs will be proven effective and safe. In addition, their workings seem to be dependent on the psychophysiology of the user, like weight, blood flow and mood. There has also not yet been any definitive research into the addictive potential of these drugs, and it is unclear what side effects are likely at doses outside the prescribed range.

Responses to smart drugs vary from very positive to very negative. Opponents consider the use of smart drugs to be unnatural, dangerous and fraudulent. The objection that smart drugs are unnatural is often based on assumptions regarding what is natural: medical treatments for the sick, exercise and education are seen as "natural" and, hence, "good." If smart drugs take us beyond "natural" limits of human functioning, their enhancements are seen as unnatural, and what is unnatural should evoke caution and skepticism (Fukuyama 2002; Kass 2002).

An objection that has been given more attention is that smart drugs are risky. Since our brain is both a very important and a highly complex organ, it is, opponents who highlight this aspect argue, very dangerous and irresponsible to use smart drugs not knowing their effects and unintended side effects in the long run. The use of methylphenidate without clinical supervision already shows, some claim, dangerous side effects such as symptoms akin to those of schizophrenia, depression and anxiety. Prolonged abuse has been associated with psychotic symptoms, including hallucinations and paranoia (Morton et al. 2000). Others have pointed out that using smart drugs comes with the risk of long-term damage to the brain's working memory and ability to move efficiently from one task to another. There may especially be a heavy cost of the use of smart drugs on the uniquely delicate, developing brain of young people, opponents claim: it might cause a long-term decrease in the brain plasticity that would be necessary for task switching, planning ahead and adaptive flexibility in behavior (Urban and Gao 2014).

Other opponents argue that healthy people who take smart drugs are not only taking a chance with the unknown long-term side effects, but also gain undue advantage over others. Smart drugs influence the

competition between those who would prefer to use them and those who would rather not, creating an incentive or even a pressure among non-users to start using too, which could lead to a situation in which all students need, or believe that they need, to use cognitive enhancers to be able to compete. If, however, they are able to resist using smart drugs, the drugs could increase factual inequalities between members of society (Sahakian and Marein-Zamir 2007; Forlini and Racine 2009; Dubljević 2013).

On the other side, advocates of human enhancement technologies like transhumanists often argue that fear for these new drugs is unnecessary and should be viewed, along with emerging technologies such as brain stimulation and prosthetic brain chips, "in the same general category as education and good health habits, and information technology – ways that our uniquely innovative species tries to improve itself" (Greely et al. 2008, 702). Although they recognize that smart drugs require relatively little effort, are invasive and are for the time being not equitably distributed, they believe that none of these provides compelling grounds for prohibition. Exercise, nutrition and sleep seem to have similar characteristics and effects as smart drugs and are, hence, morally equivalent to them, proponents of smart drugs claim, though we do not prohibit them. Considering smart drugs as cheating is, they claim, begging the question, since the rules used to consider them cheating need to distinguish currently allowed cognitive enhancement, such as double expressos and private tutors, if they are to be banned. The appeal to "natural" is rejected on the grounds that the lives of almost all human beings is "unnatural"; if we accept "unnatural" cognitive enhancers like writing and laptop computers, why should we prohibit other "unnatural" enhancers? And as for the objection on the basis of drug abuse, they do recognize drug abuse as a major social problem but they argue that regulatory rules differentiate between different types of drugs (from caffeine to heroine), which should also apply to smart drugs; the mere fact that cognitive enhancers are drugs is no reason to outlaw them. An evidence-based approach is required here, they argue. Their general take on smart drugs is that like all new technologies, they could be used well or poorly; maximizing its benefits and minimizing its harms should be our most important concern (Greely et al. 2008).

How Smart Drugs Shape our Selves

TST does not approach technologies in terms of do's and don'ts. The same applies to smart drugs, also because more traditional substances that could be considered "mood enhancers" such as chocolate, wine and cigarettes have been an accepted part of our culture for many centuries, which is not to say that these substances can be considered healthy. Just like people are changing their appearances, bodies and moods with

make-up, balding remedies, plastic surgery, steroids and caffeine, more individuals are finding their ways to more modern cognitive enhancers. If these drugs were to become broadly available and mainstream, they could increasingly become a technology that shapes the self, which implies that, as I argued repeatedly, they would no longer be simply externalized and evaluated as an external influence.

Currently, a great number of smart drugs are examined and tested. For example, donepezil, a drug approved by the US Food and Drug Administration (FDA) to slow the memory loss of Alzheimer's patients, seems to improve the memory of "healthy" people. Controlled populations of pilots who had taken donepezil remembered their training better, as evidenced by their improved performance on the maneuvers they had learned and on their response to stressful emergency situations (Yesavage et al. 2002). It is expected that many different smart drugs will make their way toward clinical trials. Of course, this does not exclude the possibility that many nootropics that looked promising in animal models will completely fail when brought to clinical trial. Another hurdle for smart drugs is that while they might enhance certain cognitive capacities, they could cause deleterious effects; for example, mice that have become more receptive for learning through cognitive enhancers seem to be also more sensitive to pain (Gazzaniga 2005, 75–79). Currently, available memory drugs are mild and marginal in their effects. Future drugs are expected to be much more powerful, which makes the question of how we should live with them more urgent.

It must be clear that the purpose of this reflection is not to review and settle debates about the usage of smart drugs in terms of allowance or prohibition. Approaching smart drugs in those terms is only possible if the question that I have addressed throughout this book was to be circumvented, namely the question of how technology is mediating and shaping the very standards that we use to analyze and evaluate its usage. As repeatedly has been elaborated, technologies that intrude in the very fabric of the self are not neutral but rather "normalize," explicitly or implicitly, what is meant by "enhancing human capacities" and "optimizing values" (see also Aydin 2015). As I have briefly discussed in Chapter 7, this also applies to smart drugs. Smart drugs, which presumably would improve cognitive capacities, are not neutral technologies, but rather express, feed into or bring about a particular conception of cognition. They do not simply measure IQ but also frame what is IQ. Generally, IQ tests measure people's analytical skills, verbal comprehension, perceptual organization, working memory and processing speed. These are skills important in different academic settings, which explains why IQ tests are reasonably good predictors of academic success. However, they are far from perfect indicators of success in the "real world." Just like the "real world" cannot be reduced to the "academic" world, intelligence cannot be reduced to IQ.

Let me explain this in more detail applying the structure that illustrates the framing and moral evaluation of technology with regard to enhancement provided in Chapter 7. I briefly recap this structure: (1) value x is identified as an intrinsic or at least as an indisputable preferred value; (2) capacity y is identified as an important capacity that affects value x, in the form: value x can be better realized by improving capacity y; technology z is identified as a technology that affects capacity y, in the form: technology z can improve capacity y. Technology z can then be qualified as a "human enhancement technology" because (1) it can improve capacity y, which in its turn (2) can significantly contribute to realizing value x, which (3) is identified as a preferred value. Applied to smart drugs, smart drugs can be considered as a human enhancement technology, because (1) smart drugs increase (*inter alia*) IQ, (2) giving more people the opportunity to increase their IQ will lead to (*inter alia*) more equality and (3) equality is identified as a preferred value. Again, the aim of this structure is not to establish whether smart drugs z can really improve capacity y, which in its turn can contribute to realizing value x. The crucial point is that this question can only be adequately understood and investigated if one has made clear how smart drugs z frames what is considered an "enhanced capacity" and an "optimized value," respectively. Characteristic of almost all smart drugs is that in framing their expected effects, "smart" describes how well one processes information and figures out tasks; smarter is frequently just another word for faster (Gazzaniga 2005, 84). So we may all get faster at figuring out new problems, but it is not clear at all what it would mean to get "smarter" beyond the given characterization, for example, with regard to making better judgments and decisions.

The structure presented in Chapter 7 rudimentarily includes the further developed Self-Formation Theory. It recognizes how technologies are increasingly "normalizing" what we consider "smart" and "good thinking," disclosing copious variations of criteria, both diachronic and synchronic, for what can be considered "cognitive enhancement." It also indicates that generalizing and substantializing particular framing of cognitive enhancement is reductive and, implicitly or explicitly, disregards other possible conceptions of "normal" and "enhanced" cognition.

Sublimation through Smart Drugs

TST could now be employed in order to understand what appropriating entails, that is, what is meant by actively and critically accepting and integrating smart drugs. In order to further specify how the theory could prove itself useful, I will focus on a particular kind of smart drug, namely memory enhancers. There have been many different smart drugs associated with memory improvement, like modafinil (provigil), d-amphetamine (adderall), methylphenidate (ritalin),

rivastigmine (exelon), galantamine (razadyne) and donepezil (aricept). For convenience, I will use donepezil as an example. As indicated, donepezil is a drug that is used to treat dementia in, for example, Alzheimer patients. Again, I do not claim here that it has been proven that this drug significantly improves memory in the long run. Instead, I take hypothetically as a starting point that it or an altered version of it does improve long-term verbal and visual episodic recall (which is confirmed in some studies and counterfeit in other; see Grön et al. 2005; Fitzgerald et al. 2008; Balsters et al. 2011), does not have severe side effects and does not come with legal issues. Let's call this improved version "donepezilplus" for convenience.

The first caveat of TST recognizes that smart drugs in general and donepezilplus in particular enables forming a preferred stable self and, at the same time, create an obstacle that prevents reaching the self's desired singularity. Donepezilplus could improve a healthy user's capacity to recall visual and verbal episodes to such a degree that she acquires a photographic memory that enables remembering everything that she experiences. Donepezilplus would (literally) open up a whole new world for this person. She would experience the world in eternal Platonic forms, not forgetting anything. Donepezilplus might, for example, offer this person who earlier worked as a policeman the possibility to become a world-famous profiler, solving crimes that no one else could solve. Instead of *phronesis*, one could say, contemplation has become her most dominant capacity. However, we need to acknowledge here the double-edged knife that characterizes donepezilplus as "other within": donepezilplus is both the necessary condition for becoming a world-renowned profiler and, at the same time, an obstacle to see and choose a different course of life. It is by virtue of donepezilplus that it becomes possible to have an extraordinary capacity, which, at the same time, becomes a capacity that can no longer be usurped and controlled, since it forms the self in a particular direction, in this case in a world-renowned profiler who structures the world the way world-renowned profilers do and only sees and actualizes possibilities that world-renowned profilers see and actualize. The more donepezilplus becomes a structural ingredient administered to her body, a permanent add-on to her mental world and an integrated part of her life, the less difficulties she will have attributing the capacity of photographic memory to herself. Until one day, donepezilplus is out of stock at her reliable pharmacy or internet store; then the "otherness" of donepezilplus will become visible again as well as the question of who she really is, the person on or off donepezilplus.

The second caveat of TST highlights how smart drugs could bring about different and new standards for determining what is "good self-formation." Donepezilplus not only allows a healthy person to better remember things but by radically transforming a healthy person's life, it also introduces new goals and expectations. Before the usage of

donepezilplus, being an excellent policeman was measured by, for example, securing the safety of the neighborhood. After popping donepezilplus, this accomplishment, one could imagine, is experienced as far from sufficient since the bar is set much higher now. Not being able to solve a big crime every month could be experienced as failure, as a waste of talent made possible by donepezilplus. In addition, this drug could introduce a "new normal" (Greely et al. 2008; Fitz et al. 2014). Nonusers of donepezilplus are not able to compete with users or not able to meet certain expectations that have become a norm and are forced to also use it. What was once a choice for people aspiring a more "adventurous" life could become a *de facto* necessity for all. What was "good performance" before the introduction of donepezilplus could be valued as "bad performance" after people started using it. Moreover, as a result of the usage of donepezilplus, photographic memory could become a highly valued capacity on the basis of which people compete and, hence, redefine what is considered a "smart" or even an "intelligent" person. Donepezilplus introduces new standards for what is "good performance" and transforms the meaning attributed to old ones.

Recognizing these caveats, TST aims to illustrate how it is possible to integrate these "intrusive technologies," which are increasingly shaping the very nature of the self while still allowing that a person can shape herself in a socially suitable and also singular way. The first step is to make visible which goals and ideals are embedded in or promoted by, in this case, the usage of the smart drug donepezilplus. We have already discussed the reasoning that makes this visible. Applied to donepezilplus, it would look like this: donepezilplus can be considered as a human enhancement technology and, hence, something that should be aspired toward, because (1) donepezilplus gives a person photographic memory, (2) giving people photographic memory will enable (*inter alia*) solving major crimes and, hence, creating a safer world and (3) safety is a preferred value. It is important to explicate here how "safety" is framed and operationalized in the context of using donepezilplus. Safety is understood as the result of solving major crimes, and that is seen as the result of, in part, revealing, mapping out and connecting major facts and events (data) related to a crime.

Second, appropriating "intrusive technologies" requires not only making visible, but also enabling the challenging of goals and ideals embedded in or promoted by the smart drug donepezilplus by virtue of other or longer-term goals and ideals. What is at stake is whether the mapped-out aims coincide with those identified after deliberation and reflection. In general, this challenge amounts to fighting reductionism. Although donepezilplus mediates our self-identification, the goals and ideals that we identify with in the process of realizing a preferred self cannot be fully captured by or reduced to the goals and ideals embedded in or promoted by this smart drug. Safety, for one, could be increased

by fighting crime using profilers with photographic memory. However, a more efficient way to fight crime and increase safety might be by offering educational programs and appropriate labor to people who live in bad neighborhoods and don't see many alternatives. After deliberate and critical reflection from different perspectives, our protagonist might also discover that using donepezilplus to solve major crimes is not her genuine passion but a symptom of trying to compensate for not meeting the expectations of her father who was a war hero. She still might decide to use donepezilplus in order to solve major crimes, now not to please her father but to create a better world for herself and her children. Or she might convince the CIA, her employer, to work with a highly sophisticated big data company that can do the profiling better than people on donepezilplus, and stop taking donepezilplus.

This case also reveals another aspect that should be considered. Having a pill that enhances memory may enable a student to get a high score on an exam or a former policeman to become a world-renowned profiler who makes the world a safer place, but it's also possible that, depending on particular circumstances, its long-term usage will lead to a whole new set of problems and disorders. Nietzsche considered what he sometimes called "active forgetting" a more valuable capacity than "memorizing." In the second essay of his *Untimely Meditations* (UM), he argued that it is possible to live almost without memory, and to live happily in that state, as animals demonstrate. However, he believed that it is altogether impossible to live at all without forgetting. Active forgetting in this context has nothing to do with memories fading away, but is a positive and active force, a capacity that an individual, a society and a culture need to live healthy and happy lives, have faith in themselves and create a new future. Active forgetting enables selective remembering. Active forgetfulness is an active ability to suppress negative and traumatic memories, which is a necessary condition for a culture to uphold and develop itself further. It is a kind of defense mechanism that blocks harmful thoughts, enables a positive spirit and makes place for the new (see also Aydin 2017). Nietzsche wrote, "The person [or group, or culture] in whom this apparatus of suppression is damaged, so that it stops working, can be compared (and not just compared) to a dyspeptic; he cannot 'cope' with anything" (GM II, 1). It is probably not a coincidence that psychiatry offices are full of people with unhappy memories they would like to get rid of. Not only for victims of horrendous emotional events such as accidental trauma, abuse and stressful relationships but also for more ordinary, happy people, memory enhancers could in the long run create more problems than they solve, precisely because forgetting seems to be a pivotal ingredient of happiness.

This digression neatly illustrates the challenge to integrate a smart drug in such a way that it contributes to realizing maximally inclusive ideals that, after thorough deliberation and critique, prove admirable.

Besides goals and ideals that concern functional and socially valued aspects, it is crucial here to also include goals and ideals that are of specific value for the particular person in question. In discovering what is good self-formation, taking into account social valuations is a necessary but not sufficient condition: taking donepezilplus might indeed greatly fit within the aspirations of one person and enable her to make the world a safer place, whereas for another person it might be a pathway to hell and the last thing that would make her feel safe. Witnessing horrors every day that remain hidden in the everyday lives of ordinary people might disclose a world where the latter does not want to live anymore, prompting her to seek another pathway.

The third aspect of TST entails sociopolitical measures that make visible and enable challenging the goals and ideals embedded in or promoted by smart drugs, in our case, the smart drug donepezilplus. Critically reflecting on and challenging these goals is not an individual matter but a collective challenge that requires the involvement and collaboration of different involved stakeholders. Doctors and pharmaceutical companies are responsible for making clear to users what the advantages and side effects of using donepezilplus are, which requires being trained to take a broader perspective and also consider social and psychological variables that might be relevant for the effect of the drug. Given the power of pharmaceuticals in our society, it is of upmost importance that economic interests do not lead to skewing the health benefits of this and other smart drugs. Lawmakers should limit abuse of this drug, and at the same time prevent creating incentives to illegally produce and sell it. Public discussion needs to increase awareness of the different, and sometimes contrasting, interests of users, sellers and producers of this kind of drugs as well as prompt reflection on the desirability and undesirability of a culture that incorporates and normalizes the use of smart drugs by healthy people. In this reflection, the motives of individuals who have idiosyncratic goals and ideals that do not completely fit in this culture should also be taken into account, though without making them dominant to such a degree that they can no longer be examined and questioned from other perspectives. As indicated repeatedly, individual goals may not always be authentic or critically acquired but could be projections of contingent "big others" inscribed in our unconsciousness.

The final step in the process of self-formation as sublimation is to secure a permanent space that enables challenging each particular goal, however beautiful it may seem. Projecting an ultimate, virtual and maximally inclusive ideal in the future creates this void. With respect to smart drugs, the ultimate ideal that can never be fully realized prevents critical reflection about how to integrate smart drugs in our society and lives to reach an end. Whether donepezilplus will make a person "smarter" and what "smarter" means will greatly depend on its particular integration in the very particular practices, outlook and preferences of the person

in question as well as her particular social context and circumstances. Some people might be able to successfully integrate a smart drug like donepezilplus in their life projects and evaluate it as enhancing their cognitive capacities. Others might value its effects very differently and even take an opposite stand, experiencing the improvement of long-term verbal and visual memory as a heavy yoke. The most important aim of the ultimate ideal is to secure room for singular valuations of smart drugs like donepezilplus, which make it possible to deviate from every particular characterization of "smartness"; "smartness" might involve getting faster at figuring out stuff, but being smart cannot be reduced to this element. The ultimate ideal enables disclosing not just functionalist values related to processing information, but also other values that prove to be important in being smart. If we want to prevent a murder, then having photographic memory that enables figuring out very fast how to find the perpetrator can be considered "smart." However, if our purpose is to go west, leave our old life behind us and start anew, then having photographic memory would not do us much good. Moreover, what can be considered "smart" will also depend on values and standards that currently do not even exist. They remain, to a great extent, virtual because they will be brought about by technologies of the future: sublime technologies that might shape sublime selves that we cannot imagine today, or monstrous technologies that might shape monstrous selves; or both.

Conclusion Part III

How ought we to form ourselves? And, more specifically, how ought we to form ourselves in a technological world? This was the primary question of the third part of this study, which contained four chapters. I opened this part arguing that transhumanists and other techno-optimists who propose approaching technological self-formation as radical self-enhancement are confronted with two problems, which I have successively discussed in the first two chapters. In Chapter 7, I demonstrated that transhumanists' and other techno-optimists' presumed notion of "enhancement" is unsustainable because radical self-transformation does not only imply enhancing the human by technologically improving her capacities and adding new functions to her constitution but also introducing new and different standards for establishing what is "enhancement." Transhumanists seem to be confused about their own conception of the posthuman. If variations of criteria, both diachronic and synchronic, of what can be considered "normal" and "enhanced" can already be observed in both past and present contexts and situations, what then gives the assurance that present standards can be sustained in the future, especially if that future is subject to radical and disruptive technological influences? I contrasted the transhumanist's notion of the posthuman with Nietzsche' notion of the Overhuman (*Übermensch*) and employed the latter as a "tuning fork" to "hear" whether transhumanists escape an essentialist conception in which one particular view of the enhanced or ideal human being is rendered absolute. Although transhumanists' visualizations of radical self-transformation claim to overcome an anti-essentialist anthropology, on reflection they display, as I have shown, the posthuman as essentially invariable, uniform and independent. This Nietzschean critique reveals the transhumanists' depiction of the posthuman is much too reassuring: it does not consider the possibility that transformations into a being that goes beyond the present human might also herald a world with radically different standards as well as the destruction of our present values, goals and ideals, including our most cherished (humanist) ideals. This prospect does not so much render the transhumanist project worthless, but it challenges

us to reflect much more soundly and seriously on the incentives and implications of the possible impact of emerging technologies and radical self-transformation. Moreover, a more adequate understanding of "technological enhancement" requires considering what the conditions are for allowing for divergent possibilities to form ourselves in different ways, which entails including in the assessment of "enhancement" other criteria besides efficiency and functionality.

   The second objection with regard to "enhancement" rhetoric, which I have discussed in Chapter 8, concerns disregarding that in attempts to technologically form ourselves we come across some-Thing within – something uncanny – that often can and does resist molding ourselves as we please. I examined this notion of the "uncanny within" through an analysis and evaluation of Masahiro Mori's the uncanny valley in robotics, reviewing different psychological hypothesis that attempt to explain the uncanny valley. Within that field, I highlighted Wang et al. (2015) *Dehumanization Hypothesis*, which takes mechanistic traits of humanoids, that is, a lack of emotion and warmth, as an explanation of uncanny responses to humanlike robots. Recognizing that this and other psychological explanations undoubtedly explain relevant aspects of the negative responses of certain humans to certain humanlike robots, I have proposed a more philosophical-existential perspective that addresses the looming "little big" question of "what makes humanness?" From this perspective, I have proposed that the uncanny valley might say at least as much about the viewer of humanoids as it says about humanoids, namely that a humanlike robot not only instigates awareness of what makes the robot different from the viewer but also what makes the viewer different from the robot as well as of the impossibility to appropriate that difference, which elicits uncanniness; it confronts the viewer of the robot with her own unfoundedness, which is and remains constitutively strange. Lacan's notion of "extimacy" was employed to further explain this experience of strangeness that is not simply the effect of something external intruding in the self; "ownness" structurally contains "otherness," we have seen. This other within was more explicitly interpreted in terms of "technology within" through complementing Lacan's notion of "extimacy" with Nancy's view of being "closed open" for intrusive technologies: in the confrontation with the familiar and, at the same time, strange robot, the self not only uncannily senses the human in the robot but also the robot in the human. If new and emerging technologies are becoming increasingly "extimate" and are shaping the very fabric of the self, understanding technologies as means that can enhance or treat the human disregards that the self cannot simply externalize technology and autonomously determine if and how it wants to use it in order to realize certain goals that it has set for itself, which, in fact, would be yet another shadow of an essentialist and dualist framework. Instead of the "faster, stronger, smarter" rhetoric, I have proposed to understand technological

self-formation in terms of "technological sublimation," which could also redefine the very notion of "human enhancement." This view was developed in the last two chapters of this part of the book.

In Chapter 9, I developed the notion of self-formation as sublimation on the basis of insights derived from Freud, Lacan, Nietzsche and Peirce. This notion of sublimation elucidates how technology is becoming an intrinsic part of our mind, body and self, and could disrupt the standards that we use to establish what we consider "normal," "healthy" or "enhanced" features and capacities. It recognizes this unsettled structure of the self and, at the same time, attempts to prevent adopting the view that the self is no more than a plaything of contingent influences and forces, unable to impart a stable and socially valued form to itself that it could consider *its own*. I showed that Freud understands sublimation as the redirecting of a repressed sexual drive toward a nonsexual and socially valued aim. Although he does not resolve the problem that social valuations could be both repressive and arbitrary, which does not warrant singularity, he does highlight an important dimension of sublimation, namely the constant tension between what society expects and what libidinal drives demand. Lacan takes up this tension and reinterprets and intensifies it on the basis of his three categories of the Real, the Imaginary and the Symbolic, accentuating that the social-cultural order cannot be simply externalized because it is a constitutive condition for developing a unified ego as well as moral and communicative subjectivity, and cannot completely absorb and nullify the stubborn drives that are also a part of the self. Complementing what Lacan says about the social-cultural order with a material-technological order, the tension between a technological culture and nature has been situated in the heart of self and self-formation: technology is increasingly becoming both the necessary condition for forming a unified, coherent and social self and an obstacle that prevents it from reaching the singularity that it seeks. From a Lacanian perspective, sublimation becomes a (paradoxical) challenge to (re)discover a void or emptiness that, instead of denying, enables usurping our drives, this way forming a singular self within a social order. I also argued that Nietzsche adds to this notion the importance of actively integrating as much facets of one's life as possible, including conflicting elements, and consolidating them in a preferred lifestyle. From this perspective, sublimation requires organizing intense internal struggle into a form that can bring about ever-new life forms, securing singularity.

Next, I have argued that Freud, Lacan and Nietzsche show considerable differences, but seem to all view society primarily as restricting the potential contained in the individual and, hence, as a danger for forming a singular self. I have employed Peirce to illustrate that society does not have to be as such a danger for singularity, but is rather a necessary condition for critical self-formation: the orientation toward societal goals

and ideals, and eventually toward an ultimate ideal, prevents the individual becoming trapped in one-sided convictions, including subjective ones, and creates room for forming a singular self that "has grown up under the influence of a course of self-criticisms and of heterocriticisms."

By bringing together the different and most important aspects of "sublimation," derived from ideas from the works of Freud, Lacan, Nietzsche and Peirce, and taking into account the role that technology plays in self-formation, I then outlined in a comprehensive theory what self-formation as sublimation in a technological world entails, a view which I have called Technological Sublimation Theory (TST): self-formation as sublimation involves, I have summarized, appropriating intrusive technologies, which are increasingly shaping the very fabric of the self – therefore, rendering complete control impossible and univocal standards unattainable – by virtue of an ongoing deliberate and critical identification with goals and ideals embedded in or promoted by those technologies, through sociopolitical measures and devotion to ever higher and more inclusive ideals that culminate in an ultimate, esthetic ideal that enables forming a stable and socially valued as well as a singular form.

In the tenth and final chapter of this part of the book, I have retrieved particular guidelines from TST and applied them to three technological fields: technological environments (shaping the self from the "outside"), brain imaging technologies (bridging and merging "inside" and "outside") and smart drugs (shaping the self from the "inside"). The smart elderly home was taken as an example of a smart environment to illustrate how self-formation as sublimation could be realized. I have tried to show how sublimation in this respect entails collectively finding out what would be a "good home" for the particular inhabitant through mapping out which goals are inscribed in the smart technologies of the house, and what are the interests, aims and limitations of the different relevant stakeholders, including those of the inhabitant, in the light of an ultimate ideal, that is, an ideal that permanently makes room for criticizing the different explicated values, goals and ideals in order to design the best house for that particular inhabitant. Neurofeedback has been taken as an example of brain imaging technologies. I have tried to show how sublimation entails a collective effort to map out which goals – in this case – neuroimaging seeks to realize, what are the interests, aims and limitations of relevant stakeholders, including those of the subject that is undergoing treatment, in the light of an ultimate ideal that enables criticizing the explicated values, goals and ideals from different perspectives and realizing what could be a healthy lifestyle for the person in question. As an example of self-formation as sublimation through smart drugs, I have taken "donepezilplus," a fictional improved version of the drug donepezil. Sublimation entailed, as we have seen, collectively making visible which goals donepezilplus seeks to realize, in the light of an ultimate, all-inclusive ideal that enables criticizing the displayed values and goals

of the different stakeholders, including the user of the drug, in order to integrate it in the very particular practices and outlook of the person in question as well as her particular social context and circumstances.

It must be clear that TST as well as the guidelines derived from it are not a philosopher's stone. The theory is programmatic, and the guidelines do not provide clear-cut directives. However, it does enable mapping out the most important conditions and provide relevant general guidelines for forming what could be considered a "good self," at the same time recognizing that that self has been influenced and shaped by particular "extimate" technologies. Displaying those conditions and offering those guidelines make, on the one hand, the paradoxical challenge of critically relating to something that cannot be fully explicated and objectified palpable; on the other hand, they indicate how to create a void that soaks off the rigid constraints that this technological order imposes on the self. They help, one could say, tame the "technological monster" that the human has become by unpacking and critically evaluating its inner drives and goals. Further elaborating and nuancing the application of TST to the proposed technological fields and examining other technological fields through the proposed lens as well as investigating the advantages and limitations observed in applying TST could contribute to sharpening the mapped-out conditions and guidelines and provide cues to sophisticate TST. In particular, the application to technologies such as big data, AI and algorithms, which are currently under development, could further test, modify and refine this theory.

Epilogue
Groundwork for a Philosophy of Existential Technology

In this study, I have gradually and progressively developed a particular notion of self-formation in a technological world. Thorough explanations were needed to show why an interactionist approach to self and self-formation beyond the inside-outside dualism is, on the one hand, difficult to grasp and sustain, and, on the other hand, philosophically tenable and doing the most justice to our current state of affairs. It is difficult to grasp, because the contested view of an invariable, independent and unified self is at the root of the development of Western philosophy and culture, and seems to be engraved in our everyday life *conceptions* of self and its relation to the world. It is difficult to sustain because attempts to overcome essentialist and dualist views of self and self-formation often quietly end up rehabilitating the very framework that they intend to overcome. At the same time, it is philosophically tenable because it is in line with and contributes to increasingly common contemporary pragmatist and phenomenological views of self and self-formation; and it also does justice to our current state of affairs because it complies with our everyday life *experience* of being strongly influenced and shaped by our increasingly technological environment. In attempts to overcome essentialist and dualist views in favor of an interactionist view, we seem to be confronted with a discrepancy between, on the one hand, how we *conceive* ourselves and our interactions with the world, and, on the other hand, how we *experience* ourselves and our interactions with the world. In theory we often are still essentialists and dualists, whereas in practice we are interactionists.

This book can be viewed as an attempt to contribute to overcoming this incongruity between theory and practice, taking as starting point our lived experience of being increasingly surrounded with and influenced by technologies that intrude more and more into our selfhood, without giving in to the view that it is impossible to provide a theory with explanatory power that exceeds a particular context, application or case study. It must be clear that the importance of particular settings, applications and case studies is by no means downplayed here, as also illustrated in the last chapter of this book. However, providing a suitable theoretical framework is of crucial importance to aptly grasp what is at

stake, recognize which normative questions a given context, application and case study raises and figure out how to tackle those questions. In order to answer subsequently what is the self, whether self-formation in a technological world is possible and how the self ought to be formed, I have engaged with a variety of authors and positions. After a critical, first-hand reading of Aristotle and Descartes' essentialist and dualist ontology and anthropology, Nietzsche's and Peirce's views were employed for establishing a basis for an interactionist perspective that displayed the self as something that in relating to itself and its environment discovers itself not as a substance but as a process and a challenge, that is, as a self impelled to form itself in a certain direction, which can be done more or less deliberately and critically.

The next step entailed, using Peirce as the main guide, going beyond the autonomy-determination opposition through debunking in different ways the inside-outside distinction. From the artifactual mind thesis, the mind, which traditionally has been viewed as an inner realm that secures freedom and autonomy, was demonstrated to be greatly determined by the "outside" world. This does not render normative questions irrelevant, but rather highlights even more the importance of critically evaluating the technologies that surround us and increasingly shape our mind and self. This normative approach, which presumes that deliberate and critical self-formation is at least not impossible, has been elaborated through adopting the distinction between, on the one hand, unconscious urges and shorter-term intentions and actions, and, on the other hand, longer-term goals and ideals. Because brain imaging technologies are often used by both advocates and opponents of free will as functional evidence, they have been examined from both a Cartesian and an interactionist perspective. From the proposed critical self-formation framework, these technologies are approached not as technologies that represent the inner world of the self, but rather, pragmatically, as technologies that could contribute to develop a preferred self by foregrounding both the neurological constraints and possibilities that underlie our (unwanted) habits, allowing the use of that extrospective knowledge to directly adjust old or develop new habits in the light of desired longer-term ideals. After illustrating how criticizing societal and institutional goals and ideals does not exclude but rather includes orientation toward (other or more inclusive) goals and ideals, the notion of freedom has been reinterpreted in terms of critical self-formation in light of evermore inclusive goals and ideals.

In the last part of the book, I criticized the view of technological self-formation as enhancement in two ways: first, I tried to illustrate that standards for "enhancement" are not univocal and, hence, are dependent on particular contexts and environments as well as on the technologies embedded in those contexts and environments, which themselves will shape the individuals interacting with those environments, including

their ethical standards. Then, I argued that being shaped by those environments, the self increasingly experiences technology as another within, as both familiar and unfamiliar, which also explains why technologies sometimes are experienced as uncanny. Because technologies are becoming extimate, the self is confronted with the challenge of how to critically relate to some-Thing within that has become a very part of its fabric. Confronted with these difficulties, next I have proposed that, building on Freud, Lacan, Nietzsche and Peirce, instead of conceiving self-formation as "enhancement," it is more feasible to understand it as "sublimation," or alternatively "enhancement" could be reinterpreted in terms of "sublimation." I have demonstrated that his view, on the one hand, attempts to recognize that our environment, which is increasingly a technological environment and subject to radical change, greatly shapes our present and future selves, and, on the other hand, resists the idea that the self is merely a product of random powers. The proposed Technological Sublimation Theory (TST) attempts to do justice to both the impact of the technological environment on the self and the potential of the self to impose a unique, singular form to itself. Applying the framework to three fields of technology (smart technological environments, brain imaging technologies and smart drugs), I have shown how the self is, respectively, increasingly becoming part of and being shaped by an outside technological infrastructure, how its "inside" world is reflected and explained by "outside" technologies and how it is shaped by outside technologies from the "inside," though without rendering critical self-formation, that is, the formation of both a socially valued and singular self, impossible.

The proposed interactionist view of self-formation in a technological world is, as repeatedly said, programmatic and in need of further elaboration in different ways and directions. I stipulate two relevant routes for further research. First of all, the ingredients used to prepare TST could be, to use a food metaphor, mixed more thoroughly in order to find out to what extent they are mixable or not mixable, which could result in strengthening the different building blocks of the interactionist view of self-formation. There might be also ingredients that need to be added to the existing ingredients. I will mention five relevant matters in that respect: first, Peirce's three categories of Firstness, Secondness and Thirdness have been related to three elements that are retrieved from Nietzsche's notions of the will to power and his idea of the Übermensch, namely potentiality, struggle and organization. A more thorough elaboration of those relations is needed to establish to what extent these dimensions could complement one another, building a stronger basis for developing an interactionist view of world and self.

Second, Lacan's triadic structure of Real Order, Imaginary Order and Symbolic Order seems to have similarities with both Nietzsche's three elements and Peirce's three categories. Comparing those lists of categories

could shed further light on especially the relation between the individual and society and provide a more sophisticated account of what it entails to form a self that, on the one hand, is and remains influenced by and indebted to the technological world that it is increasingly shaped by, and, on the other hand, could appropriate those influences and "capitalize" them in its own unique, singular fashion.

Third, with regard to similarities between Nietzsche and Lacan, examining particularly the relation between Lacan's the Thing and Nietzsche's Übermensch, might prove fruitful (see also Phillips 2015). Especially the detaching and disorienting power of both figures is striking. By comparing them and examining how they could complement one another, it might be possible to gain a more nuanced view of the normative significance of both figures as well as of the importance of building buffers that enable a "void" in moral and cultural constellations.

Fourth, the notion of the "sublime" has a tradition of its own, potentially related to monstrous, uncanny and extimate technology as well as to sublimation. Within that tradition, it is understood as something that exceeds the ordinary, which is also expressed aptly in the German word for the sublime, namely "das Erhabene" (the "exalted"). The sublime refers here to something that exceeds our very understanding, which sometimes is also associated with religious or transcendent experiences. Working out the different meanings of the "sublime" in relation to the proposed framework might produce very interesting results that enrich, among other things, what I have called the "esthetic" ideal of self-formation as well as TST (see also De Mul 2012).

Fifth, other views and concepts that have been mentioned or discussed in this book as well as views that have not been taken into account might be used to further elaborate, nuance, enrich and correct the developed framework. For example, Foucault's notion of "subjectivation" and other relevant concepts like "ethical substance" and "*telos*" (see Foucault 1984, 355) could complement and enrich (and possibly challenge) the proposed notion of technological appropriation and sublimation.

TST is intended to contribute to opening up and drawing attention to a new strand of research in the philosophy of technology, namely a philosophy of existential technology. In this study, much effort was devoted to elaborating how technologies in general, which traditionally are considered part of the "outside world," increasingly influence and intrude in our "inside world," blur the boundary between "inside" and "outside," making our "inside" to a great extent also *for us* an "outside"; hence, making technologies existential. However, TST remains to a great extent "general" and, hence, the earned "large banknotes and bills," to use a Husserlian expression (Husserl 1994, 5/56), need to be cashed out in "small change," which comprises the second route for further research. In other words, TST should contribute to a better understanding of how *particular* technologies influence and shape *particular*

existential phenomena, in traditional approaches often considered part of the human's inner world. How do particular (AI-based) technologies, such as smart phones, brain imaging technologies, navigation technologies and surveillance technologies, affect and shape particular phenomena like guilt, anxiety, hope, despair and especially desire? Turning large banknotes into small change requires a detailed and meticulous case-by-case examination and evaluation. TST could provide normative guidelines for how to deliberately and critically regulate those particular technologies and integrate them in our lives, imposing both a stable and a singular form to our existential moods. It goes without saying that the scope of this research is large; so, there is no shortage of work for the future.

Text Abbreviations

Plato

CR   Cratylus
ED   Euthydemus
ER   Euthyphro
PA   Parmenides
PD   Phaedo
PB   Philebus
SY   Symposium
TM   Timaeus
Notation in text: (work) (Stephanus number)
Work used: Plato (1997). J. M. Cooper (ed.). *Plato: Complete Works*. Indianapolis: Hackett.

Aristotle

AP   Analytica Posteriora
DA   De Anima
DC   De Caelo
GA   Generations of Animals
MT   Metaphysica
PY   Physica
Notation in text: (work) (Bekker number)
Work used: Aristotle. (1984). J. Barnes (ed.). *The Complete Works of Aristotle*. Princeton, NJ: Princeton University Press.

Descartes

Reg. Regulae ad directionem ingenii
Disc. Discours de la Methode
Med. Mediationes de prima Philosophia
Princ. Principia philosophiae
Resp. Responsio ad secundas objectiones
Notation in text: (work) (volume), (page): (Adam and Tannery [AT]) (volume), (page)

Work used: Descartes (1984, 1985, 1991). J. Cottingham, R. Stoothoff, and D. Murdoch (trans.). *The Philosophical Writings of Descartes*, Vol. 1, 2, 3. Cambridge: Cambridge University Press.

Nietzsche

KSA   Samtliche Werke, Kritische Studenausgabe
CW   The Case of Wagner
GM   Genealogy of Morals
UM   Untimely Meditations
GS   The Gay Science
BG   Beyond Good and Evil
WN   Writings from the Late Notebooks
SZ   Thus spoke Zarathustra
EH   Ecce Homo
TI   Twilight of the Idols
Notation in text: (work) (part) (fragment), (page)
Works used: Nietzsche, Friedrich. (1988). G. Colli and M. Montinari (eds.). *Friedrich Nietzsche, Sämtliche Werke. Kritische Studienausgabe in 15 Bänden.* München/Berlin/New York: De Gruyter; Nietzsche, Friedrich. (1967). *The Birth of Tragedy and the Case of Wagner* (W. Kaufmann, trans.). New York: Penguin Random House; the Nietzsche Cambridge University Press series.

Peirce

CP   Charles Peirce – Collected Papers
EP   Charles Peirce – Essential Peirce
M   Charles Peirce – Peirce Papers in the Houghton Library
Notation in text: (work) (volume), (page)
Works used: Peirce, Charles. (1931–1935, 1958). Ch. Hartshorne and P. Weiss and A. Burks (eds.). *Collected Papers of C.S. Peirce.* Cambridge: Harvard University Press; Peirce, Charles. (1973). R.S. Robins (ed.). *Peirce Papers in the Houghton Library.* Cambridge: Harvard University, Houghton Library; Peirce, Charles. (1992 and 1998). N. Hauser and N. Kloesel (eds.). *The Essential Peirce.* Bloomington and Indianapolis: Indiana University Press.

References

Aarts, Emile and Boris de Ruyter. (2009). "New Research Perspectives on Ambient Intelligence." *Journal of Ambient Intelligence and Smart Environments* 1: 5–14.

Abel, Gunter. (1984). *Nietzsche: Die Dynamik der Willen zur Macht un die Ewige Wiederkehr.* Berlin/New York: De Gruyter.

Adams, Fred and Kenneth Aizawa. (2001). "The Bounds of Cognition." *Philosophical Psychology* 14(1): 43–64.

Alborn, Timothy. (1989). "Peirce's Evolutionary Logic: Continuity, Indeterminacy, and the Natural Order." *Transactions of the Charles S. Peirce Society* 25: 1–28.

Andacht, Fernando and Mariela Michel. (2005). "A Semiotic Reflection on Self–interpretation and Identity." *Theory and Psychology* 15(1): 51–75.

Anders, Günther. (1956/2002). *Die Antiquiertheit des Menschen, Band 1: Über die Seele im Zeitalter der zweiten industriellen Revolution.* Munich: Beck.

Ansell-Pearson, Keith. (2011). "The Future Is Superhuman: Nietzsche's Gift." *The Agonist.* http://www.nietzschecircle.com/Pdf/The_Future_is_Superhuman.pdf

Aristotle. (1984). J. Barnes (ed.). *The Complete Works of Aristotle.* Princeton, NJ: Princeton University Press.

Armstrong, David. (1968/1993). *A Materialist Theory of Mind.* New York: Routledge.

Arpaly, Nomy. (2003). *Unprincipled Virtue: An Inquiry into Moral Agency.* Oxford: Oxford University Press.

Aydin, Ciano. (2002). "De Betekenis van Müller-Lauters Nietzsche-interpretatie." *Tijdschrift voor Filosofie* 64(4): 743–764.

Aydin, Ciano. (2003). *Zijn en Worden. Nietzsches Omduiding van het Substantiebegrip.* Maastricht: Shaker Publishing.

Aydin, Ciano. (2004) "Chaos." In P. van Tongeren, G. Schank, H.W. Siemens, Nietzsche Research Group (eds.). *Nietzsche-Wörterbuch*, Vol I. Berlin/New York: De Gruyter, 449–468.

Aydin, Ciano. (2006). "Beyond Essentialism and Relativism: Nietzsche and Peirce on Reality." *Cognitio* 7(1): 25–47.

Aydin, Ciano. (2007a). "Charles S. Peirce: Fenomenologie van Een, Twee en Drie." In C. Aydin (ed.). *De vele Gezichten van de Fenomenologie.* Kampen/Kapellen: Klement/Pelckmans, 197–218.

Aydin, Ciano. (2007b). "Naar een proces–pragmatische grondslag voor het identiteitsbegrip: Peirce over potentialiteit, interactie en regulariteit." *Tijdschrift voor Filosofie* 69(1): 35–78.

Aydin, Ciano. (2007c). "Nietzsche on Reality as Will to Power: Toward an 'Organization-Struggle Model.'" *Journal of Nietzsche Studies* 33: 25–48.

Aydin, Ciano. (2008). "The Struggle between Ideals: Nietzsche, Schmitt, and Lefort on the Politics of the Future." In H.W. Siemens and V. Rood (eds.). *Nietzsche, Power & Politics: Rethinking Nietzsche's Legacy for Political Thought*. Berlin/New York: De Gruyter, 801–816.

Aydin, Ciano. (2009). "On the Significance of Ideals. Peirce and the Good Life." *Transactions of the C.S. Peirce Society* 45: 422–443.

Aydin, Ciano. (2015). "The Artifactual Mind: Overcoming the 'Inside-Outside' Dualism in the Extended Mind Thesis and Recognizing the Technological Dimension of Cognition." *Phenomenology and the Cognitive Sciences* 14(7): 3–94.

Aydin, Ciano. (2017). "The Posthuman as Hollow Idol: A Nietzschean Critique of Human Enhancement." *Journal of Medicine and Philosophy* 42(3): 304–327.

Aydin, Ciano. (2018a). "From Camera Obscura to fMRI: How Brain Imaging Technologies Mediate Free Will." In J. Aagaard, J.K. Berg Friis, J. Sorenson, O. Tafdrup, and C. Hasse (eds.). *Postphenomenological Methodologies: New Ways in Mediating Techno-Human Relationships*. Lanham/Boulder, CO/New York/London: Lexington Books, 103–122.

Aydin, Ciano. (2018b). "World Oriented Self-formation as Sublimation: or Why Postphenomenology Needs Peircean Pragmatism." *Cognitio* 19(2): 204–219.

Aydin, Ciano, Margot González-Woge, and Peter-Paul Verbeek. (2019). "Technological Environmentality: Conceptualizing Technology as a Mediating Milieu." *Philosophy and Technology* 32(2): 321–338.

Babich, Babette. (2011). "Nietzsche's Post-human Imperative: On the 'All-Too-Human' Dream of Transhumanism." *The Agonist*. http://www.nietzschecircle.com/AGONIST/2011_08/Dream_of_Transhumanism.pdf

Bagdasaryan, Juliana and Michel Le Van Quyen. (2013). "Experiencing Your Brain: Neurofeedback as a New Bridge between Neuroscience and Phenomenology." *Frontiers in Human Neuroscience* 7: 680.

Bailey, Lee Worth. (1989). "Skull's Darkroom: The Camera Obscura and Subjectivity." In P.T. Durbin (ed.). *Philosophy of Technology: Practical, Historical and Other Dimensions*. Dordrecht: Kluwer Academic Publishers, 63–79.

Baker, Gordon. (1998). "The Private Language Argument." *Language and Communication* 18: 325–356.

Baker, Lynne Rudder. (2000). *Persons and Bodies*. Cambridge: Cambridge University Press.

Balaguer, Mark. (2015). "Determinism and Its Relevance to the Free-Will Question." In J. Clausen and N. Levy (eds.). *Handbook of Neuroethics*. Springer: Dordrecht, 231–252.

Balsters, Joshua H., Redmond G. O'Connell, Mary Martin, Alessandra Galli, Sarah Cassidy, Sophia Kilcullen, Sonja Delmonte, Sabina Brennan, Jim F. Meaney, Andrew J. Fagan, Arun L.W. Bokde, Neil Upton, Robert Lai, Marc Laruelle, Brian Lawlor, and Ian H. Robertson. (2011). "Donepezil Impairs Memory in Healthy Older Subjects: Behavioral, EEG and Simultaneous EEG/fMRI Biomarkers." *PLoS One* 6(9): e24126.

Barad, Karen. (2007). *Meeting the Universe Halfway: Quantum Physics and the Entanglement of Matter and Meaning*. Durham, NC/London: Duke University Press.

Baron, Robert A. (1997). "The Sweet Smell of... Helping: Effects of Pleasant Ambient Fragrance on Prosocial Behavior in Shopping Malls." *Personality and Social Psychology Bulletin* 23: 498–503.

Bartholdy, Savani, Peter Musiat, Iain C. Campbell, and Ulrike Schmidt. (2013). "The Potential of Neurofeedback in the Treatment of Eating Disorders: A Review of the Literature." *European Eating Disorders Review* 21(6): 456–463.

Bartneck, Christoph, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita. (2007). "Is the Uncanny Valley an Uncanny Cliff?" In *Proceedings of the 16th IEEE International Symposium on Robot and Human Interactive Communication*. New York: IEEE Press, 368–373.

Bateson, Gregory. (1973). *Steps to an Ecology of Mind*. London: Granada.

Baumeister, Roy F., E.J. Masicampo, and Kathleen D. Vohs. (2011). "Do Conscious Thoughts Cause Behavior?" *Annual Review of Psychology* 62: 331–361.

Beaulieu, Alain. (2002). "L'experience Deleuzienne du Corps." *Revue Internationale de Philosophie* 4: 511–522.

Berger, Herman. (1968). *Op zoek naar Identitet. Het Aristotelische Substantiebegrip en de Mogelijkheid van een Hedendaagse Metafysiek*. Nijmegen/Utrecht: Dekker en Vegt.

Bernstein, Richard. (1972). *Praxis and Action*. London: Duckworth.

Bernstein, Richard. (1980). "Action, Conduct, and Self-Control." In R. Bernstein (ed.). *Perspectives on Peirce*. Westport, CT: Greenwood Press, 66–91.

Berthelot, Rene. (1908). "Sur le Pragmatisme de Nietzsche." *Revue de Metaphysique et de Morale* 16: 403–447.

Berthelot, Rene. (1909). "Sur le Pragmatisme de Nietzsche." *Revue de Metaphysique et de Morale* 17: 654–702.

Berthelot, Rene. (1912). "Un Romantisme Utilitaire. Le pragmatisme chez Nietzsche et chez Poincaré." *Revue Philosophique de la France Et de l'Etranger* 73: 211–217.

Bibri, Simon E. (2015). *The Shaping of Ambient Intelligence and the Internet of Things*. Paris: Atlantis Press.

Block, Ned. (1978). "Troubles with Functionalism." In C.W. Savage (ed.). *Perception and Cognition: Issues in the Foundations of Psychology*. Minneapolis: University of Minnesota Press, 261–325.

Bostrom, Nick. (2002). "Existential Risks: Analyzing Human Extinction Scenarios and Related Hazards." *Journal of Evolution and Technology* 9. http://www.jetpress.org/volume9/risks.html

Bostrom, Nick. (2003a). "Human Genetic Enhancements: A Transhumanist Perspective." *Journal of Value Inquiry* 37: 493–506.

Bostrom, Nick. (2003b). "Transhumanist Values." In F. Adams (ed.). *Ethical Issues for the 21st Century*. Charlottesville, VA: Philosophical Documentation Center Press, 3–14.

Bostrom, Nick. (2003c). "The Transhumanist FAQ: v 2.1." *World Transhumanist Association*. https://www.nickbostrom.com/views/transhumanist.pdf

Bostrom, Nick. (2005). "A History of Transhumanist Thought." *Journal of Evolution and Technology* 14. http://jetpress.org/volume14/bostrom.html

Bostrom, Nick. (2008). "Why I Want to Be a Posthuman When I Grow Up." In B. Gordijn and R. Chadwick (eds.). *Medical Enhancement and Posthumanity*. New York: Springer, 107–137.

Bostrom, Nick. (2013). "Existential Risk Prevention as Global Priority." *Global Policy* 4: 15–31.

Bostrom, Nick and Rebecca Roache. (2008). "Ethical Issues in Human Enhancement." In J. Ryberg, T.S. Petersen, and C. Wolf (eds.). *New Waves in Applied Ethics*. New York: Pelgrave Macmillan, 120–152.

Bowell, Rebecca and Hedy Kober. (2016). "Food Cue Reactivity and Craving Predict Eating and Weight Gain: A Meta-Analytic Review." *Obesity Reviews* 17(2): 159–177.

Brey, Philip. (2000). "Technology as Extension of Human Faculties." In C. Mitcham (ed.). *Research in Philosophy and Technology*, Vol. 19. London: Elsevier/JAI Press, 59–78.

Brey, Philip. (2008). "Human Enhancement and Personal Identity." In J.K. Berg Olsen, E. Selinger, and S. Riis (eds.). *New Waves in Philosophy of Technology*. New York: Palgrave Macmillan, 169–185.

Brink, David. (1997). "Moral Motivation." *Ethics* 108: 4–32.

Burge, Tyler. (1988) "Individualism and Self-Knowledge." *Journal of Philosophy* 85(11): 649–663.

Buss, Sarah. (2012). "Autonomous Action: Self-determination in the Passive Mode." *Ethics* 122: 647–691.

Cacioppo, John. T., Gary. G. Berntson and Howard C. Nusbaum. (2008). "Neuroimaging as a New Tool in the Toolbox of Psychological Science." *Current Directions in Psychological Science* 17(2): 62–67.

Callon, Michel. (1986). "The Sociology of an Actor-Network: The Case of the Electric Vehicle." In M. Callon, J. Law, and A. Rip (eds.). *Mapping the Dynamics of Science and Technology: Sociology of Science in the Real World*. Houndmills, UK: Macmillan, 19–34.

Camerer, Colin, George Loewenstein and Drazen Prelec (2005). "Neuroeconomics: How Neuroscience Can Inform Economics." *Journal of Economic Literature* 43(1): 9–64.

Campbell, Scott. (2006). "The Conception of a Person as a Series of Mental Events." *Philosophy and Phenomenological Research* 73: 339–358.

Campbell, Tim and Jeff McMahan. (2010). "Animalism and the Varieties of Conjoined Twinning." *Theoretical Medicine and Bioethics* 31: 285–301.

Carruthers, Peter. (2011). *The Opacity of Mind: An Integrative Theory of Self-Knowledge*. Oxford: Oxford University Press.

Chen, Yuan-Jyue, Neil Dalchau, Niranjan Srinivas, Andrew Phillips, Luca Cardelli, David Soloveichik, and Georg Seelig. (2013). "Programmable Chemical Controllers Made from DNA." *Nature Nanotechnology* 8: 755–762.

Choudhury, Suparna, Saskia Kathi Nagel, and Jan Slaby. (2009). "Critical Neuroscience: Linking Neuroscience and Society through Critical Practice." *BioSocieties* 4(1): 61–77.

Christoff, Kalina, Diego Cosmelli, Dorothe Legrand, and Evan Thomspon. (2011). "Specifying the Self for Cognitive Neuroscience." *Trends in Cognitive Sciences* 15(3): 104–112.

Churchland, Paul. (1981). "Eliminative Materialism and the Propositional Attitudes." *The Journal of Philosophy* 78(2): 67–90.

Clark, Andy. (1997). *Being There*. Cambridge: The MIT Press.

Clark, Andy. (2003). *Natural-Born Cyborgs: Mind, Technologies, and the Future of Human Intelligence*. Oxford: Oxford University Press.

Clark, Andy. (2008). *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. Oxford: Oxford University Press.

Clark, Andy. (2010). "Coupling, Constitution, and the Cognitive Kind: A Reply to Adams and Aizawa." In R. Menary (ed.). *The Extended Mind*. Cambridge: The MIT Press, 81–100.

Clark, Andy and David Chalmers. (1998). "The Extended Mind." *Analysis* 58: 7–19.

Clark Julia and Marilyn McGee-Lennon. (2011). "A Stakeholder-Centred Exploration of the Current Barriers to the Uptake of Home Care Technology in the UK." *Journal of Assistive Technologies* 5(1): 12–25.

Clowes, Robert W. (2015). "Thinking in the Cloud: The Cognitive Incorporation of Cloud–Based Technology." *Philosophy and Technology* 28(2): 261–296.

Colapietro, Vincent. (1989). *Peirce's Approach of the Self. A Semiotic Perspective on Human Subjectivity*, Albany: State University of New York Press.

Colapietro, Vincent. (2006). "Toward a Pragmatic Conception of Practical Identity." *Transactions of the Charles Peirce Society* 42(2): 173–205.

Conant, James. (2004). "Why Worry About the Tractatus?" In B. Stocker (ed.). *Post-Analytic Tractatus*. Aldershot: Ashgate, 167–192.

Dalibert, Lucie. (2014). *Posthumanism and Somatechnologies: Exploring the Intimate Relations between Humans and Technologies*. Enschede: University of Twente (CTIT PhD thesis series No 14–303).

Danto, Arthur C. (1964). "Nietzsche." In D.J.O. O'Conner (ed.). *A Critical History of Western Philosophy*. London: The Free Press, 384–401.

Danto, Arthur C. (1965/1980). *Nietzsche as Philosopher*. New York: Columbia University Press.

Danziger, Kurt. (1994). *Constructing the Subject: Historical Origins of Psychological Research*. Cambridge: Cambridge University Press.

Darley, John M. and Bibb Latane. (1968). "Bystander Intervention in Emergencies: Diffusion of Responsibility." *Journal of Personality and Social Psychology* 8: 377–383.

Dartnall, Terry. (2005). "Does the World Leak into the Mind? Active Externalism, 'Internalism,' and Epistemology." *Cognitive Science* 29: 135–143.

Davey, Christopher, Jesus Pujol and Ben J. Harrison. (2016). "Mapping the Self in the Brain's Default Mode Network." *Neuroimage* 132: 390–397.

Davidson, Donald. (1984). *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press.

Debrock, Guy. (1992). "Peirce, a Philosopher for the 21st Century: Introduction." *Transactions of the Charles S. Peirce Society* 28: 1–18.

Debrock, Guy. (2003a). "Introduction: Process Pragmatism." In G. Debrock (ed.). *Process Pragmatism. Essays on a Quiet Philosophical Revolution*. Amsterdam/New York: Rodopi, 1–10.

Debrock, Guy. (2003b). "Ethics and Pragmatic Process Philosophy." In G. Debrock (ed.). *Process Pragmatism. Essays on a Quiet Philosophical Revolution*. Amsterdam/New York: Rodopi, 159–170.

Dehaene, Stanislas, Elizabeth Spelke, Philippe Pinel, Ruxandra Stanescu and Sanna Tsivkin (1999). "Sources of Mathematical Thinking: Behavioral and Brain Imaging Evidence." *Science* 284: 970–974.

De Mul, Jos. (2012). "The (Bio)Technological Sublime." *Diogenes* 59: 32–40.

Dennett, Dan. (1991). *Consciousness Explained*. London: Penguin Books.

Dennett, Dan. (2003). "Who's on First? Heterophenomenology Explained." *The Journal of Consciousness Studies* 10: 19–30.

de Preester, Helena. (2011). "Technology and the Body: The (Im)Possibilities of Re-Embodiment." *Foundations of Science* 16: 119–137.

de Waal, Cornelis. (2006). "Science Beyond the Self. Remarks on Charles S. Peirce's Social Epistemology." *Cognitio* 7(1): 149–163.

Deretić, Irina and Stefan Sorgner, eds. (2016). *From Humanism to Meta-, Post- and Transhumanism?* New York: Peter Lang.

Descartes, Rene. (1965). *Discourse on Method, Optics, Geometry, and Meteorology* (P.J. Olscamp, trans). New York/Kansas City, MO: The Bobbs-Merrill Company, Inc.

Descartes (1984, 1985, 1991). J. Cottingham, R. Stoothoff, and D. Murdoch (trans.). *The Philosophical Writings of Descartes*, Vol. 1, 2, 3. Cambridge: Cambridge University Press.

Descartes, Rene. (1641/1985). "Meditations on First Philosophy." In J. Cottingham, R. Stoothoff, and D. Murdoch (trans.). *The Philosophical Writings of Descartes*, Vol. 2. Cambridge: Cambridge University Press, 1–62.

Diels, Hermann and Walther Kranz. (eds.) (1956). *Die Fragmente der Vorsokratiker*, Vol. 2. Berlin: Weidmann.

Doliński, Dariusz, Tomasz Grzyb, Michal Folwarczny, Patrycja Grzybała, Karolina Krzyszycha, Karolina Martynowska, and Jakub Trojanowski. (2017). "Would You Deliver an Electric Shock in 2015? Obedience in the Experimental Paradigm Developed by Stanley Milgram in the 50 Years Following the Original Studies." *Social Psychological and Personality Science* 8(8): 927–933.

Donadio, Stephen (1978). *Nietzsche, Henry James and the Artistic Will*. New York: Oxford University Press.

Donald, Merlin. (1991). *Origins of the Modern Mind: Three Stages in the Evolution of Culture and Cognition*. Cambridge: Harvard University Press.

Doris, John M. (2002). *Lack of Character: Personality and Moral Behavior*. New York: Cambridge University Press.

Doris, John M. (2015). *Talking to Our Selves: Reflection, Ignorance, and Agency*. Oxford: Oxford University Press.

Dretske, Fred. (1988). *Explaining Behavior: Reasons in a World of Causes*. Cambridge: MIT Press.

Dretske, Fred. (1995). *Naturalizing the Mind*. Cambridge: The MIT Press.

Dretske, Fred. (1996). "Phenomenal Externalism, or if Meanings Ain't in the Head, Where Are Qualia?" In E. Villanueva (ed.). *Philosophical Issues 7*. Atascadero: Ridgeview Publishing Company, 143–158.

Dretske, Fred. (1999). "The Mind's Awareness of Itself." *Philosophical Studies* 95: 103–124.

Dubljević, Velijko. (2013). "Cognitive Enhancement, Rational Choice and Justification." *Neuroethics* 6(1): 179–187.

Dumit, Joseph. (2004). *Picturing Personhood: Brain Scans and Biomedical Identity*. Princeton, NJ: Princeton University Press.

Durkheim, Emile. (1993/lectures:1913/1914) "Pragmatismus und Soziologie." In H. Joas (ed.) and M. Bischoff (trans.). *Schriften zur Soziologie der Erkenntnis*. Frankfurt a.M.: Suhrkamp, 1–168.

Durt, Christoph, Thomas Fuchs, and Christian Tewes. (eds.) (2017). *Embodiment, Enaction, and Culture: Investigating the Constitution of the Shared World*. Cambridge: MIT Press.

Eastman, Nigel and Colin Campbell. (2006). "Neuroscience and Legal Determination of Criminal Responsbility." *Nature Reviews Neuroscience* 7: 311–318.

Ehrenhard, Michel Leon, Bjosrn Kijl, and Lambertus Johannes Maria Nieuwenhuis. (2014). "Market Adoption Barriers of Multi-Stakeholder Technology: Smart Homes for the Aging Population." *Technological Forecasting and Social Change* 89: 306–315.

Ellul, Jacques. (1977). *Le système technician*. Paris: Calmann-Lévy.

Engelbart, Doug C. (1962). "Augmenting Human Intellect: A Conceptual Framework." *Summary Report,* SRI Project No. 3578, AFOSR–3223, Contract AF 49(638)–1024.

Erny, Nicola. (2005). *Konkrete Vernünftigkeit*. Tübingen: Mohr Siebeck.

Fahmy, Melissa Seymour. (2011). "On the Supposed Moral Harm of Selecting for Deafness." *Bioethics* 25: 128–136.

Faraci, David and David Shoemaker. (2010). "Insanity, Deep Selves, and Moral Responsibility: The Case of JoJo." *Review of Philosophy and Psychology* 1: 319–332.

Farrer, Chloe, and Chris D. Frith. (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.

Feenberg, Andrew. (2002). *Transforming Technology*. New York: Oxford University Press.

Fenichel, Otto. (1945). *The Psychoanalytic Theory of Neurosis*. New York: Norton.

Ferrando, Francesca. (2013). "Posthumanism, Transhumanism, Antihumanism, Metahumanism, and New Materialisms: Differences and Relations." *Existenz* 8(2): 26–32.

Ferrari, Francesco, Maria Paola Paladino, and Joland Jetten. (2016). "Blurring Human–Machine Distinctions: Anthropomorphic Appearance in Social Robots as a Threat to Human Distinctiveness." *International Journal of Social Robotics* 8(2): 287–302

Ferrey, Anne, Tyler J. Burleigh, and Mark J. Fenske. (2015). "Stimulus-Category Competition, Inhibition, and Affective Devaluation: A Novel Account of the Uncanny Valley." *Frontiers in Psychology* 6(249): 35–49.

Fitz, Nicholas S., Roland Nadler, Praveena Manogaran, Eugene W.J. Chong, and Peter B. Reiner. (2014). "Public Attitudes Toward Cognitive Enhancement." *Neuroethics* 7(2): 173–188.

Fitzgerald, David B., Gregory P. Crucian, Jeannine Mielke, Brian V. Shenal, David Burks, Kyle B. Womack, Georges Ghacibeh, Valeria Drago, Paul S. Foster, Edward Valenstein, and Kenneth M. Heilman. (2008). "Effects of Donepezil on Verbal Memory After Semantic Processing in Healthy Older Adults." *Cognitive and Behavioral Neurology* 21(2): 57–64.

Floridi, Luciano. (2014). *The Fourth Revolution: How the Infosphere is Reshaping Human Reality*. Oxford: Oxford University Press.

Foot, Philippa. (1957). "Free Will as Involving Determinism." *The Philosophical Review* 66: 439–450.

Forlini, Cynthia and Eric Racine. (2009). "Autonomy and Coercion in Academic 'Cognitive Enhancement' Using Methylphenidate: Perspectives of Key Stakeholders." *Neuroethics* 2(3): 163–177.

Foucault, Michel. (1984). "On the Genealogy of Ethics: An Overview of Work in Progress." In P. Rabinow (ed.). *The Foucault Reader*. New York: Pantheon, 340–372.

Foucault, Michel. (1988). "Technologies of the Self." In L.H. Martin, H. Gutman and P.H. Hutton (eds.). *Technologies of the self: A seminar with Michel Foucault*. Amherst: University of Massachusetts Press, 16–49.

Foucault, Michel. (1989). *Resume des course, 1970–1982*. Paris: Julliard.

Foucault, Michel. (1997). "The Ethics of Concern for Self as a Practice of Freedom." In P. Rabinow (ed.). *Ethics, Subjectivity and Truth*. New York: The New Press, 281–302.

Frank, M. (2009). "Reading My Mind." *CBS News*. http://www.cbsnews.com/news/reading-my-mind/

Frankfurt, Harry G. (1971). "Freedom of the Will and the Concept of a Person." *The Journal of Philosophy* 68: 5–20.

Freud, Sigmund. (1981). J. Strachey (ed.). *The Standard Edition of the Complete Works of Sigmund Freud*. London: Hogarth Press.

Fried, Itzhak, Roy Mukamel and Gabriel Kreiman. (2011). "Internally Generated Preactivation of Single Neurons in Human Medial Frontal Cortex Predicts Volition." *Neuron* 69: 548–562.

Friedman, Lesley. (1995). "C.S. Peirce's Transcendental and Immanent Realism." *Transactions of the Charles S. Peirce Society* 31(2): 374–392.

Friedman, Lesley. (1997). "Peirce's Reality and Berkeley's Blunders." *Journal of the History of Philosophy* 35(2): 253–268.

Frischmann, Brett and Evan Selinger. (2018). *Re-engineering Humanity*. Cambridge: Cambridge University Press.

Frood, Arran. (2018). "Use of 'Smart Drugs' on the Rise." *Nature*. https://www.nature.com/articles/d41586-018-05599-8

Fukuyama, Francis. (2002). *Our Posthuman Future: Consequences of the Biotechnology Revolution*. New York: Farrar, Straus, and Giroux.

Fuss, Diana. (1989). *Essentially Speaking: Feminism, Nature, and Difference*. New York: Routledge.

Gallagher, Shaun. (2007). "Neurophilosophy and Neurophenomenology." In L. Embree and T. Enon (eds.). *Phenomenology 2005*, Vol. 5. Bucharest: Zeta Press, 293–316.

Gallagher, Shaun. (2011). "Interpretations of Embodied Cognition." In W. Tschacher and C. Bergomi (eds.). *The Implications of Embodiment: Cognition and Communication*. Exeter: Imprint Academic, 59–70.

Gallagher, Shaun. (2013). "A Pattern Theory of Self." *Frontiers in Human Neuroscience* 7: 443.

Gallagher, Shaun and Anthony Crisafi. (2009). "Mental Institutions." *Topoi* 28: 45–51.

Gallagher, Shaun and Anya Daly. (2018). "Dynamical Relations in the Self-Pattern." *Frontiers in Psychology* 9: 664.

Gallagher, Shaun and Francisco J. Varela. (2003). "Redrawing the Map and Resetting the Time: Phenomenology and the Cognitive Sciences." *Canadian Journal of Philosophy* 29(Suppl.): 93–132.

Gallie, Walter Bryce. (1966). *Peirce and Pragmatism*. New York: Dover Publications.

Garreau, Joel. (2005). *Radical Evolution: The Promise and Peril of Enhancing our Minds, our Bodies – and What it Means to be Human*. New York: Doubleday.

Garret, Brian. (1998). *Personal Identity and Self-Consciousness*. London: Routledge.

Garrisson, Kathleen A., Dustin Scheinost, Patrick D. Worhunsky, Hani M. Elwafi, Thomas A. Thornhill, Evan Thompson, Clifford Saron, Gaelle Desbordes, Kober Hedy, Michelle Hampson, Jeremy R. Gray, R. Todd Constable, Xenophon Papademetris, and Judson A. Brewer. (2013). "Real-time fMRI links Subjective Experience with Brain Activity during Focused Attention." *Neuroimage* 81: 110–118.

Gazzaniga, Michael S. (2005). *The Ethical Brain*. Washington, DC: Dana Press.

Geller, Tom. (2008). "Overcoming the Uncanny Valley." *IEEE Computer Graphics and Applications* 28: 11–17.

Gemes, Ken. (2009). "Freud and Nietzsche on Sublimation." *Journal of Nietzsche Studies* 38: 38–59.

Gevensleben, Holger, Birgit Holl, Bjoern Albrecht, Dieter Schlamp, Oliver Kratz, Petra Studer, Susanne Wangler, Aribert Rothenberger, Gunther H. Moll, and Hartmut Heinrich. (2009). "Distinct EEG Effects Related to Neurofeedback Training in Children with ADHD: A Randomized Controlled Trial." *International Journal of Psychophysiology* 74: 149–157.

Gibson, James J. (1979). *The Ecological Approach to Visual Perception*. Boston, MA: Houghton Mifflin.

Giddens, Anthony. (1991). *Modernity and Self-identity. Self and Society in the Late Modern Age*. Cambridge: Polity Press.

Gilihan, Seth J. and Martha J. Farah. (2005). "Is Self Special? A Critical Review of Evidence from Experimental Psychology and Cognitive Neuroscience." *Psychological Bulletin* 131(1): 76–97.

Glock, Hans-Johann. (1986). "Vygotsky and Mead on the Self, Meaning and Internalisation." *Studies in Soviet Thought* 31: 131–148.

Godfrey-Smith, Peter (2014). *Philosophy of Biology*. Princeton, NJ: Princeton University Press.

Goldman, Rachel L., Jeffrey J. Borckardt, Heather A. Frohman, Patrick M. O'Neil, Alok Madan, Laura K. Campbell, Amanda Budak, and Mark S. George. (2011). "Prefrontal Cortex Transcranial Direct Current Stimulation (Tdcs) Temporarily Reduces Food Cravings and Increases the Self-Reported Ability to Resist Food in Adults with Frequent Food Craving." *Appetite* 56(3): 741–746.

Gordon, Robert M. (2007). "Ascent Routines for Propositional Attitudes." *Synthese* 159: 151–165.

Granier, Jean. (1966). *Le Problème de la Vérité dans la Philosophie de Nietzsche*. Paris: Éditions du Seuil.

Gray, Heather M., Kurt Gray, and Daniel M. Wegner. (2007). "Dimensions of Mind Perception." *Science* 315(5812): 619.

Greely, Henry, Barbara Sahakian, John Harris, Ronald C. Kessler, Michael Gazzaniga, Philip Campbell, and Martha J. Farah. (2008). "Towards Responsible Use of Cognitive-Enhancing Drugs by the Healthy." *Nature* 456(7223): 702–705.

Greene, Joshua D., Leigh E. Nystrom, Andrew D. Engell, John M. Darley, and Jonathan D. Cohen. (2004). "The Neural Bases of Cognitive Conflict and Control in Moral Judgment." *Neuron* 44(2): 389–400.

Greeno, James G. (1994). "Gibson's Affordances." *Psychological Review* 101: 336–342.

Grön, Georg, Matthias Kirstein, Axel Thielscher, Matthias Riepe, and Manfred Spitzer. (2005). "Cholinergic Enhancement of Episodic Memory in Healthy Young Adults." *Psychopharmacology* 182: 170–179.

Guthrie, Stewart E. (1993). *Faces in the Clouds*. New York: Oxford University Press.

Haack, Susan. (1992). "'Extreme Scholastic Realism:' Its Relevance to Philosophy of Science Today." *Transactions of the Charles S. Peirce Society* 28(1): 19–50.

Haack, Susan. (1997). "Pragmatism and Ontology: Peirce and James." *Revue Internationale de Philosophy* 31(121/122): 377–400.

Habermas, Jurgen. (2003). *The Future of Human Nature*. Cambridge, UK: Polity Press.

Hacking, Ian. (1986). "Making Up People." In T.L. Heller, M. Sosna and D.E. Wellbery (eds.). *Reconstructing Individualism: Autonomy, Individuality, and the Self in Western Thought*. Stanford, CA: Stanford University Press, 222–236.

Haggard, Patrick. (2017). "Sense of Agency in the Human Brain." *Nature Reviews Neuroscience* 18: 196–207.

Hallman, Max O. (1985). "Nietzsche and Pragmatism." *Kinesis* 14: 63–78.

Hanson, David. (2005). "Expanding the Aesthetic Possibilities for Humanoid Robots." In *IEEE-RAS International Conference on Humanoid Robots*. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.472.2518andrep=rep1andtype=pdf

Hanson, Mark J. (1999). "Indulging Anxiety: Human Enhancement from a Protestant Perspective." *Christian Bioethics* 5: 121–138.

Happ, Heinz. (1971). *Hyle: Studien zum aristotelischen Materie-Begriff*. Berlin/New York: De Gruyter.

Haraway, Donna. (1991). "A Cyborg Manifesto: Science, Technology, and Socialist-Feminism in the Late Twentieth Century." In D. Haraway (ed.). *Simians, Cyborgs and Women: The Reinvention of Nature*. New York: Routledge, 149–181.

Harman, Gilbert. (1999). "Moral Philosophy Meets Social Psychology: Virtue Ethics and the Fundamental Attribution Error." *Proceedings of the Aristotelian So*ciety 99: 315–331.

Harman, Gilbert. (2002). "The Nonexistence of Character Traits." *Proceedings of the Aristotelian Society* 100: 223–226.

Harman, Gilbert. (2003). "No Character or Personality." *Business Ethics Quarterly* 13: 87–94.

Harris, John. (2007). *Enhancing Evolution*. Princeton, NJ/Oxford: Princeton University Press.

Harris, Sam. (2012). *Free Will*. New York: Free Press.

Hartshorne, Charles. (1964). "Charles Peirce's 'One contribution to Philosophy' and His Most Serious Mistake." In Ch. Hartshorne, E.C. Moore, and R. Robin (eds.). *Studies in the Philosophy of Charles Sanders Peirce*. Amherst: The University of Massachusetts Press.

Hartshorne, Charles. (1980). "A Revision of Peirce's Categories." *Monist* 63(3): 277–289.

Hassin, Ran R., James S. Uleman and John A. Bargh. (eds.) (2005). *The New Unconscious*. New York: Oxford University Press.

Hauskeller, Michael. (2010). "Nietzsche, The Overhuman and The Posthuman: A Reply to Stefan Sorgner." *Journal of Evolution and Technology* 21. https://jetpress.org/v21/hauskeller.htm

Hausman, Carl. (1990). "In and Out of Percepts." *Transactions of the Charles S. Peirce Society* 26: 271–308.

Hausman, Carl. (1993). *Charles S. Peirce's Evolutionary Philosophy.* Cambridge: Cambridge University Press.

Hausman, Carl. (2007). "Metaphorical Semeiotic Referents: Dynamic Objects as Dyads." *The Transactions of the Charles S. Peirce Society* 43(2): 276–287.

Haynes, John-Dylan and Geraint Rees. (2006). "Decoding Mental States from Brain Activity in Humans." *Nature Reviews Neuroscience* 7: 523–534.

Hegel, Georg Wilhelm Friedrich. (1949) *Philosophy of Right* (T.M. Knox, trans.). Oxford: Clarendon Press.

Heidegger, Martin. (1961). *Nietzsche*. Volume I and II. Pfullingen: Neske.

Heimsoeth, Heinz. (1958). *Die sechs grossen Themen der abendländischen Metaphysik und der Ausgang des Mittelalters.* Darmstadt: Wissenschaftliche Buchgesellschaft.

Hingst, Kai-Michael. (1998). *Perspektivismus und Pragmatismus. Ein Vergleich auf der Grundlage der Wahrheitsbegriff und der Religionphilosophien von Nietzsche und James.* Würzburg: Königshausen & Neumann.

Hilpinen, Risto. (1993). *Authors and Artifacts. Proceedings of the Aristotelian Society* 93: 155–178.

Ho, Chin-Chang, Karl F. MacDorman, and Z.A. Dwi Pramono. (2008). "Human Emotion and the Uncanny Valley: A GLM, MDS, and Isomap Analysis of Robot Video Ratings." In *Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction,* 169–176.

Hongladarom, Soraj. (2016). *The Online Self: Externalism, Friendship and Games*. Dordrecht: Springer.

Hookway, Christopher. (1985). *Peirce.* London/Boston/Melbourne/Henley: Routledge/Kegan Paul.

Hookway, Christopher. (2000). *Truth, Rationality, and Self-Control: Themes from Peirce*. Oxford: Oxford University Press.

Hudson, Hud. (2001). *A Materialist Metaphysics of the Human Person.* Ithaca: Cornell University Press.

Hulswit, Menno. (2002). *From Cause to Causation: A Peircean Perspective.* Dordrecht: Kluwer Academic Publishers.

Hume, David. (1978). *Treatise of Human Nature*. Oxford: Clarendon Press.

Hurley, Susan L. (2003). *Justice, Luck, and Knowledge*. Cambridge, Mass.: Harvard University Press.

Hurley, Susan L. (2010). "The Varieties of Externalism." In R. Menary (ed.). *The Extended Mind*. Cambridge: The MIT Press, 101–154.

Husserl, Edmund. (1994). K. Schuhmann (ed.). *Briefwechsel*. Husserliana Dokumente III/1–10. Dordrecht: Kluwer.

Hutchins, Ed. (2011). "Enculturating the Supersized Mind." *Philosophical Studies* 152: 437–446.

Hutchinson, Phil and Rupert Read. (2005). "*Memento*: A Philosophical Investigation." In R. Read and J. Goodenough (eds.). *Film as Philosophy: Essays in Cinema After Wittgenstein and Cavell*. Hampshire: Palgrave Macmillan.

Ihde, Don. (1990). *Technology and the Lifeworld* (The Indiana Series in the Philosophy of Technology, ed). Bloomington and Minneapolis: Indiana University Press.

Ihde, Don. (2007). "Art Precedes Science." In P. Kockelkoren (ed.). *Mediated Vision*. Arnhem: ArtEZ/Veenman Publishing, 25–37.

Ingold, Tim. (2000). *The Perception of the Environment: Essays on Livelihood, Dwelling and Skill*. London/New York: Routledge.

James, Klem. (2009). "Breton, Bataille and Lacan s Notion of Transgressive Sublimation." *E-pisteme* 2(1): 53–66.

Jentsch, Ernst. (1906/2008). "On the Psychology of the Uncanny." In J. Collins and J. Jervis (eds). *Uncanny Modernity: Cultural Theories, Modern Anxieties*. London: Palgrave Macmillan, 216–228.

Johnston, Mark. (2007). "'Human Beings' Revisited: My Body is not an Animal." In D. Zimmerman (ed.). *Oxford Studies in Metaphysics*, Vol. 3. Oxford: Oxford University Press, 33–74.

Jones, Owen and Francis Shen. (2012). "Law and Neuroscience in the United States." *International Neurolaw: A Comparative Analysis*: 349. SSRN. https://ssrn.com/abstract=2001085

Julien, Philippe. (1990). *Le retour à Freud de Jacques Lacan*. Paris: E.P.E.L.

Kaipayil, Joseph. (2009). *Relationalism: A Theory of Being*. Bangalore: JIP Publications.

Kant, Immanuel. (1785/2003). *Fundamental Principles of the Metaphysics of Morals* (T.M. Abbott, trans.). Rockville, MD: Arc Manor.

Kant, Immanuel. (1997). *Groundwork of the Metaphysics of Morals* (M. Gregor, trans.). Cambridge: Cambridge University Press.

Kaplan, Frederic. (2004). "Who is Afraid of the Humanoid? Investigating Cultural Differences in the Acceptance of Robots." *International Journal of Humanoid Robotics* 3: 465–480.

Kapp, Ernst. (1877). *Grundlinien einer Philosophie der Technik: Zur Entstehungsgeschichte der Cultur aus Neuen Gesichtpunkten*. Braunschweig: Westermann.

Kass, Leon. (2002). *Life, Liberty and the Defense of Dignity: The Challenge for Bioethics*. San Francisco: Encounter Books.

Kiran, Asle. (2012) "Technological Presence: Actuality and Potentiality in Subject Constitution." *Human Studies* 35(1): 77–93.

Kiran, Asle H. and Peter-Paul Verbeek. (2010). "Trusting Our Selves to Technology." *Knowledge, Technology & Policy* 23(3–4): 409–427.

Kirchhoff, Michael D. (2012). "Extended Cognition and Fixed Properties: Steps to a Third-wave Version of Extended Cognition." *Phenomenology and the Cognitive Sciences* 11: 287–308.

Klein, Colin. (2010). "Images are not the Evidence in Neuroimaging." *The British Journal for the Philosophy of Science* 61(2): 265–278.

Klein, Colin. (2011) "The Dual Track Theory of Moral Decision-Making: A Critique of the Neuroimaging Evidence." *Neuroethics* 4(2):143–162.

Klein, Colin. (2017). Brain Regions as Difference-Makers. *Philosophical Psychology* 30(1/2): 1–20.

Kleinpeter, Hans. (1912). "Die Erkenntnislehre Friedrich Nietzsches." *Wissenschaftliche Rundschau* 3: 5–9.

Klemm, William R. (2010). "Free Will Debates: Simple Experiments are Not So Simple." *Advances in Cognitive Psychology* 6: 47–65.

Knappett, Carl. (2005). *Thinking Through Material Culture: An Interdisciplinary Perspective*. Philadelphia, PA: University of Pennsylvania Press.

Kockelkoren, Petran. (2003). *Technology: Art, Fairground and Theatre*. Rotterdam: NAi Publishers.

Kockelkoren, Petran. (2007). "The Artist as Researcher?" In P. Kockelkoren (ed.). *Mediated Vision*. Rotterdam: Veenman Publishers/ArtEZ Press, 129–152.

Koechlin, Etienne, and Hyafil, Alexandre (2007). Anterior Prefrontal Function and the Limits of Human Decision-Making. *Science* 318(5850): 594–598.

König, Josef. (1967). "Einige Bemerkungen über den formalen Charakter des Unterschieds von Ding und Eigenschaft." In E. Fries (ed.). *Festschrift für Joseph Klein zum 70. Geburtstag*. Göttingen.

Korsgaard, Christine. (2009). *Self-Constitution: Agency, Identity, and Integrity*. Oxford: Oxford University Press.

Kosko, Bart. (1999). *The Fuzzy Future: From Society and Science to Heaven in a Chip*. New York: Random House/Harmony Books.

Kurzweil, Ray. (2005). *The Singularity is Near: When Humans Transcend Biology*. New York: Penguin.

Kutas, Marta and Kara D. Federmeier. (1998). "Minding the Body." *Psychophysiology* 35(2): 135–150.

Lacan, Jacques. (1988). *The Seminar, Book I, Freud's Papers on Technique, 1953–1954* (J. Forrester, trans.). New York: Norton & Co.

Lacan, Jacques. (1997). J.A. Miller (ed.) and D. Porter (trans.). *The Ethics of Psychoanalysis 1959–1960. The Seminar of Jacques Lacan Book VII*. New York: Norton.

Lacan, Jacques. (2005). *Écrits: A Selection* (A. Sheridan, trans.). London: Routledge.

Lacan, Jacques. (2006). *Écrits* (B. Fink, trans.). New York and London: Norton.

Lacan, Jacques. (2014). J.A. Miller (ed.) and A.R. Price (trans.). *Anxiety. The Seminar of Jacques Lacan Book X*. Cambridge: Polity Press.

Lachs, John. (1980). "Peirce, Santayana, and the Large Facts." *Transactions of the Charles S. Peirce Society* 16: 3–13.

Laland, Kevin N., John Odling-Smee, and Marcus W. Freedman. (2000). "Niche Construction, Biological Evolution and Cultural Change." *Behavioral and Brain Sciences* 23: 131–146.

Lamme, Victor. (2006). "Towards a True Neural Stance on Consciousness." *Trends in Cognitive Sciences* 10: 494–501.

Lamme, Victor. (2010). *De Vrije wil bestaat niet. Over wie er echt de baas is in het brein*. Amsterdam: Bert Bakker.

Laplanche, Jean and Jean-Bertrand Pontalis. (1973). *The Language of Psychoanalysis*. New York: Norton.

Latane, Bibb and John Darley (1968). "Bystander Intervention in Emergencies: Diffusion of Responsibility." *Journal of Personality and Social Psychology* 8: 377–383.

Latane, Bibb and John Darley (1970). *The Unresponsive Bystander: Why Doesn't He Help?* New York: Appleton and Century Crofts.

Latane, Bibb and Judith Rodin (1969). "A Lady in Distress: Inhibiting Effects of Friends and Strangers or Bystander Intervention." *Journal of Experimental Psychology* 5: 189–202.

Latour, Bruno. (1992). "Where Are the Missing Masses? The Sociology of a Few Mundane Artifacts." In Wiebe E. Bijker and John Law (eds.). *Shaping Technology/Building Society*. Cambridge: MIT Press, 225–258.

Latour, Bruno. (1993). *We Have Never Been Modern* (C. Porter, trans.). Cambridge, MA: Harvard University Press.

Lee, Younghwa, Jintae Lee, and Zoonky Lee (2006). "Social Influence on Technology Acceptance Behavior: Self-identity Theory Perspective." *ACM SIGMIS Database* 37(2–3): 60–75.

Legrand, Dorothée and Perrine Ruby (2009). "What is Self-Specific? Theoretical Investigation and Critical Review of Neuroimaging Results." *Psychological Review* 116(1): 252–282.

Lettow, Susanne. (2011). "Somatechnologies: Rethinking the Body in Philosophy of Technology." *Techné: Research in Philosophy and Technology* 15(2):110–117.

Levinas, Emmanuel. (1961/1969). *Totality and Infinity, An Essay on Exteriority* (A. Lingis, trans.). Pittsburgh: Duquesne University Press.

Levinson, Paul. (1988). *Mind at Large: Knowing in the Technological Age*. Greenwich, CT: JAI Press.

Lewis, Clarence Irving. (1929), *Mind and the World-Order*. New York: Dover.

Lewis, Clarence Irving. (1946). *An Analysis of Knowledge and Valuation*. La Salle, IL: Open Court.

Libet, Benjamin, Curtis A. Gleason, Elwood W. Wright, and Dennis Keith Pearl. (1983). "Time of Conscious Intention to Act in Relation to Onset of Cerebral Activity (Readiness-Potential). The Unconscious Initiation of a Freely Voluntary Act." *Brain A Journal of Neurology* 106: 623–642.

Libet, Benjmain. (2011). "Do We Have Free Will?" In W. Sinott-Armstrong and L. Nadel (eds.). *Conscious Will and Responsibility*. New York: Oxford University Press.

Littlefield, Melissa Monique. "Constructing the Organ of Deceit: The Rhetoric of fMRI and Brain Fingerprinting in Post-9–11 America." *Science Techology and Human Values* 34(3): 365–392.

Ljubisavljevic, Milos, K. Maxood, Jovana Bjekic, Joji Oommen, and N. Nagelkerke. (2016). "Long-Term Effects of Repeated Prefrontal Cortex Transcranial Direct Current Stimulation (tDCS) on Food Craving in Normal and Overweight Young Adults." *Brain Stimulation* 9(6): 826–833.

Logothetis, Nikos K. (2008) "What We Can Do and What We Cannot Do with fMRI." *Nature* 453: 869–878. doi:10.1038/nature06976.

Lokhorst, Gert-Jan. (2008). "Descartes and the Pineal Gland." In E.N. Zalta (ed.). *The Stanford Encyclopedia of Philosophy.* http://plato.stanford.edu/archives/sum2011/entries/pineal-gland/

Löwith, Karl. (1967). *Gott, Mensch und Welt in der Metaphysik von Descartes bis zu Nietzsche*. Göttingen: Vandenhoeck & Ruprecht.

Lutz, Antoine and Evan Thompson. (2003). "Neurophenomenology: Integrating Subjective Experience and Brain Dynamics in the Neuroscience of Consciousness." *Journal of Consciousness Studies* 10(9/10): 31–52.

Macdonald, Cynthia. (2007). "Introspection and Authoritative Self-Knowledge." *Erkenntnis* 67: 355–372.

MacDorman, Karl. (2006). "Subjective Ratings of Robot Video Clips for Human Likeness, Familiarity, and Eeriness: An Exploration of the Uncanny Valley." In *ICCS/CogSci-2006 Long Symposium: Toward Social Mechanisms of Android Science.* http://www.coli.uni-saarland.de/courses/agentinteraction/contents/papers/MacDorman06short.pdf

MacDorman, Karl F. and Hiroshi Ishiguro. (2006). "The Uncanny Advantage of Using Androids in Cognitive and Social Science Research." *Interaction Studies: Social Behaviour and Communication in Biological and Artificial Systems* 7: 297–337.

MacDorman, Karl F., Robert Green, Chin Chang Ho, and Clinton T. Koch. (2009). "Too Real for Comfort? Uncanny Responses to Computer Generated Faces." *Computers in Human Behavior* 25: 695–710.

Majumder, Sumit, Emad Aghayi, Moein Noferesti, Hamidreza Memarzadeh-Tehran, Tapas Mondal, Zhibo Pang, and M Jamal Deen. (2017). "Smart Homes for Elderly Healthcare – Recent Advances and Research Challenges." *Sensors* 17(11): 2496.

Malabou, Catherine. (2008). *What Should We Do with Our Brain?* (S. Rand, trans.). New York: Fordham University Press.

Malafouris, Lambros. (2004). "The Cognitive Basis of Material Engagement: Where Brain, Body and Culture Conflate." In E. DeMarrais, C. Gosden, and C. Renfrew (eds.). *Rethinking Materiality: The Engagement of Mind with the Material World*. Cambridge: McDonald Institute for Archaeological Research, 53–62.

Malafouris, Lambros. (2008a). "At the Potter's Wheel: An Argument *for* Material Agency." In C. Knappett, and L. Malafouris (eds.). *Material Agency: Towards a Non-Anthropocentric Approach.* New York: Springer, 19–36.

Malafouris, Lambros. (2008b). "Beads for a Plastic Mind: the 'Blind Man's Stick' (BMS) Hypothesis and the Active Nature of Material Culture." *Cambridge Archaeological Journal* 18(3): 401–414.

Malafouris, Lambros. (2013). *How Things Shape the Mind: A Theory of Material Engagement*. Cambridge: MIT Press.

Marsh, Jeffrey L. (2006). "To Cut or Not to Cut?" In E. Parens (ed.). *Surgically Shaping Children*. Baltimore, MA: John Hopkins University Press, 113–124.

Marty, Robert. (1997). *76 Definitions of the Sign by C. S. Peirce*. http://www.cspeirce.com/menu/library/resources/76defs/76defs/htm

Mathur, Maya B. and David B. Reichling. (2016). "Navigating a Social World with Robot Partners: A Quantitative Cartography of the Uncanny Valley." *Cognition* 146: 22–32.

Matthews, Paul and Peter Jezzard. (2004). "Functional Magnetic Resonance Imaging." *Journal of Neurology, Neurosurgery, and Psychiatry* 75(1): 6–12.

Mayer, Robert. (1867) *Mechanik der Wärme*. Stuttgart: Cotta.

McCluhan, Marshall. (1966/1964). *Understanding Media: The Extensions of Man*. New York: McGraw-Hill. Paperback edition.

McGinn, Robert E. (1980). "Nietzsche on Technology." *Journal of the History of Ideas* 41: 679–691.McLuhan, Marshall and Quentin Fiore. (1996). *The Medium is the Message*. London: Penguin Modern Classics.

McLuhan, Marshall and Terence Gordon. (2003). *Understanding Media: The Extensions of Man*. Corte Madera, CA: Gingko Press.

Mele, Alfred R. (2006). "Free Will. Theories, Analysis, and Data." In S. Pockett, W.P. Banks and S. Gallagher (eds.). *Does Consciousness Cause Behavior?* Cambridge: MIT Press, 187–205.

Menary, Richard. (2006). "Attacking the Bounds of Cognition." *Philosophical Psychology* 19: 329–344.

Menary, Richard. (2007). "Writing as Thinking." *Language Sciences* 29: 621–632.

Menary, Richard. (2008). "Embodied Narratives." *Journal of Consciousness Studies* 15(6): 63–84.

Menary, Richard. (2009). "Intentionality, Cognitive Integration and the Continuity Thesis." *Topoi* 28: 31–43.

Merleau-Ponty, Maurice. (1962). *Phenomenology of Perception* (C. Smith, trans.). London: Routledge.

Metzinger, Thomas. (2003). *Being No One*. Cambridge: MIT Press.

Meyers, Robert. (1999). "Pragmatism and Peirce's Externalist Epistemology." *Transactions of the Charles S. Peirce Society* 35(4): 638–653.

Micoulaud-Franchi, Jean-Arthur, Clelia Quiles, Guillaume Fond, Michel Cermolacce and Jean Vion-Dury. (2014). "The Covariation of Independent and Dependant Variables in Neurofeedback: A Proposal Framework to Identify Cognitive Processes and Brain Activity Variables." *Consciousness and Cognition* 26: 162–168.

Milgram, Stanley. (1974). *Obedience to Authority: An Experimental View*. New York: Harper and Row.

Milne, Bruce. (1982). *Know the Truth: A Handbook of Christian Belief*. Leicester, UK: Intervarsity Press.

Mitchell, Wade J., Kevin A. Szerszen, Sr., Amy Shirong Lu, Paul W. Schermerhorn, Matthias Scheutz, and Karl F. Macdorman. (2011). "A Mismatch in the Human Realism of Face and Voice Produces an Uncanny Valley." *i-Perception* 2: 10–12.

Mobbs, Dean, Hakwan Lau, Owen D. Jones, and Chris D. Frith. (2007). "Law, Responsibility and the Brain." *PloS Biology* 5(4): 103.

Moles, Alistair. (1990). *Nietzsche's Philosophy of Nature and Cosmology*. New York: Peter Lang.

Monroe, Andrew E. and Bertram Malle. (2010). "From Uncaused Will to Conscious Choice: The Need to Study, not Speculate about People's Folk Concept of Free Will." *Review of Philosophy and Psychology* 1: 211–224.

Moran, Richard. (2001) *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton, NJ: Princeton University Press.

Moravec, Hans. (1999). *Robot: Mere Machine to Transcendent Mind*. New York: Oxford University Press.

More, Max. (1990). "Transhumanism: Towards a Futurist Philosophy." *Extropy* 6: 6–12.

More, Max. (2010). "The Overhuman in the Transhuman." *Journal of Evolution and Technology* 21. https://jetpress.org/v21/more.htm

Mori, Masahiro. (1970/2012). "The Uncanny Valley." K. MacDorman and N. Kageki (trans.). *IEEE Robotics and Automation Magazine* 19: 98–100.

Moriarty, Jane Campbell. (2008). "Flickering Admissability: Neuroimaging Evidence in US Courts." *Behavioral Sciences and the Law* 26: 29–49.

Morton, W. Alexander and Gwenolyn G. Stockton. (2000). "Methylphenidate Abuse and Psychiatric Side Effects." *Primary Care Companion to the Journal of Clinical Psychiatry* 2(5): 159–164.

Müller-Freienfels, Richard. (1912). "Nietzsche und der Pragmatismus." *Archiv für Geschichte der Philosophie* 26: 339–358.

Müller Lauter, Wolfgang. (1971). *Nietzsche. Seine Philosophie der Gegensätze und die Gegensätze seiner Philosophie.* Berlin/New York: De Gruyter.

Müller-Lauter, Wolfgang. (1999a). *Über Freiheit und Chaos.* Berlin/New York: De Gruyter.

Müller-Lauter, Wolfgang. (1999b), *Über Werden und Wille zur Macht. Nietzsche-Interpretationen I.* Berlin/New York: De Gruyter.

Müller-Lauter, Wolfgang. (2000). *Heidegger und Nietzsche.* Berlin/New York: De Gruyter.

Muoio, Patricia A. (1984). "Peirce on the Person." *Transactions of the Charles S. Peirce Society* 20(2): 169–181

Myers, Ella. (2008). "Resisting Foucauldian Ethics: Associative Politics and the Limits of the Care of the Self." *Contemporary Political Theory* 7: 127.

Naam, Ramez. (2004). *More than Human: Embracing the Promise of Biological Enhancement.* New York: Joseph Henry Press.

Näätänen, Ristro, Anne Lehtokoski, Mietta Lennes, Marie Cheour, Minna Huotilainen, Antti Livonen, Martti Vainio, Paavo Alku, Risto J. Ilmoniemi, Aavo Luuk, Juri Allik, and Janne Sinkkonen. (1997). "Language-specific Phoneme Representations Revealed by Electric and Magnetic Brain Responses." *Nature* 385: 432–434.

Nair, Dinesh G. (2005). About being BOLD. *Brain Research Reviews* 50: 229–243.

Nancy, Jean-Luc. (2008). "The Intruder." In R. Rand (trans.). *Corpus.* New York: Fordham University Press, 161–170.

Natorp, Paul. (1921). *Platos Ideenlehre. Eine Einführung in den Idealismus.* Leipzig: Felix Meiner.

Nietzsche, Friedrich. (1967). *The Birth of Tragedy and the Case of Wagner* (W. Kaufmann, trans.). New York: Penguin Random House.

Nietzsche, Friedrich. (1988). G. Colli and M. Montinari (eds.). *Friedrich Nietzsche, Sämtliche Werke. Kritische Studienausgabe in 15 Bänden.* München/Berlin/New York: De Gruyter.

Nietzsche, Friedrich. (1997a). K. Ansell-Pearson (ed.). *On the Genealogy of Morality.* Cambridge: Cambridge University Press.

Nietzsche, Friedrich. (1997b). D. Breaseale (ed.) and R. Hollingdale (trans.). *Untimely Meditations.* Cambridge: Cambridge University Press.

Nietzsche, Friedrich. (2001). B. Williams (ed.). *The Gay Science: with a Prelude in German Rhymes and an Appendix of Songs.* Cambridge: Cambridge University Press.

Nietzsche, Friedrich. (2002). R.P. Horstmann and J. Norman (eds.). *Beyond Good and Evil.* Cambridge: Cambridge University Press.

Nietzsche, Friedrich. (2003). R. Bittner (ed.). *Writings from the Late Notebooks.* Cambridge: Cambridge University Press.

Nietzsche, Friedrich. (2006). R. Pipin (ed.). and A. del Caro (trans.). *Thus Spoke Zarathustra.* Cambridge: Cambridge University Press.

Nisbett, Richard E. Timothy DeCamp Wilson. (1977). "Telling More Than We Can Know: Verbal Reports on Mental Processes." *Psychological Review* 84: 231–259.

Noonan, Harold W. (1989). *Personal Identity.* New York: Routledge.

Northoff, Georg, Alexander Heinzel, Moritz de Greck, Felix Bermpohl, Henrik Dobrowolnu, and Jaak Panksepp. (2006). "Self-Referential Processing in our

Brain – A Meta-Analysis of Imaging Studies on the Self." *Neuroimage* 31: 440–457.

Olson, Eric. (1997). *The Human Animal: Personal Identity Without Psychology.* New York: Oxford University Press.

Olson, Eric (2003) An Argument for Animalism. In R. Martin and J. Barresi (eds.) *Personal Identity. Blackwell Readings in Philosophy.* Blackwell: Oxford, 318–33.

Papacharissi, Zizi. (ed.) (2011). *A Networked Self: Identity, Community, and Culture on Social Network Sites.* New York: Routledge.

Parens, Erik. (ed.) (1998). *Enhancing Human Traits: Ethical and Social Issues.* Washington, DC: Georgetown University Press.

Parfit, Derek. (2012). "We Are Not Human Beings." *Philosophy* 87: 5–28.

Pariser, Eli. (2011). *The Filter Bubble: What the Internet is Hiding from You.* New York: The Penguin Press.

Pauen, Michael. (2004). "Does Free Will Arise Freely?" *Scientific American Mind* 14: 40–47.

Pauen, Michael. (2007). "Self-Determination: Free Will, Responsibility, and Determinism." *Synthesis Philosophica* 44: 455–475.

Peirce, Charles. (1887). "Logical Machines." *American Journal of Psychology* 1(1): 165–170.

Peirce, Charles. (1901). "Review of Fraser's Berkeley." *Nation* 73: 95–96.

Peirce, Charles. (1931–1935, 1958). Ch. Hartshorne and P. Weiss and A. Burks (eds.). *Collected Papers of C.S. Peirce.* Cambridge: Harvard University Press.

Peirce, Charles. (1973). R.S. Robins (ed.). *Peirce Papers in the Houghton Library.* Cambridge: Harvard University, Houghton Library.

Peirce, Charles. (1992 and 1998). N. Hauser and N. Kloesel (eds.). *The Essential Peirce.* Bloomington and Indianapolis: Indiana University Press.

Phillips, Luke. (2015). "Sublimation and the Übermensch." *Journal of Nietzsche Studies* 46(3): 349–366.

Plato (1997). J. M. Cooper (ed.). *Plato: Complete Works.* Indianapolis: Hackett.

Plessner, Helmuth. (1975). *Die Stufen des Organischen und der Mensch: Einleitung in die philosophische Anthropologie.* Berlin/New York: De Gruyter.

Poldrack, Russell A. (2006). "Can Cognitive Processes Be Inferred from Neuroimaging Data?" *Trends in Cognitive Sciences* 10: 59–63.

Poldrack, Russell A. (2011). "Inferring Mental States from Neuroimaging Data: From Reverse Inference to Large-Scale Decoding." *Neuron* 72(5), 692–697.

Poletti, Anna and Julie Rak. (eds.) (2014). *Identity Technologies: Constructing the Self Online. Madison.* Madison: The University of Wisconsin Press.

Poliakoff, Ellen, Natalie Beach, Rebecca Best, Toby Howard, and Emma Gowen. (2013). "Can Looking at a Hand Make your Skin Crawl? Peering into the Uncanny Valley for Hands." *Perception* 42: 998–1000.

Popper, Karl. (1972). *Objective Knowledge: An Evolutionary Approach.* Oxford: Oxford University Press.

Potter, Vincent. (1997). *Charles S. Peirce. On Norms and Ideals.* New York: Fordham University Press.

Quine, Willard van Orman. (1960). *Word and Object.* Cambridge: MIT Press.

Quine, Willard van Orman. (1969). *Ontological Relativity.* New York: Columbia University Press.

Quinton, Anthony. (1962). "The Soul." *Journal of Philosophy* 59: 393–403; reprinted in Perry (ed.), 1975.

Rasmussen, David. (1995). "Rethinking Subjectivity: Narrative Identity and the Self." *Philosophy and Social Criticism* 21(5–6): 159–172.

Richardson, John. (1996). *Nietzsche's System*. Oxford: Oxford University Press.

Ricoeur, Paul. (1990). *Soi-même Comme un Autre*. Paris: Le Seuil.

Rieff, Philip. (1965). *Freud: The Mind of the Moralist*. London: Methuen.

Ringo, Allegra. (2013). "Understanding Deafness: Not Everyone Wants to Be 'Fixed.'" *The Atlantic*. http://www.theatlantic.com/health/archive/2013/08/understanding-deafness-not-everyone-wants-to-be-fixed/278527/

Robertson, Brian. (2015). *Lacanian Antiphilosophy and the Problem of Anxiety: An Uncanny Little Object*. New York: Palgrave Macmillan.

Rorty, Richard. (1972). "The World Well Lost." *Journal of Philosophy* 69: 649–665.

Rorty, Richard. (1982). *Consequences of Pragmatism*. Minneapolis, MN: Harvester.

Rorty, Richard. (1989). *Contingency, Irony and Solidarity*. Cambridge: Cambridge University Press.

Rorty, Richard. (1991). *Essays on Heidegger and Others*. Cambridge: Cambridge University Press.Rosenberger, Robert and Verbeek, Peter-Paul. (eds.) (2015). *Postphenomenological Investigations: Essays on Human-Technology Relations*. London: Lexington Books.

Rosenthal, Sandra B. (1994). *Charles Peirce's Pragmatic Pluralism*. Albany: State University of New York Press.

Rosenthal, Sandra B. (1996). "Continuity, Contingency, and Time: The Divergent Intuitions of Whitehead and Pragmatism." *Transactions of the Charles S. Peirce Society* 32(4): 542–567.

Rosenthal, Sandra B. (2001). "Firstness and the Collapse of Universals." *The Commens Encyclopedia: The Digital Encylcopedia of Peirce Studies*. http://www.digitalpeirce.fee.unicamp.br/firros.htm

Roskies, Adina L. (2007). "Are Neuroimages Like Photographs of the Brain?" *Philosophy of Science* 74(5): 860–872.

Rouse, Joseph. (1996). *Engaging Science: How to Understand its Practices Philosophically*. Ithaca: Cornell University Press.

Rupert, Robert D. (2004). "Challenges to the Hypothesis of Extended Cognition." *Journal of Philosophy* 101: 389–428.

Russell, Bertrand. (1910). "Knowledge by Acquaintance and Knowledge by Description." *Proceedings of the Aristotelian Society* 11: 108–128.

Russell, Bertrand. (1985). In D. Pears (ed.). *The Philosophy of Logical Atomism*. La Salle, IL: Open Court, 35–155.

Ryle, Gilbert. (1949). *The Concept of Mind*. New York: Barnes and Noble.

Sahakian, Barbara and Sharon Morein-Zamir. (2007). "Professor's Little Helper." *Nature* 450(20): 1157–1159.

Sartre, Jean-Paul. (1943). *Being and Nothingness: An Essay on Phenomenological Ontology* (H.E. Barnes, trans.). London: Routledge.

Savulescu, Julian. (2001). "Procreative Beneficence: Why We Should Select the Best Children." *Bioethics* 15: 413–426.

Saygin, Ayse Pinar, Thierry Chaminade, Hiroshi Ishiguro, Jon Driver, and Chris Frith. (2012). "The Thing that Should Not Be: Predictive Coding and

the Uncanny Valley in Perceiving Human and Humanoid Robot Actions." *Social Cognitive and Affective Neuroscience* 7: 413–422.Scanlon, Tim. (1998). *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.

Schaefer, G. Owen, Guy Kahane and Julian Savulescu. (2014). "Autonomy and Enhancement." *Neuroethics* 7: 123–136.

Schechtman, Marya. (ed.). (1996). *The Constitution of Selves*. Ithica: Cornell University Press.

Scheler, Max. (1926/1977). In M.S. Frings (ed.). *Erkenntnis und Arbeit. Eine Studie über Wert und Grenzen des pragmatischen Motivs in der Erkenntnis der Welt*. Frankfurt a.M.: Klostermann.

Schick, Kathy D. and Nicholas Toth. (1993). *Making Silent Stones Speak: Human Evolution and the Dawn of Technology*. New York/London: Simon and Schuster.

Seidl, Horst. (1989). *Aristoteles' Metaphysik*. Hamburg: Meiner Felix Verlag.

Selinger, Evan. and Timothy Engström. (2008). "A Moratorium on Cyborgs: Computation, Cognition, and Commerce." *Phenomenology and the Cognitive Sciences* 7: 327–341.

Settles, Burr. (2012). *Active Learning (Synthesis Lectures on Artificial Intelligence and Machine Learning)*. San Rafael, CA: Morgan and Claypool Publishers.

Seyama, Jun'ichiro and Ruth S. Nagayama. (2007). "The Uncanny Valley: Effect of Realism on the Impression of Artificial Human Faces." *Presence* 16: 337–351.

Shoemaker, Sidney. (2011). "On What We Are." In S. Gallagher (ed.). *The Oxford Handbook of the Self*. Oxford: Oxford University Press.

Shoemaker, Sidney. (1984). "Personal Identity: A Materialist Account." In S. Shoemaker and R. Swinburne (eds.) *Personal Identity*. Oxford: Blackwell, 67–132.

Simondon, Gilbert. (1995). *L'Individu et sa genèse physico-biologique*. Grenoble: Millon.

Sinnott-Armstrong, Walter, Adina Roskies, Teneille Brown, and Emily Murphy. (2008). "Brain Images as Legal Evidence." *Episteme: A Journal of Social Philosophy* 5(3): 359–373.

Skagestad, Peter. (1993). "Thinking with Machines: Intelligence Augmentation, Evolutionary Epistemology, and Semiotic." *The Journal of Social and Evolutionary Systems* 16(2): 157–180.

Skagestad, Peter. (1999). "Peirce's Inkstand as an External Embodiment of Mind." *Transactions of the Charles S. Peirce Society* 35(3): 551–561.

Slaby, Jan. (2010). "Steps towards a Critical Neuroscience." *Phenomenology and the Cognitive Sciences* 9(3): 97–416.

Slatman, Jenny. (2007). "Grenzen aan het Vreemde." *Wijsgerig Perspectief* 47: 6–16.

Slors, Marc. (2012). *Dat Had Je Gedacht! Brein, Bewustzijn En Vrije Wil In Filosofisch Perspectief*. Amsterdam: Boom.

Slors, Marc. (2015). "Conscious Intending as Self-Programming." *Philosophical Psychology* 28: 94–113.

Sloterdijk, Peter. (1998). *Sphären I: Blasen*. Frankfurt: Suhrkamp.

Smart, Paul R. (2017). "Situating Machine Intelligence Within the Cognitive Ecology of the Internet." *Minds and Machines* 27(2): 357–380.

Smart, Paul, Richard Heersmink and Robert Clowes. (2017). "The Cognitive Ecology of the The Internet." In S.J. Cowley and F. Vallée-Tourangeau (eds.). *Cognition Beyond the Brain*, 2nd ed. Dordrecht: Springer, 251–282.

Smith, Angela. (2005). "Responsibility for Attitudes: Activity and Passivity in Mental Life." *Ethics* 115(2): 236–271.

Smith, Angela. (2008). "Control, Responsibility, and Moral Assessment." *Philosophical Studies* 138(3): 367–392.

Snead, Carter O. (2007) "Neuroimaging, Entrapment, and the Predisposition to Crime." *American Journal of Bioethics* 7: 60–61

Snowdon, Paul F. (1990). "Persons, Animals, and Ourselves." In Ch. Gill (ed.). *The Person and the Human Mind*. Oxford: Oxford University Press, 83–107.

Snowdon, Paul F. (2014). *Persons, Animals, Ourselves*. Oxford: Oxford University Press.

Soon, Chun Siong, Marcel Brass, Hans-Jochen Heinze, and John-Dylan Haynes. (2008). "Unconscious Determinants of Free Decisions in the Human Brain." *Nature Neuroscience* 11(5): 543–545.

Sorgner, Stefan. (2009). "Nietzsche, the Overhuman, and Transhumanism." *Journal of Evolution and Technology* 20. http://jetpress.org/v20/sorgner.htm

Sparrow, Robert. (2007). "Revolutionary and Familiar, Inevitable and Precarious: Rhetorical Contradictions in Enthusiasm for Nanotechnology." *NanoEthics* 1(1): 57–68.

Sponselee, Anna-mie A.G., Ben Schouten, D.G. Bouwhuis, and Charles G. Willems. (2008). "Smart Home Technology for the Elderly: Perceptions of Multidisciplinary Stakeholders." In M. Mühlhäuser, A. Ferscha, and E. Aitenbichler (eds.). *Constructing Ambient Intelligence*. Berlin: Springer, 314–326.

Sripada, Chandra. (2016). "Self-expression: A Deep Self Theory of Moral Responsibility." *Philosophical Studies* 173: 1203–1232.

Stack, George J. (1983). *Lange and Nietzsche*. Berlin: Walter de Gruyter.

Stanovich, Keith E. (2009). *What Intelligence Tests Miss: The Psychology of Rational Thought*. New Haven, CT: Yale University Press.Stiegler, Bernard. (1998). *Technics and Time, 1. The Fault of Epimetheus*. Stanford, CA: Stanford University Press.

Stegmaier, Werner. (1977). *Substanz. Grundbegriff der Metaphysik*. Stuttgart-Bad Cannstatt: Frommann-Holzboog.

Stegmaier, Werner. (1992). *Philosophie der Fluktuanz. Dilthey und Nietzsche*. Göttingen: Vandenhoeck and Ruprecht.

Stein, Ludwig. (1908). "Die Neupositivistische Bewegung (Der 'Pragmatismus' von William James)." In L. Stein (ed.). *Philosophische Strömungen der Gegenwart*. Stuttgart: Ferdinand Enke, 33–75.

Steilberg, Hays Alan. (1996). *Die amerikanische Nietzsche-Rezeption von 1896–1950*. Berlin/New York: Walter de Gruyter.

Sterelny, Kim. (2010). "Minds: Extended or Scaffolded?" *Phenomenology and the Cognitive Sciences* 9(4): 465–481.

Sternberg, Robert J. and Karin Sternberg. (2012). *Cognitive Psychology*, 6th ed. Belmont, CA: Wadsworth.

Stillman, Tyler F., Roy Baumeister, and Alfred Mele. (2011). "Free Will in Everyday Life: Autobiographical accounts of Free and Unfree Actions." *Philosophical Psychology* 24: 381–394.

Stock, Gregory. (2002). *Redesigning Humans: Our Inevitable Genetic Future*. Boston, MA: Houghton Miffin Company.

Stocker, Michael. (1979). "Desiring the Bad." *Journal of Philosophy* 76: 738–753.

Strakowski, Stephen, James C. Eliassen, Martine Lamy, Michael A. Cerullo, Jane B. Allendorfer, Michelle Madore, Jing-Huei Lee, Jeffrey A. Welge, Melissa P. DelBello, David E. Fleck, and Caleb M. Adler. (2011). "fMRI Brain Activation in Bipolar Mania: Evidence for Disruption of the Ventrolateral Prefrontal-Amygdala Emotional Pathway." *Biological Psychiatry* 69(4): 381–388.

Sulzer, James Samuel, Ranganatha Sitaram, Maria Laura Blefari, SPyros Kollias, Niels Birbaumer, Klaas Enno Stephan, Andreas Luft, and Roger Gassert. (2013). "Neurofeedback-Mediated Self-Regulation of the Dopaminergic Midbrain." *Neuroimage* 83: 176–184.

Sutton, John. (2008). "Material Agency, Skills and History: Distributed Cognition and the Archaeology of Memory." In C. Knappett, L. Malafouris (eds.). *Material Agency: Towards a Non-Anthropocentric Approach*. New York: Springer, 37–55.

Sutton, John. (2009). "'The Feel of the World': Exograms, Habits, and the Confusion of Types of Memory." In K. Kania (ed.). *Philosophers on Memento*. London: Routledge, 65–86.

Sutton, John. (2010). "Exograms and Interdisciplinarity: History, the Extended Mind, and the Civilizing Process." In R. Menary (ed.). *The Extended Mind*. Cambridge: The MIT Press, 189–226.

Sutton, John, Celia B. Harris, Paul G. Keil and Amanda J. Barnier. (2010). "The Psychology of Memory, Extended Cognition, and Socially Distributed Remembering." *Phenomenology and the Cognitive Sciences* 9(4): 521–560.

Svavarsdóttir, Sigrun. (1999). "Moral Cognitivism and Motivation." *Philosophical Review* 108: 161–219.

Swaab, Dick. (2010). *Wij zijn ons Brein. Van Baarmoeder tot Alzheimer*. Amsterdam: Uitgeverij Contact.

Swinburne, Richard. (1984). "Personal Identity: The Dualist Theory." In S. Shoemaker and R. Swinburne (eds.). *Personal Identity*. Oxford: Blackwell, 1–66.

Taylor, Charles. (1976). "Responsibility for Self." In A.O. Rorty (ed.). *The Identities of Persons*. Berkeley: University of California Press, 281–299.

Taylor, Charles. (1989). *Sources of the Self: The Making of Modern Identity*. Cambridge: Harvard University Press.

Taylor, Charles. (1992). *The Ethics of Authenticity*. Cambridge: Harvard University Press.

Tenner, Edward. (2003). *Our Own Devices: The Past and Future of Body Technology*. New York: Knopf.

Thibault, Robert T., Amanda MacPherson, Michael Lifshitz, Raquel R. Roth, and Amir Raz. (2018). "Neurofeedback with fMRI: A Critical Systematic Review." *Neuroimage* 172: 786–807.

Thurston, Luke. (2003). "Meaning on Trial: Sublimation and the Reader." In P. Adams (ed.). *Art: Sublimation or Symptom*. New York: Other Press, 29–48.

Trevena Judy, and Miller Jeff (2010). Brain Preparation Before a Voluntary Action: Evidence Against Unconscious Movement Initiation. *Consciousness and Cognition* 19: 447–456.

Turkle, Sherry. (2004). "Whither Psychoanalysis in Computer Culture?" *Psychoanalytic Psychology* 21: 16–30.

Turner, Danielle, Trevor Robbins, Luke Clark, Adam R. Aron, Jonathan Dowson, and Barbara J. Sahakian. (2003). "Cognitive Enhancing Effects of Modafinil in Healthy Volunteers." *Psychopharmacology* 165: 260–269.

Ulmer, Karl. (1953). *Wahrheit, Kunst und Natur bei Aristoteles. Ein Beitrag zur Aufklärung der metaphysischen Herkunft der modernen Technik*. Tübingen: Walter De Gruyter.

Unger, Peter. (1979) "Why There Are No People." *Midwest Studies in Philosophy* 4(1): 177–222.

Urban, Kimberly R. and Wen-Jun Gao. (2014). "Performance Enhancement at the Cost of Potential Brain Plasticity: Neural Ramifications of Nootropic Drugs in the Healthy Developing Brain." *Frontiers in Systems Neuroscience* 8(38). https://www.frontiersin.org/articles/10.3389/fnsys.2014.00038/full

Van den Eede, Yoni. (2011). "In between Us: On the Transparency and Opacity of Technological Mediation." *Foundations of Science* 16(2): 139–159.

Van der Veer, Rene. (1987). "The Relation between Vygotsky and Mead Reconsidered. A Comment on Glock." *Studies in Soviet Thought* 34: 91–93.

Van Est, Rinie, Virgil Rerimassie, Ira van Keulen, and Gaston Dorren. (2014). *Intimate Technology: The Battle for Our Body and Behavior*. The Hague: Rathenau Institute.

Van Inwagen, Peter. (1990). *Material Beings*. Ithaca: Cornell University Press.

Van Tongeren, Paul. (1989). *Die Moral von Nietzsches Moralkritik*. Bonn: Bouvier Verlag.

Varela, Francisco. (1996). "Neurophenomenology: A Methodological Remedy for the Hard Problem." *Journal for Consciousness Studies* 3(4): 330–349.

Varela, Francisco, Evan Thompson, and Eleanor Rosch. (1991). *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge: MIT Press.

Velleman, David. (1992). "What Happens When Someone Acts?" *Mind* 101(403): 461–481.

Vennix, Ad C.M. (1996). *Metafysica*. Syllabus Metafysica. Nijmegen.

Vennix, Ad C.M. (1999). "De Cartesiaanse Filosofie in Vogelvlucht." *Filosofie* 9(1): 4–20.

Verbeek, Peter-Paul. (2005). *What Things Do: Philosophical Reflections on Technology, Agency, and Design*. University Park: Pennsylvania State University Press.

Verbeek, Peter-Paul. (2008). "Cyborg Intentionality: Rethinking the Phenomenology of Human-Technology Relations." *Phenomenology and the Cognitive Sciences* 7(3): 387–395.

Verbeek, Peter-Paul. (2011). *Moralizing Technology: Understanding and Designing the Morality of Things*. Chicago: University of Chicago Press.

Verbeek, Peter-Paul. (2014). "Designing the Public Sphere: Information Technologies and the Politics of Mediation." In L. Floridi (ed.). *The Onlife Manifesto: Being Human in a Hyperconnected Era*. Dordrecht: Springer, 217–227.

Vidal, Fernando and Francisco Ortega. (2017). *Being Brains: Making the Cerebral Subject*. New York: Fordham University Press.

Vincent, Nicole A. (ed.). (2013). *Neuroscience and Legal Responsibility*. Oxford: Oxford University Press.

Visker, Rudi. (2005). "The Strange(r) Within Me." *Ethical Perspectives* 12(4): 425–441.

Vogeley, Kai and Shaun Gallagher. (2011). "Self in the Brain." In Shaun Gallagher (ed.). *The Oxford Handbook of the Self*. Oxford: Oxford University Press, 111–138.

Volkow, Nora D., Gene-Jack Wang and Ruben D. Baler. (2011). "Reward, Dopamine and the Control of Food Intake: Implications for Obesity." *Trends in Cognitive Science* 15(1): 37–46.

Waibel, Edwin P.B. (1915). *Der Pragmatismus in der Geschichte der Philosophie*. Bonn: Ludwig.

Wang, Shensheng, Scott O. Lilienfeld, and Philippe Rochat. (2015). "The Uncanny Valley: Existence and Explanations." *Review of General Psychology* 19: 393–407.

Warbeke, John M. (1909). "Friedrich Nietzsche, Antichrist, Superman and Pragmatist." *Harvard Theological Review* 2: 366–385.

Warwick, Kevin. (2000). "Cyborg 1.0." Wired. http://archive.wired.com/wired/archive/8.02/warwick.html

Watson, Gary. (1975). "Free Agency." *The Journal of Philosophy* 72: 205–220.

Watts, Christopher M. (2008). "On Mediation and Material Agency in the Peircean Semeiotic." In C. Knappett and L. Malafouris (eds.). *Material Agency: Towards a Non-Anthropocentric Approach*. New York: Springer, 187–207.

Wegner, Daniel M. and Thalia Parker Wheatley. (1999). "Apparent Mental Causation: Sources of the Experience of Will." *American Psychologist* 54: 480–492.

Weingarten, Harvey P. and Dawn Elston. (1990). "The Phenomenology of Food Cravings." *Appetite* 15(3): 231–246.

Weisberg, Deena Skolnick, Frank C. Keil, Joshua Goodstein, Elizabeth Rawson and Jeremy R. Gray. (2008). "The Seductive Allure of Neuroscience Explanations." *Journal of Cognitive Neuroscience* 20: 470–477. doi: 10.1162/jocn.2008.20040.

Weiser, Mark. (1991). "The Computer for the Twenty-First Century." *Scientific American* 265(3): 94–104.

Wheeler, Michael. (2004). "Is Language the Ultimate Artefact?" *Language Sciences* 26: 693–715.

Whitehead, Alfred North. (1925). *Science and the Modern World*. New York: The Macmillan Company.

Whitehead, Alfred North. (1929/1978). *Process and Reality. An Essay in Cosmology*. New York: The Free Press.

Wiggins, David. (1980). *Sameness and Substance*. Oxford: Blackwell.

Wiggins, David. (2001). *Sameness and Substance Renewed*. Cambridge: Cambridge University Press.

Wilcox, John T. (1980). "A Note on Correspondence and Pragmatism in Nietzsche." *International Studies in Philosophy* 12: 77–80.

Wilcox, John T. (1986). "Nietzsche Scholarship and 'The Correspondence Theory of Truth': The Danto Case." *Nietzsche Studien* 15: 337–357.

Wildman, Wesley J. (2010). "An Introduction to Relational Ontology." In J. Polkinghorne and J. Zizioulas (eds.). *The Trinity and an Entangled World: Relationality in Physical Science and Theology*. Grand Rapids: Eerdmans, 55–73.

Wiley, Norbert. (2006). "Pragmatism and the Dialogical Self." *International Journal for Dialogical Science* 1: 5–21.

Williams, Lawrence E., and John A. Bargh. (2008). "Experiencing Physical Warmth Promotes Interpersonal Warmth." *Science* 322: 606–607.

Wilsdon, James and Paul Miller. (eds.) (2006). *Better Humans: The Politics of Human Enhancement and Life Extension*. London: Demos.

Withy, Katherine. (2015). *Heidegger on Being Uncanny*. Cambridge: Harvard University Press.

Wittgenstein, Ludwig. (1922). *Tractatus Logico-Philosophicus*. London: Kegan Paul.

Wittgenstein, Ludwig. (1958). *The Blue and Brown Books*. Oxford: Blackwell.

Wittgenstein, Ludwig. (2001). *Philosophical Investigations*. Oxford: Blackwell.

Wolf, Susan M. (2008). "Neurolaw: The Big Question." *The American Journal of Bioethics* 8(1): 21–22.

Wolf, Susan. (1990). *Freedom Within Reason*. Oxford: Oxford University Press.

Wolf, Susan. (2003). "Sanity and the Metaphysics of Responsibility." In G. Watson (ed.). *Free Will*. Oxford: Oxford University Press, 372–387.

Wong, James. (2013). "Self and Others: The Work of 'Care' in Foucault's Care of the Self." *Philosophy Faculty Publications* 6. https://scholars.wlu.ca/phil_faculty/6

Yamada, Yuki, Takahiro Kawabe, and Keiko Ihaya. (2013). "Categorization Difficulty is Associated with Negative Evaluation in the 'Uncanny Valley' Phenomenon." *Japanese Psychological Research* 55: 20–32.

Yesavage, Jerome A., Marco S. Mumenthaler, July L. Taylor, Lisa A. Friedman, Ruth O'Hara, Javaid Sheikh, Jared R. Tinklenberg and Peter J. Whitehouse. (2002). "Donepezil and Flight Simulator Performance: Effects on Retention of Complex Skills." *Neurology* 59(1): 123–125.

Young, Kymberly, Vadim Zotev, Raquel Phillips, Maaya Misaki, Han Yuan, Wayce C. Drevets, and Jerzy Bodurka. (2014). "Real-time fMRI Neurofeedback Training of Amygdala Activity in Patients with Major Depressive Disorder." *PLoS One* 9(2): e88785.

Zahavi, Dan. (2005). "Being Someone." *Journal Psyche* 11(5). http://journal-psyche.org/files/0xaad8.pdf

Zahavi, Dan. (2008). "Internalism, Externalism, and Transcendental Idealism." *Synthese* 160: 355–374.

Zahavi, Dan. (2015). "Self and Other: From Pure Ego to Co-constituted We." *Continental Philosophy Review* 48(2): 143–160.

Zangwill, Nick. (2003). "Externalist Moral Motivation." *American Philosophical Quarterly* 40: 143–54.

Zizek, Slavoj. (1989). *The Sublime Object of Ideology*. London: Verso.

Zizek, Slavoj. (1996). "From Desire to Drive: Why Lacan is not Lacanian." http://theoryleaks.org/text/articles/slavoj-zizek/why-lacan-is-not-lacanian/

Zizek, Slavoj. (2009). *The Parallax View*. Cambridge: The MIT Press

Zizek, Slavoj. (2010). *Living in the End Times*. London: Verso.

Zizek, Slavoj. (2012). *Organs Without Bodies: On Deleuze and Consequences*. London: Routledge.

Zwart, Hub. (2015). "Extimate Technologies: Empowerment, Intrusiveness, Surveil- lance: The Fate of the Human Subject in the Age of Intimate Technologies and Big Data." In *Emerging Technologies and Human Rights* (proceedings). Strasbourg: Council of Europe, 40–45.

Zwart, Hub. (2017). "'Extimate' Technologies and Techno-Cultural Discontent: A Lacanian Analysis of Pervasive Gadgets." *Techné: Research in Philosophy and Technology* 21(1): 24–54.

Index

Note: **Bold** page numbers refer to tables, *italics* page numbers refer to figures and page numbers followed by "n" refer to notes.