**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2020

Neuromodulation of the “Moral Brain” – Evaluating Bridges Between Neural Foundations of Moral Capacities and Normative Aims of the Intervention

Ineichen, Christian ; Christen, Markus

Abstract: The question of whether neuroscience has normative implications or not becomes practically relevant when neuromodulation technologies are used with the aim of pursuing normative goals. The historical burden of such an endeavor is grave and the current knowledge of the neural foundations of moral capacities is surely insufficient for tailored interventions. Nevertheless, invasive and non-invasive neuromodulation techniques are increasingly used to address complex health disturbances and are even discussed for enhancement purposes, whereas both aims entail normative objectives. Taking this observation as an initial position, our contribution will pursue three aims. First, we summarize the potential of neuromodulation techniques for intervening into the “moral brain” using deep brain stimulation as a paradigmatic case and show how neurointerventions are changing our concepts of agency and personality by providing a clearer picture on how humans function. Second, we sketch the “standard model” explanations with respect to ethically justifying such interventions, which rely on a clear separation between normative considerations (“setting the goals of the intervention” or “the desired condition”) and empirical assessments (“evaluating the outcome of the intervention” or “the actual condition”). We then analyze several arguments that challenge this “standard model” and provide bridges between the empirical and normative perspective. We close with the observation that maintaining an analytical distinction between the normative and empirical perspective is reasonable, but that the practical handling of neuromodulation techniques that involve normative intervention goals is likely to push such theoretical distinctions to their limits.

DOI: https://doi.org/10.1007/978-3-030-56134-5_9

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-194178>

Book Section

Accepted Version

Originally published at:

Ineichen, Christian; Christen, Markus (2020). Neuromodulation of the “Moral Brain” – Evaluating Bridges Between Neural Foundations of Moral Capacities and Normative Aims of the Intervention. In: Holtzman, G S; Hildt, E. Does Neuroscience Have Normative Implications? Cham: Springer, 165-185.

DOI: https://doi.org/10.1007/978-3-030-56134-5_9

Metadata of the chapter that will be visualized online

Chapter Title	Neuromodulation of the “Moral Brain” – Evaluating Bridges Between Neural Foundations of Moral Capacities and Normative Aims of the Intervention	
Copyright Year	2020	
Copyright Holder	Springer Nature Switzerland AG	
Author	Family Name	Ineichen
	Particle	
	Given Name	Christian
	Suffix	
	Division	Preclinical Laboratory for Translational Research into Affective Disorders (PLaTRAD), Department of Psychiatry, Psychotherapy and Psychosomatics & Department of Neurology
	Organization/University	Psychiatric & University Hospital Zurich
	Address	Zürich, Switzerland
	Email	christian.ineichen@uzh.ch
Corresponding Author	Family Name	Christen
	Particle	
	Given Name	Markus
	Suffix	
	Division	
	Organization/University	University of Zurich (UZH)
	Address	Zürich, Switzerland
	Email	christen@ifi.uzh.ch
	Email	christen@ethik.uzh.ch

Abstract	<p>The question of whether neuroscience has normative implications or not becomes practically relevant when neuromodulation technologies are used with the aim of pursuing normative goals. The historical burden of such an endeavor is grave and the current knowledge of the neural foundations of moral capacities is surely insufficient for tailored interventions. Nevertheless, invasive and non-invasive neuromodulation techniques are increasingly used to address complex health disturbances and are even discussed for enhancement purposes, whereas both aims entail normative objectives. Taking this observation as an initial position, our contribution will pursue three aims. First, we summarize the potential of neuromodulation techniques for intervening into the “moral brain” using deep brain stimulation as a paradigmatic case and show how neurointerventions are changing our concepts of agency and personality by providing a clearer picture on how humans function. Second, we sketch the “standard model” explanations with respect to ethically justifying such interventions, which rely on a clear separation between normative considerations (“setting the goals of the intervention” or “the desired condition”) and empirical assessments (“evaluating the outcome of the intervention” or “the actual condition”). We then analyze several arguments that challenge this “standard model” and provide bridges between the empirical and normative perspective. We close with the observation that maintaining an analytical distinction between the normative and empirical perspective is reasonable, but that the practical handling of neuromodulation techniques that involve normative intervention goals is likely to push such theoretical distinctions to their limits.</p>
Keywords (separated by “ - ”)	<p>Neuromodulation - Deep brain stimulation - Is-ought-gap - Agency - Personality - Self-regulation</p>

Chapter 9

Neuromodulation of the “Moral Brain” – Evaluating Bridges Between Neural Foundations of Moral Capacities and Normative Aims of the Intervention

Christian Ineichen and Markus Christen

Abstract The question of whether neuroscience has normative implications or not becomes practically relevant when neuromodulation technologies are used with the aim of pursuing normative goals. The historical burden of such an endeavor is grave and the current knowledge of the neural foundations of moral capacities is surely insufficient for tailored interventions. Nevertheless, invasive and non-invasive neuromodulation techniques are increasingly used to address complex health disturbances and are even discussed for enhancement purposes, whereas both aims entail normative objectives. Taking this observation as an initial position, our contribution will pursue three aims. First, we summarize the potential of neuromodulation techniques for intervening into the “moral brain” using deep brain stimulation as a paradigmatic case and show how neurointerventions are changing our concepts of agency and personality by providing a clearer picture on how humans function. Second, we sketch the “standard model” explanations with respect to ethically justifying such interventions, which rely on a clear separation between normative considerations (“setting the goals of the intervention” or “the desired condition”) and empirical assessments (“evaluating the outcome of the intervention” or “the actual condition”). We then analyze several arguments that challenge this “standard model” and provide bridges between the empirical and normative perspective. We close with the observation that maintaining an analytical distinction between the normative and empirical perspective is reasonable, but that the practical handling of neu-

C. Ineichen

Preclinical Laboratory for Translational Research into Affective Disorders (PLaTRAD),
Department of Psychiatry, Psychotherapy and Psychosomatics & Department of Neurology,
Psychiatric & University Hospital Zurich, Zürich, Switzerland
e-mail: christian.ineichen@uzh.ch

M. Christen (✉)

University of Zurich (UZH), Zürich, Switzerland
e-mail: christen@ifi.uzh.ch; christen@ethik.uzh.ch

© Springer Nature Switzerland AG 2020

G. S. Holtzman, E. Hildt (eds.), *Does Neuroscience Have Normative Implications?*, The International Library of Ethics, Law and Technology 22,
https://doi.org/10.1007/978-3-030-56134-5_9

romodulation techniques that involve normative intervention goals is likely to push such theoretical distinctions to their limits.

Keywords Neuromodulation · Deep brain stimulation · Is-ought-gap · Agency · Personality · Self-regulation

9.1 Introduction

The potential impact of the natural sciences on normative inquiries has been the basis of considerable debate. After the collapse of the Thomistic worldview supporting the view that the good, the true and the being are three perspectives of the same reality (e.g. Rfner 1964), the scientific and the moral domain have been carefully separated. It has been a rather recent development that we now again discuss interconnections of these two domains (see e.g. Harris 2011). That neuroscience started to infiltrate “morality” as a research area by investigating the neural foundations of human moral behavior at least since the tragic case of Phineas Gage, a railroad worker whose skull was brutally perforated by an iron rod during an accidental explosion, can be considered one reason for this rapprochement. Gage, who showed sudden asocial behavioral manifestations secondary to the accidental explosion, set the stage for scientifically inquiring into the “moral” brain. And although a closer look at the story revealed that many of the accounts of Gage’s life after 1848 are strange mixtures of slight fact, considerable fancy and downright fabrication (Macmillan 2000), Gage symbolized a paradigmatic description of how brain and moral behavior are related – namely in the sense that the dysfunction of some parts of the brain (the right orbitofrontal or ventromedial prefrontal cortex) can lead to major aberrations in moral behavior.

Interestingly, this “deterministic connection” of lesions in certain brain regions with amoral behavior has already been questioned in the first “review paper” (using today’s terminology) on the effect of frontal lesions on behavior by Leonore Welt, published in 1888 (2009). Welt presented 58 cases (of various forms) of frontal lesion patients (including the Phineas Gage case), of whom 47 did not show character changes (“Charaktervernderungen”) after the lesions (Christen and Regard 2012). Certainly, degree and localization of these injuries were much harder to describe when neuroimaging was not yet available. Nevertheless, Welt urged caution for those seeking to deterministically associate brain lesions with character changes. Irrespective of this, the case of Phineas Gage is frequently cited in the introduction of papers published today that discuss the relationship between brain and morality.

This emerging knowledge about the connection between brain and (moral) behavior also triggered the tragic period between 1935 and 1950, characterized by the dramatic rise and subsequent crashing fall of psychosurgery to treat mental

illness. The neurophysiologist and co-recipient of the Nobel Prize in Medicine in 1949, Egaz Moniz, first proposed in 1935 to treat a patient’s pathological anxiety states and paranoid ideas by disrupting the connections between the frontal lobes and the rest of the brain, a procedure called “prefrontal leucotomy” (Valenstein 1986)—a term that was later changed to lobotomy by Walter Freeman. With only anecdotal information to support his theory, no evidence from preclinical experimentation, no tests of safety or effectiveness and without any follow-up of patients beyond the immediate postoperative period, Moniz saw his intervention quickly adopted by several other medical practitioners. In particular, Walter Freeman almost single-handedly popularized the widespread use of psychosurgery throughout the U.S. between his first prefrontal lobotomy performed on the 14th September in 1936 and his last lobotomy in 1967 by performing about 3400 transorbital lobotomies (Valenstein 1986; El-Hai 2005). Despite critics describing lobotomized patients to be confused, lacking in affect and without normal response to common human interactions, there was a remarkable lack of critical evaluation by rapidly declaring many of the patients treated as completely cured. The tragic rise and belated fall of psychosurgery in the 1970s elucidates how science attempted to pursue normative goals by modulating people’s brain, aiming at “recalibrating” their behavior and affect so as to conform to certain societal norms.

Meanwhile, the more recent advent of trying to empirically grab ahold of the foundations of human moral behavior has snowballed into numerous inquiries that are linked to the promise of providing solid scientific answers not only to questions of moral behavior, but also to questions of morality as such. Contributing to this emerging aura of authority has been the close collaboration of researchers from psychology, philosophy and neuroscience, with the aim of illuminating the distinct neuronal pathways that underlie our moral intuitions, motivations, judgments and behaviors. For example, Joshua Greene, a philosopher by training, joined forces with neuroscientist Jonathan Cohen and analyzed functional magnetic resonance images of people’s brains as they responded to hypothetical moral dilemmas (Greene et al. 2001). Another example includes social psychologist Jonathan Haidt, who, based on a set of neuropsychological experiments, questioned the role reasoning plays in moral judgments (Haidt 2007). In his view, reason rather serves only a secondary role as a post-hoc means of justifying unreflective and quick intuitions. Moreover, many philosophers (beginning with Foot (1967) and Thomson (1985)) and experimental psychologists (starting with Greene et al. 2001) have been especially taken with a series of moral conundrums referred to as “trolleyology” – a provoking series of thought experiments circling around the question of whether one should (intentionally or accidentally) sacrifice a human being to prevent that a runaway trolley kills several human beings (Cathcart 2013). For this empirical inquiry, numerous techniques, including neuroimaging, have been endeavored.

The story above, about neuroscientific investigations and their ability to provide stringent answers to questions of morality, would be one-sided if one would not mention that this scientific advance ignited a countermovement, typified by e.g. philosopher Selim Berker. Berker has pointed out in his paper “the normative

insignificance of neuroscience” the claim that neural and psychological processes that track personal factors cannot be relied on to support moral propositions or guide moral decisions (Berker 2009). Consistently, others have convincingly claimed that an understanding of the neural correlates of reasoning cannot tell anything significant about whether the outcome of this reasoning is justified. The discussion above already touches on unjustifiably bridging the is-ought gap, that is, suggesting that knowledge gained through descriptive natural science can itself determine what is morally right and wrong.

The previously sketched focus of neuroscientific studies investigating the neural foundations of morality has led to a first and incomplete mapping of the “moral brain”. Meanwhile, a large body of research has tried to identify and probe “moral circuits” that underlie human processing both in healthy and pathological states. Consequently, various brain regions have been implicated in emotional processing and social cognition, including theory of mind. Those include the ventromedial prefrontal cortex (including the medial orbitofrontal cortex and Brodmann area 10 and 11), the orbitofrontal cortex, but also the amygdala, superior temporal sulcus, bilateral temporoparietal junction, posterior cingulate cortex, and precuneus (Young and Dungan 2012; Fumagalli and Priori 2012; Dolan 1999; Casebeer 2003). However, and even though many brain areas have successfully been associated with moral motivation, judgment and behavior, it is important to highlight that insight on the interconnection between moral abilities and brain regions is still rather scarce and unspecific, as the associated networks and brain regions are involved in a broad range of human behaviors.

What is more relevant in the context of the neuroscience of morality is that advances in contemporary neuroscience have made it possible to probe and modulate brain functions with an unprecedented level of precision and with hypotheses gleaned from preclinical (using laboratory animals) and clinical research. Some of the most recent neuromodulation technologies have made it possible to probe causal relationships between the different dysfunctional socio-moral pathways underlying many complex (e.g. neuropsychiatric) pathologies. Examples of such technologies include deep brain stimulation (DBS), transcranial magnetic stimulation (TMS), transcranial direct current stimulation (tDCS), and optogenetics, to name a few. All these technologies manipulate neural networks in a rather specific, evidence-driven way. While the level of precision with which neural elements can be modulated varies between the different approaches, they all surely transcend the intervention possibilities commonly employed by oral medication. Network science, as an approach to model and analyze the dynamics of communication on networks, has proven useful for the prediction of emergent network states. Hence, such approaches can offer insight into the mechanisms by which brain networks transform and process information (Avena-Koenigsberger et al. 2018). The aim of many researchers of injecting test signals into an electrical neural circuit for subsequent recording of the output at distinct points also constitutes an engineering ideal. This portrays nicely the image of a more technical notion of the human brain.

The question of whether neuroscience has normative implications or not becomes relevant when neuromodulation technologies are used with the aim of pursuing

normative goals. As outlined previously in the discussion of psychosurgery, the historical burden of such an endeavor is grave and the current knowledge of the neural foundations of moral capacities is surely insufficient for tailored interventions. Nevertheless, invasive and non-invasive neuromodulation techniques are increasingly used to address complex health disturbances and are even discussed for enhancement purposes (Earp et al. 2017), and both of these aims entail normative goals. In what follows, we describe DBS, a neuromodulation technology promising much hope in the treatment of numerous complex health disturbances, and its potential of intervening into the “moral brain”. As will be seen, research on DBS has provided insights into the neurophysiology of the basal ganglia and how a specific area of the brain, namely the subthalamic nucleus (STN), is involved in various computations necessary to establish agentic states. Apart from agency, DBS-research has been associated with complex behavioral alterations that encompass personality-constituting processes. After describing complex psychological alterations following DBS interventions, we provide a theoretic link that tries to bring together notions of agency and self-regulation on the ground of the neurophysiology of the basal ganglia. We then outline how both notions have relevance in the moral sphere. Finally, we provide normative aspects of neuromoral interventions that will focus on the presumed separability of the normative and empirical perspective that have practical relevance for medical decision-making processes.

9.2 Neuromodulating the “Socio-moral Brain” 174

In what will follow, we shall turn towards deep brain stimulation as an exemplary case for intervening into complex health disturbances. We will extend this notion by alluding to complex disorders that inevitably involve complex psychological aberrations with social and moral relevance.

9.2.1 *DBS as an Exemplary Case of Intervening Into the “Moral Brain”* 179

It was 1971 when psychologist B. F. Skinner, the father of behaviorism, expressed his hope that some of the most abominable of the humanly created problems that derange our lives, such as wars and famines, could all be solved by new “technologies of behavior.” Skinner wanted to “maximize the achievements of which the human organism is capable” – an aim that already subsumes normative goals pertaining to social engineering (Skinner et al. 1972). Albeit Skinner understood these “technologies of behavior” as manipulation of the external environment at that time, recent advances have emerged to the point where manipulations of the internal environment of the brain are possible. One such possibility is exemplified by the

electrical modulation of brain circuits through DBS, a symptomatic neurosurgical intervention (i.e. leading to substantial symptom relief in well-selected patients) that includes electrode implantation to apply electrical currents to target structures.

It was 1976 when Jean Siegfried started DBS in pain patients and 1986, at the meeting of the American Society for Stereotactic and Functional Neurosurgery, when Alim-Louis Benabid presented his abstract on stimulation of the ventral intermediate nucleus of the thalamus (Vim) for Parkinson's disease (PD). Both events marked the beginning of modern deep brain stimulation (DBS) (Christen and Müller 2012; Hariz 2012). From this moment on, intense research activities marked an era of probing new anatomical targets, resulting in a subsequent increase in the therapeutic spectrum. The early success of approved and exempted indications has led to the use of DBS to treat a non-undisputed variety of other emerging indications, including e.g. pain, obsessive compulsive disorder, major depression, Tourette syndrome, obesity, anorexia nervosa, substance addiction, epilepsy, pathological aggression, dementia, apart from other movement disorders (Youngerman et al. 2016). In particular, the early success and the subsequent research focus on treating PD through DBS led to multiple randomized clinical trials that have shown DBS to be more effective than the best medical treatment in well-selected patients (Deuschl et al. 2006; Follett et al. 2010; Weaver et al. 2009). The recognition that DBS was better at improving the motor symptoms of PD compared to best medical treatment was final evidence for DBS to be taken as the gold standard therapeutic intervention for advanced PD that can no longer be successfully treated with L-Dopa.

Meanwhile, the ongoing expansion of DBS indications from movement disorders such as dystonia, essential tremor and PD to targeting mood and mind (in e.g. depression and Alzheimer's disease) marked a time that was characterized not only by intense research activities but also by serendipitous discoveries. For example, a failed attempt to treat obesity in a patient showing vivid memory recollection during stimulation (Hamani et al. 2008) expeditiously resulted in a first clinical trial for Alzheimer's disease (Laxton et al. 2010). As a result of intended and serendipitous discoveries, DBS is now used for treating various indications reaching from cognition to volition to behavior and affect. Owing to improvements in navigation and imaging techniques, researchers were further able to pursue hypothesis driven empirical research. In parallel, often observed side-effects built the basis for a broadening of the therapeutic spectrum. Some of the observed side effects encompassed complex affective and behavioral changes relating to personality that are more difficult to assess and to describe. Unsurprisingly, the observation and description of these complex alterations has taken longer compared to more focally described changes such as paresthesias (abnormal sensation such as tingling, numbness) or disconjugate gaze (unpaired movements of the eyes).

Currently, DBS is still most frequently used for treating patients suffering from PD. PD represents a complex, chronic, neurodegenerative, multi-systemic disorder that not only affects the dopaminergic system but also serotonergic, glutamatergic and GABAergic neurotransmission (Tremblay et al. 2015), leading to motor and non-motor dysfunction (Schapira et al. 2017). Accordingly, PD additively involves non-motor symptoms such as neuropsychiatric disturbances (depression, anxiety,

apathy), sleep-wake dysfunction (e.g. REM sleep behavior disorder and excessive daytime sleepiness), cognitive problems (e.g. dysexecutive syndrome, dementia) and autonomic dysfunctions (e.g. sexual dysfunction, orthostasis) (Olanow et al. 2011).

Mounting evidence suggests that DBS interventions, primarily geared to treating PD, can result in complex behavioral and affective changes such as hypomania, new onset impulse control disorders (ICDs), including hypersexuality, pathological buying, pathological gambling, and addiction to levodopa (Ballanger et al. 2009; Hack et al. 2014; Castrioto et al. 2014), logorrhea, irritability, aggression, mirthful laughter, egocentrism, lying, and acute sadness and crying (summarized in Ineichen et al. 2016; but see also: Jahanshahi et al. 2015; Müller and Christen 2011). Currently, it is difficult to pinpoint whether such complex non-motor side effects emerge from stimulation, ongoing pathological processes of the disorder being treated, pharmacological therapy, or their interrelation. Nevertheless, through sophisticated study designs, such neuromodulation interventions have the potential to provide causal explanations by demonstrating whether modulation of a certain node is necessary and sufficient for generating the hypothesized modification in function. Therefore, just as the description of side effects that paved the way for the exploration of DBS in areas beyond movement disorders, there is a possibility that these DBS-related complex side effects could provide a basis for future interventions that target socio-moral abilities of patients. In 2012 for example, the journal *Brain* published an article titled “Functional and Clinical Neuroanatomy of Morality” in which the authors allude to the functional role of subcortical structures in the context of morality (Fumagalli and Priori 2012). They also suggest that DBS might be used to treat pathological antisocial behaviors.

Evidently, the above-mentioned alterations following DBS interventions pertain to personality-related changes comprising changes in socio-moral information processing. Given that the treatment of patients suffering from different disorders with electrodes implanted in various nuclei of the brain resulted in a more refined understanding of computations underlying various neural processes, the same is likely to take place in terms of understanding neural computations that underlie socio-moral behavior. As knowledge increases, it is conceivable that similar practices could take place with diseases that directly affect moral behavior, and such engaging in such practices could be the first step toward a more general understanding of the interaction between processes in the brain and moral behavior. Although we have to acknowledge that current DBS interventions interfere with complex socio-moral networks (illuminating the present impossibility to deliver current to target in an appropriate and specific way; Ineichen et al. 2018), they also demonstrate the potential to modulate our social and moral infrastructure through invasive brain technologies. Even though e.g. neuroimaging has provided much insight into neural processes, simply building strong correlations through sophisticated modeling and data analysis still does not provide causal mechanistic insights (Poldrack and Farah 2015). Because of their potential to probe brain networks, neuromodulation technologies, such as DBS, are gaining an increased amount of attraction.

279 **9.2.2 Agency, Control and Self-Regulation**

280 Because neuroscience has just started investigating the mechanisms that generate
 281 agency, personality and complex socio-moral computations in general, and moral
 282 behavior specifically, there is not yet precise knowledge of how to intervene into
 283 these capacities. Nevertheless, we can succumb to the idea that it might only be a
 284 matter of time until precise knowledge on this front will surface. Hence, in what
 285 follows let us assume that there will be precise knowledge of the neural infrastruc-
 286 ture of moral capacities available in the near future, and that brain technologies
 287 could in principle be used to provide tailored interventions that aim at altering
 288 human moral capacities. For in fact, making humans more “ethical” by synthesizing
 289 knowledge on the neural basis of moral decision-making or the neural infrastructure
 290 of self-regulation really might soon become a realistic goal. Notwithstanding its
 291 logical content, the assumption of altering moral capacities may appear less eco-
 292 logically invalid if some delineations are made regarding the human capacity to
 293 self-regulate, to exert control and to be an agent of one’s life. Self-regulation more-
 294 over represents a paradigmatic case of a moral capacity itself. As will be seen, DBS
 295 of the often-targeted basal ganglia, e.g. in the context of PD, can play an important
 296 role with respect to such a capacity.

297 **9.2.2.1 Agency & Self-regulation: Concepts and Relation to Morality**

298 In the sense of the term which we use here, “agency” refers to the ability to translate
 299 mental states such as desires, beliefs or intentions into executive plans of action and
 300 encompasses a set of sensorimotor, cognitive, affective and volitional components.
 301 Neurological and neuropsychiatric disorders impose internal constraints on a num-
 302 ber of these components and hamper the ability to act. While PD can partly be
 303 characterized as a hypokinetic movement disorder with patients showing difficulty
 304 to execute movements because dysfunction of the basal ganglia undermines primar-
 305 ily the sensorimotor component of agency, major depression and generalized anxi-
 306 ety, for example, impair the cognitive, affective and volitional components of
 307 agency. Compared to hyperagentic disorders that show an excessive experience of
 308 one’s own causation and control over events, as evidenced in obsessive compulsive
 309 disorder (OCD) and schizophrenia, PD and major depression can be termed “hypo-
 310 agentic” disorders leading to loss of control and an inability to carry out action plans
 311 (Ineichen and Christen 2017; Glannon and Ineichen 2016; Haggard 2017). DBS can
 312 restore some degree of agency in some of these conditions by allowing executive
 313 control. As will be seen, agency can be considered a prerequisite for self-regulation
 314 because any pathophysiological process that undermines the patients’ ability to con-
 315 vert volitions, thoughts and emotions into actions will typically impede some forms
 316 of self-regulation. Insofar as autonomy can include competency and authenticity
 317 (Mele 2001) and as such represents an important personality trait, various

neurological and psychiatric states can impede patients’ abilities to critically reflect, carry out, identify with or endorse certain mental states (Glannon and Ineichen 2016).

The theme of willpower, primarily focusing on cognitive and volitional subsets of agency at that time, features strongly in ancient mythical stories reaching as far back as Homer’s *Odyssey*. Self-control also emerged as a unifying theme during the Victorian era in the nineteenth century when the building of character and the exercise of willpower were considered synonymous with one another. In fact, self-regulation is an important topic in evolutionary anthropology and also in models trying to describe the genealogy of morality. Apart from the notion that the capability of inhibiting certain instincts might be a true *Specificum Humanum*, it is adaptive to possess the ability not only to give room for the stronger instinct (e.g. in cases of competing goals) but to be able to consider prospect of success and the like. Sometimes, it might be better to wait for a while and to follow another, yet less important goal in the meanwhile. In the tradition of Arnold Gehlen, an important figure in philosophical anthropology, this phenomenon was termed “the hiatus of self-control” (Gehlen and Mensch 1950). Being able to inhibit certain goals for the sake of other executive plans—possibly only experienced in a modus of anticipation—implies the development of a complex sense of time including anticipation and foresight. In fact, not only future actions have to be anticipated but also the motivational moods and motives, which help coordinating the respective action. As a result, executive control can be conceptualized as a defining characteristic of morality, as putting aside own desires is required if perceiving that intentions and goals of others might be harmed. Also Wallace (Wallace 1994) articulated that free autonomous agents must possess the powers of “reflective self-control”. The idea, that self-regulation involves the capacity to stand back from and evaluate one’s first-order desires is furthermore widely shared by philosophers of both a Humean and a Kantian cast (e.g. Bratman 2000 or Kennett 2001).

9.2.2.2 Current Understanding of Self-regulation in a Moral Context

Similarly, today the capacity of the human mind to alter its own responses is regarded as an important foundation for culture, achievement, individual success and morality. The ability to suppress or override competing responses and to alter one’s responses so as to bring them into line with individual goals, ideals, moral values, social norms, laws, and other standards is an important personality process (Kahneman and Chajczyk 1983; Jonides and Nee 2006; Tangney et al. 2004; Gross 1998). It is therefore a vital foundation for the building of character and a core feature of human agency (Baumeister et al. 2006; Bandura et al. 2003). Self-regulation strongly connects to morality insofar as it refers to the “capacity to exercise moral management of the temptations and provocations the individual encounters in a setting” (Wikström 2005, p. 217). Further, people with stronger self-regulation capabilities are expected to be less likely to rely on conscious active self-control because the more energy conserving implicit self-control mechanisms are already robustly installed and therefore are more likely to behave morally.

9.2.2.3 Basal Ganglia and the Neurophysiology of Self-regulation

Meanwhile, a number of brain areas have been implicated in cognitive control, conflict monitoring, response selection and -inhibition and self-regulation in general – among others, the prefrontal cortex (PFC) (Petrides and Pandya 1999), the basal ganglia (Nambu 2008) and the anterior cingulate cortex (ACC) (e.g. Somerville et al. 2006). Low levels of self-control have also been associated with a less refined connectivity (e.g. less myelination and orientation regularity) in frontostriatal (Liston et al. 2006; Casey et al. 2007) and frontoparietal circuitry (Jonides et al. 2000), critical for effective cognitive control. In the classic basal ganglia model, the output nuclei (globus pallidus internus and substantia nigra pars reticulata) hold the cortex and superior colliculus under tonic inhibition to prevent inappropriate (i.e. maladaptive) movements and can phasically release this inhibitory control to allow movements if medium spiny neurons are in the physiological upstate and enough dopamine is present (Jahanshahi et al. 2015; Da Cunha et al. 2015). Executive control is important to override habitual or prepotent responses, to control emotions, focus attention and exert self-regulation necessary for social interaction. Mounting evidence suggests that the basal ganglia coordinate behavioral output through neuronal inhibition and disinhibition, thereby enabling flexible interactions and executive control relevant in cognitive and emotional processing. Therefore, adaptive behavior owes as much to taking appropriate action as to inhibiting or suppressing contextually inappropriate or socially unacceptable behavior (Jahanshahi et al. 2015). Consistently, failure of context-appropriate inhibitory control through e.g. frontostriatal dysfunction leads to many impulsive manifestations relevant in neuropsychiatric disorders. The STN of the basal ganglia, the physiological function of which is to raise the response threshold temporarily and to provide time for information accumulation before decision making and responding, has important connections to the ventral tegmental area and ventral striatum, implicated in drug craving, emotion regulation and impulsivity (Kober et al. 2010; Frank et al. 2007). In fact, the STN is nowadays generally considered to be not only a regulator of motor function, but also an important gateway for cognitive and emotional signal processing. A recent review including tracing, cytoarchitectonic, imaging and electrophysiologic studies furthermore points toward a topographic segregation (motor, associative and limbic) within the STN (Lambert et al. 2015; Temel et al. 2005). It is therefore possible that interference with the STN by disease (PD) or intervention (drugs, DBS), can modulate associative and limbic processing (Tremblay et al. 2015; Kopell and Greenberg 2008; Sesack and Grace 2010). As has been described, STN-DBS, albeit primarily intended to treat motor symptoms of PD, also modulates associative and limbic basal ganglia networks with behavioral and psychological consequences. Complex alterations including personality changes postoperatively may in fact point to aberrant self-regulation that might have been overshadowed by a generalized lack of agency. Studies corroborate this hypothesis insofar as they describe e.g. new onset ICDs following DBS. In sum, stimulation of the basal ganglia nuclei through DBS could alter one's ability to self-regulate as the applied current intervenes into the neural basis of emotional, motor and cognitive self-regulation.

This exemplifies the potential of DBS to intervene into the complex infrastructure of the human brain with the aim of altering moral capacities including e.g. self-regulatory abilities.

9.3 Normative Considerations of Interventions into the “Moral Brain”

Characteristic of any medical intervention is the need to conduct a harm-benefit¹ analysis. Typically, harm-benefit analyses are performed in the pre-intervention phase with the aim of determining whether the upcoming procedure is justified. In turn, it is used as a basis to retrospectively assess the success of the therapy for the individual patient. In addition, a post-interventional assessment is often used to collect data on the effectiveness and risk of the performed procedure for future patients. Generally, this involves a thorough analysis of the aims for which the intervention is applied followed by rigorously examining whether these aims have been fulfilled. This widely-adopted process also addresses the question of whether the therapeutic aim is at all desirable and therapeutically reachable.

Such a comprehensive ethical harm-benefit assessment certainly also holds for interventions with the aim to change moral behavior. While we have spent some time to substantiate that next generation interventions might soon be installed for changing moral abilities of patients, the ethical question of the desirability of the intervention still needs clarification. As can be anticipated, the pre- and post-interventional harm-benefit analysis will be different depending on whether neuro-moral interventions are at stake. Consistent with the argument presented above, it is important to stress that understanding neurophysiological processes does not replace the evaluation of outcomes. In what follows, we first outline what we call the “standard model” for ethically justifying interventions, and then question the perceived inseparability of intervention aims and outcomes if moral capacities become part of such intervention practices.

9.3.1 *Roots of the “Standard Model”*

What we call the “standard model” of any ethical evaluation of a medical intervention is the clear-cut separation between the definition of (and agreement upon) the intervention goals and the objective verification of whether these goals have been met. The former is usually seen as a normative endeavor, whereas the latter is considered an empirical task.

¹ We deliberately use the expression “harm-benefit analysis” instead of the more commonly used term “risk-benefit” because the latter is misleading.

This “standard model” reflects a basic analytic distinction in ethics – namely the distinction between an “is” (the actual condition; measured in the evaluation) and an “ought” (the desired condition, determined as the therapeutic goal). This requirement of a clear-cut distinction between “is” and “ought” is often referred to in the work of David Hume. In his 1740 published “Treatise of Human Nature” he pointed to the importance to distinguish between indicative or descriptive and imperative or prescriptive statements and that one cannot justify normative claims solely by relying on empirical (descriptive) premises. Furthermore, in contrast to the inspection of the truth-value of descriptive statements, prescriptive statements are not intended to provide verification or falsification, but instead have the character of a legitimation. Further, Hume attests that verification alone can never lead to a legitimation. Accordingly, the is-ought-fallacy is committed if one derives an “ought” from an “is”. Of course, there is a reciprocal relationship in that one is equally wrong when deriving an “is” from an “ought” (termed the moralistic fallacy). To avoid falling into this trap, a clear separation between normative and empirical claims is advocated and most people defend the idea of not confusing these two, presumably separate, domains.

However, not all philosophers agree with that conclusion. For example Hilary Putnam (2002) has argued that the ‘fact–value’ dichotomy originates from an impoverished empiricist conception of ‘fact’ and an equally impoverished positivist understanding of ‘value’. Both ideas seem to be entangled and “a proper understanding of social and scientific change requires the abandonment of this dichotomy” (Callon 1986). We will discuss this point further in the conclusion. [AUS]

9.3.2 Problems of the “Standard Model”

The clear-cut analytic distinction between descriptive and prescriptive statements does not map one-to-one to the distinction between the steps “goal setting” and “goal verification” in the medical domain. Two problems have to be discussed here: First, both steps involve normative and empirical considerations. Second, interventions that concern capacities relevant for goal setting such as self-regulation outlined above provide further complications that undermine the significance of the clear analytical distinction between normative and empirical statements.

The first problem is of less relevance to the core question of whether neuroscientific findings have normative implications. It just points to the fact that both “goal setting” and “goal evaluation” involve empirical and normative practices – although the weights are distributed unevenly, as Table 9.1 outlines. Goal setting is primarily a normative endeavor, as the key questions to answer are normative ones: Is the goal desirable (for the patient, but also taking into account a broader perspective that involves the patient’s social embedding and maybe also societal aspects such as costs)? Does this benefit outweigh risks associated with the intervention? Those questions require a “yes” in order to justify an intervention. However, also determining the therapeutic goal involves an empirical component—in particular, an

Table 9.1 Empirical and normative elements of goal setting and goal verification t1.1

	Normative challenge	Empirical challenge	t1.2
Step 1: Goal setting	Is the goal desirable?	Is reaching the goal possible?	t1.3
	Does this benefit outweigh the harm associated with reaching the goal?		t1.4
			t1.5
Step 2: Goal verification	Are the threshold values that determine “positive” or “negative” verification adequate?	Are the testing tools adequate to measure goal fulfillment?	t1.6
			t1.7
	Are the verification tools ethically acceptable?	Is the goal actually met?	t1.8
			t1.9
			t1.10

assessment of whether the goal actually can be reached in the single case with sufficient probability. But the empirical component by itself does not answer the normative questions mentioned above. For instance, interventions with a low probability of success can be justified in certain cases (assuming that aspects related to distributive justice have been taken into account; Bunnik et al. 2018).

Goal evaluation is primarily an empirical endeavor. One has to use the correct tools in order to assess the success of the intervention and one indeed has to determine that success has been reached in order to claim that the intervention actually worked. However, goal verification also implies normative questions on the methodological level; e.g. when determining threshold levels for “abnormal” values the measurement device generates. In addition, tools for assessing an intervention should be ethically acceptable (e.g., by not inflicting unnecessary pain).

Usually, the intertwining of empirical and normative aspects in goal setting and goal verification can be handled on the practical level of most medical decision-making processes. However, the situation becomes more complex when the intervention concerns moral competences of the person that are directly linked to the normative challenges mentioned above. By “moral competences” we refer to a rich set of abilities that can be summarized by the model of “moral intelligence” (Tanner and Christen 2013) and that involve several of the mechanisms discussed in Sect. 9.2, in particular self-regulation and agency. All four normative challenges (see Table 9.1) can be affected by this second problem: There may be significant disagreement among the involved stakeholders (medical experts, patient, and next of kin) regarding the question, whether the intervention goal (e.g. increasing agency or self-regulation, reducing hypomania, increasing sexual drive and the like) is desirable or regarding the evaluation of risks (of e.g. becoming morally hypersensitive, hypersexual, impulsive). The assessment of risks and benefits as well as the determination of threshold values may strongly depend on societal considerations where no consensus exists. In addition, the ethical acceptability of evaluation tools may suffer from divergent evaluations from both stakeholders and societal considerations.

In particular, the following problems could emerge: First, the affected person (patient) does not think that the goal is desirable – but other persons believe that approval of desirability would be present *after* the intervention (just as in cases of certain manic states characterized by excessive involvement in pleasurable activities

that have a high potential for negative consequences such as engaging in gambling, sexual indiscretions and the like (see e.g. Synofzik et al. 2012)). Then the intervention would require disregarding the capacity of the patient to make an informed decision with respect to the intended behavioral change due to his illness. Second, the affected person thinks that the goal is desirable – but the intervention leads to a situation where the person *post factum* disagrees with the desirability of the goal – for example because the person thinks that one should reach moral improvement through right reasons and not through brain interventions. In other words, the person may disagree with the means that led to the behavioral change after the intervention. This is particularly true, if she follows a deontological, virtue ethics, or other non-consequentialist understanding of morality, whereas a consequentialist may be more willing to *post factum* accept the kind of intervention used. This imposes a conflict concerning goal verification: the person will not agree with what the verification tells. Third, interaction effects may occur, in particular if the goal of the intervention is to benefit a whole system (e.g., social environment of the person, maybe even societal institutions, by improving impulse control). The problem here is that there is not a “one-way road” from intervening on the biological system all the way up to a desired social change. Changes on all levels (biological, psychological, social, societal) will feed back onto other levels. What looks like a simple feed-forward process from biological intervention up to desired societal changes is actually a complex system that requires a much more expanded view of the system that is changed by an intervention (see the considerations in the next section).

9.3.3 Challenges of Neuro-modulating Moral Behavior

In Sect. 9.2, we argued that it seems conceivable that in the future, we can selectively interfere with the neural circuits to improve moral capacities possibly relating to self-regulation, sense of agency, which may also have positive effects on cooperation, trust, generosity or moral judgment and behavior in general. Albeit the time is not now ripe for targeted interventions to improve self-regulatory abilities, but we base our considerations on the assumption that neuromodulation of the “moral brain” might soon become a realistic intervention option.

9.3.3.1 Justifying Targeted Neuro-moral Interventions

Probably the most pressing consideration against the use of neuromoral interventions involves potential harms to the individual. The underlying notion suggests that there are simply no safe means such that moral changes are reliable and free from severe side effects. While the experiences made with psychotropic drugs such as ataractics make it plausible that pharmacological interventions are likely to have side effects and may fail in a considerable number of cases, it is not impossible that other means of interventions (e.g. DBS) may have more reliable effects.

Strikingly, the domain of violent behavioral manifestations is probably one of the few contexts for which a widely shared consensus on the acceptability of the intervention exists (De Ridder et al. 2009). In particular, if aggressive tendencies are at stake that are linked to direct physical harm towards third parties, as in some forms of antisocial personality disorders. Justified interventions appeal therefore just to interventions that comprise more or less uncontroversial “bad behaviors”. Needless to say, the number of contexts for which one anticipates such a consensual classification is presumably rather low. Taken together, interventions with a moral aim may only be justified in cases where one addresses severe moral deficiencies with acute risks for third parties, such as attempts to correct severe aggression through DBS (Benedetti-Isaac et al. 2015).

9.3.3.2 Problems Beyond Patient Harm

However, given the numerous disastrous circumstances that are mainly man-made, the aim of improving humankind towards being more cooperative and peaceable is appealing (Persson and Savulescu 2012; Shook 2012). Therefore, it seems tempting to pursue the goals already chronicled by Skinner and many others before and after him. However, it is important to note that for many such ethically desirable behaviors, the situation is substantially more complex as many of these terms are inherently vague, depend on cultural contexts and have dramatically changed over time. Here, the “standard model” is confronted with additional problems when it comes to targeted interventions into the neural system for changing moral behavior.

The first, general point to make is that the neuropsychiatric domain poses exceptional challenges when making clear-cut distinctions between the empirical and the normative domain. This refers to psychiatry’s subliminal claims about the definition of socially acceptable and desirable behaviors. No other medical discipline operates in such an area of conflict between biology, philosophy and human value judgment. Accordingly, diagnostic criteria have not only been adjusted due to cultural change but in addition, have been the object of considerable controversies. It is plausible to assume that extending the therapeutic spectrum of neuropsychiatric interventions explicitly into the moral domain will lead to much stronger controversies. The question “Is the goal desirable?” will be hard to answer as soon as the expression “improve ethical behavior” is filled with content.

Then, on the validation side, there would be a substantial difficulty of assessing moral capacities in a reliable and valid way, given the scarcity of instruments in that domain. The reason for this scarcity is that the empirical postoperative analysis is confronted with a difficult measurement problem. Complex interventions may work best if tailored to local circumstances rather than being completely standardized. But lack of standardization would confound empirically robust research designs. The rationale for a complex intervention, the changes that are expected, and how change is to be achieved may not be clear at the outset, particularly if moral capacities are part of the intervention goals.

In addition to this methodological challenge, assessing the outcome of changes to peoples' moral capacities might very well be undermined by problems of acceptability (political or ethical objections to the intervention goals), recruitment, replication, randomization, instrument scarcity and smaller than expected effect sizes (see e.g. Craig et al. 2008). Moreover, the issue of the putative reversibility of interventions such as DBS would be an important factor for the acceptability of research into this direction. Given the vagueness of the primary study outcome (moral capacities), statistical analysis would most probably have to be replaced or adjusted by a hermeneutical, philosophical and psychological evaluation. A long-term follow-up period might also be necessary to determine whether outcomes predicted by interim or surrogate measures do occur or whether short term changes persist (Craig et al. 2008).

A further caveat relates to intervention optimization: what is the optimal level of a moral capacity such as moral sensitivity (the ability to perceive moral issues when they arise in practice; Ineichen et al. 2017) between the poles of moral blindness to moral hypersensitivity? The key to be clear about is how much change or adaption is permissible and to record outcome variation. Owing to the problem of reliably measuring such complex changes as well as strict standardization, measurement fidelity is not straightforward in the context of such complex interventions.

In summary, major prerequisites for developing and evaluating complex interventions into the "moral brain" generally include a solid theoretical understanding of both the intervention goal (which is hard to achieve for moral capacities, due to their pluralistic notions) and how the intervention causes change (which is also hard to achieve, given the scarce knowledge on the neural underpinnings of moral capacities). As a consequence, lack of effectiveness may reflect implementation failure, genuine ineffectiveness or outcome measurement insensitivity. Due to higher-level processes involved in moral capacities, variability in individual outcomes is highly probable. As a consequence, sample sizes would need to be large to take into account extra variability, and mixed-methods designs should be considered. Given the vagueness of the primary outcome, multiple measures will be needed rather than a single outcome measurement.

9.4 Conclusion

Recent neurotechnological advances have shown their potential in modulating aberrant physiological processes. In future, reliable and safe neurointerventions will be present for targeted procedures aiming at altering moral abilities of individuals. Through the provision of insight in identifying important neuroanatomical nodes that achieve neural computations relevant for moral abilities of patients, deep brain stimulation has offered a first understanding of how to intervene into the complex infrastructure of socio-moral signal processing. In particular, the impact of neurointerventions on agency and self-regulation as two relevant constituents of one's personality that may influence moral behavior, has been outlined. The normative

analysis revealed that the intertwining of empirical and normative aspects in goal setting and goal verification, a phenomenon that can be handled rather well in most medical decision-making processes, will face substantial problems when altering moral capacities becomes a therapeutic practice. Even if maintaining an analytical distinction between the normative and empirical perspective appears reasonable, the practical handling of neuromodulation techniques that involve normative intervention goals is likely to show the limit of this theoretical distinction. It remains an open question whether the unity of the Thomistic worldview that provided impressive coherence to the medieval era and that was later replaced by a modern trichotomy regains relevance. In particular, the distinction between the moral and empirical domain is challenged when facing the increased intertwining of empirical and normative aspects in goal setting and goal verification in the context of neuromodulating moral behavior.

We close with the realization that through recent advances in neurotechnologies, neuroscience has shown to provide a new and refined basis for understanding the neural foundations of human moral abilities. Regarding the normative implications of neuroscience it can be concluded that neuroscience does have normative implications in at least two respects: when therapeutic interventions alter moral capacities of patients and when neuro-modulatory practices impact on decision-making processes for justifying medical interventions.

References

- Avena-Koenigsberger, A., B. Misic, and O. Sporns. 2018. Communication Dynamics in Complex Brain Networks. *Nature Reviews Neuroscience* 19 (1): 17.
- Ballanger, B., T. van Eimeren, E. Moro, A.M. Lozano, C. Hamani, P. Boultinguez, G. Pellecchia, S. Houle, Y.Y. Poon, A.E. Lang, and A.P. Strafella. 2009. Stimulation of the Subthalamic Nucleus and Impulsivity: Release Your Horses. *Annals of Neurology* 66 (6): 817–824.
- Bandura, A., G.V. Caprara, C. Barbaranelli, M. Gerbino, and C. Pastorelli. 2003. Role of Affective Self-Regulatory Efficacy in Diverse Spheres of Psychosocial Functioning. *Child Development* 74 (3): 769–782.
- Baumeister, R.F., M. Gailliot, C.N. DeWall, and M. Oaten. 2006. Self-Regulation and Personality: How Interventions Increase Regulatory Success, and How Depletion Moderates the Effects of Traits on Behavior. *Journal of Personality* 74 (6): 1773–1802.
- Benedetti-Isaac, J.C., M. Torres-Zambrano, A. Vargas-Toscano, E. Perea-Castro, G. Alcalá-Cerra, L.L. Furlanetti, T. Reithmeier, T.S. Tierney, C. Anastasopoulos, E.T. Fonoff, and W.O. Contreras Lopez. 2015. Seizure Frequency Reduction After Posteromedial Hypothalamus Deep Brain Stimulation in Drug-Resistant Epilepsy Associated with Intractable Aggressive Behavior. *Epilepsia* 56 (7): 1152–1161.
- Berker, S. 2009. The Normative Insignificance of Neuroscience. *Philosophy & Public Affairs* 37 (4): 293–329.
- Bratman, M.E. 2000. Reflection, Planning, and Temporally Extended Agency. *Philosophical Review*: 35–61.
- Bunnik, E.M., N. Aarts, and S. van de Vathorst. 2018. Little to Lose and No Other Options: Ethical Issues in Efforts to Facilitate Expanded Access to Investigational Drugs. *Health Policy*. 2018 Jun 18. pii: S0168-8510(18)30184-2. doi: 10.1016/j.healthpol.2018.06.005. [Epub ahead of print].

- Callon, M. 1986. The Sociology of an Actor-Network: The Case of the Electric Vehicle. In *Mapping the Dynamics of Science and Technology*, 19–34. London: Palgrave Macmillan.
- Casebeer, W.D. 2003. Moral Cognition and Its Neural Constituents. *Nature Reviews Neuroscience* 4 (10): 840.
- Casey, B.J., J.N. Epstein, J. Buhle, C. Liston, M.C. Davidson, S.T. Tonev, J. Spicer, S. Niogi, A.J. Millner, A. Reiss, A. Garrett, S.P. Hinshaw, L.L. Greenhill, K.M. Shafritz, A. Vitolo, L.A. Kotler, M.A. Jarrett, and G. Clover. 2007. Frontostriatal Connectivity and its Role in Cognitive Control in Parent-Child Dyads with ADHD. *American Journal of Psychiatry* 164 (11): 1729–1736.
- Castrioto, A., E. Lhommée, E. Moro, and P. Krack. 2014. Mood and Behavioural Effects of Subthalamic Stimulation in Parkinson's Disease. *The Lancet Neurology* 13 (3): 287–305.
- Cathcart, T. 2013. *The Trolley Problem, or Would You Throw the Fat Guy Off the Bridge?: A Philosophical Conundrum*. Workman Publishing.
- Christen, M., and S. Müller. 2012. Current Status and Future Challenges of Deep Brain Stimulation in Switzerland. *Swiss Medical Weekly* 142: w13570.
- Christen, M., and M. Regard. 2012. Der 'unmoralische Patient'. Eine Analyse der Nutzung hirnerkrankter Menschen in der Moralforschung. *Nervenheilkunde* 31: 209–214.
- Craig, P., P. Dieppe, S. Macintyre, S. Michie, I. Nazareth, and M. Petticrew. 2008. Developing and Evaluating Complex Interventions: The New Medical Research Council Guidance. *BMJ* 337: a1655.
- Da Cunha, C., S.L. Boschen, A.A. Gómez, E.K. Ross, W.S. Gibson, H.K. Min, K.H. Lee, and C.D. Blaha. 2015. Toward Sophisticated Basal Ganglia Neuromodulation: Review on Basal Ganglia Deep Brain Stimulation. *Neuroscience & Biobehavioral Reviews* 58: 186–210.
- De Ridder, D., B. Languth, M. Plazier, and T. Menowsky. 2009. Moral Dysfunction and Potential Treatments. In *The Moral Brain*, ed. J. Verplaetse et al., 155–183. Berlin: Springer.
- Deuschl, G., C. Schade-Brittinger, P. Krack, J. Volkmann, H. Schäfer, K. Bötzel, C. Daniels, A. Deuschländer, U. Dillmann, W. Eisner, and D. Gruber. 2006. A Randomized Trial of Deep-Brain Stimulation for Parkinson's Disease. *New England Journal of Medicine* 355 (9): 896–908.
- Dolan, R.J. 1999. On the Neurology of Morals. *Nature Neuroscience* 2 (11): 927.
- Earp, B.D., T. Douglas, and J. Savulescu. 2017. Chapter 11: Moral Neuroenhancement. In *The Routledge Handbook of Neuroethics*, ed. Johnson LSM and K.S. Rommelfanger. New York: Routledge.
- El-Hai, J. 2005. *The Lobotomist: A Maverick Medical Genius and His Tragic Quest to Rid the World of Mental Illness*. Hoboken, NJ: Wiley.
- Follett, K.A., F.M. Weaver, M. Stern, K. Hur, C.L. Harris, P. Luo, W.J. Marks Jr., J. Rothlind, O. Sagher, C. Moy, and R. Pahwa. 2010. Pallidal Versus Subthalamic Deep-Brain Stimulation for Parkinson's Disease. *New England Journal of Medicine* 362 (22): 2077–2091.
- Foot, P. (1967). *The Problem of Abortion and the Doctrine of Double Effect*.
- Frank, M.J., J. Samanta, A.A. Moustafa, and S.J. Sherman. 2007. Hold Your Horses: Impulsivity, Deep Brain Stimulation, and Medication in Parkinsonism. *Science* 318 (5854): 1309–1312.
- Fumagalli, M., and A. Priori. 2012. Functional and Clinical Neuroanatomy of Morality. *Brain* 135 (7): 2006–2021.
- Gehlen, A., and D. Mensch. 1950. *Seine Natur und seine Stellung in der Welt*. Bonn: Verbesserte Auflage.
- Glannon, W., and C. Ineichen. 2016. Philosophical Aspects of Closed-Loop Neuroscience. In *Closed Loop Neuroscience*, 259–270.
- Greene, J.D., R.B. Sommerville, L.E. Nystrom, J.M. Darley, and J.D. Cohen. 2001. An fMRI Investigation of Emotional Engagement in Moral Judgment. *Science* 293 (5537): 2105–2108.
- Gross, J.J. 1998. The Emerging Field of Emotion Regulation: An Integrative Review. *Review of General Psychology* 2 (3): 271.
- Hack, N., U. Akbar, A. Thompson-Avila, S.M. Fayad, E.M. Hastings, E. Moro, K. Nestor, H. Ward, M. York, and M.S. Okun. 2014. Impulsive and Compulsive Behaviors in Parkinson Study

- Group (PSG) Centers Performing Deep Brain Stimulation Surgery. *Journal of Parkinson's Disease* 4 (4): 591–598. 730
- Haggard, P. 2017. Sense of Agency in the Human Brain. *Nature Reviews Neuroscience* 18 (4): 196. 732
- Haidt, J. 2007. The New Synthesis in Moral Psychology. *Science* 316 (5827): 998–1002. 733
- Hamani, C., M.P. McAndrews, M. Cohn, M. Oh, D. Zumsteg, C.M. Shapiro, R.A. Wennberg, and A.M. Lozano. 2008. Memory Enhancement Induced by Hypothalamic/Fornix Deep Brain Stimulation. *Annals of Neurology* 63 (1): 119–123. 734
- Hariz, M. 2012. Twenty-Five Years of Deep Brain Stimulation: Celebrations and Apprehensions. *Movement Disorders* 27 (7): 930–933. 737
- Harris, S. 2011. *The Moral Landscape: How Science Can Determine Human Values*. Simon and Schuster. 738
- Ineichen, C., and M. Christen. 2017. Hypo- and Hyperagentic Psychiatric States, Next-Generation Closed-Loop DBS, and the Question of Agency. *AJOB Neuroscience* 8 (2): 77–79. 741
- Ineichen, C., H. Baumann-Vogel, and M. Christen. 2016. Deep Brain Stimulation: In Search of Reliable Instruments for Assessing Complex Personality-Related Changes. *Brain Sciences* 6 (3): 40. 742
- Ineichen, C., M. Christen, and C. Tanner. 2017. Measuring Value Sensitivity in Medicine. *BMC Medical Ethics* 18 (1): 5. 743
- Ineichen, C., N.R. Shepherd, and O. Sürücü. 2018. Understanding the Effects and Adverse Reactions of Deep Brain Stimulation: Is It Time for a Paradigm Shift Towards a Focus on Heterogenous Biophysical Tissue Properties Instead of Electrode Design Only? *Frontiers in Human Neuroscience*. under review. 744
- Jahanshahi, M., I. Obeso, C. Baunez, M. Alegre, and P. Krack. 2015. Parkinson's Disease, the Subthalamic Nucleus, Inhibition, and Impulsivity. *Movement Disorders* 30 (2): 128–140. 745
- Jonides, J., and D.E. Nee. 2006. Brain Mechanisms of Proactive Interference in Working Memory. *Neuroscience* 139 (1): 181–193. 746
- Jonides, J., C. Marshuetz, E.E. Smith, P.A. Reuter-Lorenz, R.A. Koeppe, and A. Hartley. 2000. Age Differences in Behavior and Pet Activation Reveal Differences in Interference Resolution in Verbal Working Memory. *Journal of Cognitive Neuroscience* 12 (1): 188–196. 747
- Kahneman, D., and D. Chajczyk. 1983. Tests of the Automaticity of Reading: Dilution of Stroop Effects by Color-Irrelevant Stimuli. *Journal of Experimental Psychology: Human Perception and Performance* 9 (4): 497. 748
- Kennett, J. 2001. *Agency and Responsibility: A Common-Sense Moral Psychology*. Oxford University Press. 749
- Kober, H., P. Mende-Siedlecki, E.F. Kross, J. Weber, W. Mischel, C.L. Hart, and K.N. Ochsner. 2010. Prefrontal–Striatal Pathway Underlies Cognitive Regulation of Craving. *Proceedings of the National Academy of Sciences* 107 (33): 14811–14816. 750
- Kopell, B.H., and B.D. Greenberg. 2008. Anatomy and Physiology of the Basal Ganglia: Implications for dbS in Psychiatry. *Neuroscience & Biobehavioral Reviews* 32 (3): 408–422. 751
- Lambert, C., L. Zrinzo, Z. Nagy, A. Lutti, M. Hariz, T. Foltynie, B. Draganski, J. Ashburner, and R. Frackowiak. 2015. Do We Need to Revise the Tripartite Subdivision Hypothesis of the Human Subthalamic Nucleus (STN)? Response to Alkemade and Forstmann. *NeuroImage* 110: 1–2. 752
- Laxton, A.W., D.F. Tang-Wai, M.P. McAndrews, D. Zumsteg, R. Wennberg, R. Keren, et al. 2010. A Phase I Trial of Deep Brain Stimulation of Memory Circuits in Alzheimer's Disease. *Annals of Neurology* 68 (4): 521–534. 753
- Liston, C., M.M. Miller, D.S. Goldwater, J.J. Radley, A.B. Rocher, P.R. Hof, and B.S. McEwen. 2006. Stress-Induced Alterations in Prefrontal Cortical Dendritic Morphology Predict Selective Impairments in Perceptual Attentional Set-Shifting. *Journal of Neuroscience* 26 (30): 7870–7874. 754
- Macmillan, M. 2000. *An Odd Kind of Fame. Stories of Phineas Gage*. Cambridge, MA: MIT-Press. 755
- Mele, A.R. 2001. *Autonomous Agents: From Self-Control to Autonomy*. Oxford University Press on Demand. 756

- Müller, S., and M. Christen. 2011. Deep Brain Stimulation in Parkinsonian Patients—Ethical Evaluation of Cognitive, Affective, and Behavioral Sequelae. *AJOB Neuroscience* 2 (1): 3–13.
- Nambu, A. 2008. Seven Problems on the Basal Ganglia. *Current Opinion in Neurobiology* 18 (6): 595–604.
- Olanow, C.W., F. Stocchi, and A. Lang, eds. 2011. *Parkinson's Disease: Non-motor and Non-dopaminergic Features*. Wiley.
- Persson, I., and J. Savulescu. 2012. *Unfit for the Future: The Need for Moral Enhancement*. Uehiro Series in Practical Ethics. Oxford: Oxford University Press.
- Petrides, M., and D.N. Pandya. 1999. Dorsolateral Prefrontal Cortex: Comparative Cytoarchitectonic Analysis in the Human and the Macaque Brain and Corticocortical Connection Patterns. *European Journal of Neuroscience* 11 (3): 1011–1036.
- Poldrack, R.A., and M.J. Farah. 2015. Progress and Challenges in Probing the Human Brain. *Nature* 526 (7573): 371.
- Putnam, H. 2002. *The Collapse of the Fact/Value Dichotomy and Other Essays*. Harvard University Press.
- Rüfner, V. 1964. Das Personsein im Lichte gestalthaft-genetischer Betrachtungsweise: Im Hinblick auf das religiöse Erleben. *Archiv für Religionspsychologie/Archive for the Psychology of Religion*: 231–248.
- Schapira, A.H., K.R. Chaudhuri, and P. Jenner. 2017. Non-motor Features of Parkinson Disease. *Nature Reviews Neuroscience* 18 (7): 435.
- Sesack, S.R., and A.A. Grace. 2010. Cortico-Basal Ganglia Reward Network: Microcircuitry. *Neuropsychopharmacology* 35 (1): 27–47.
- Shook, J.R. 2012. Neuroethics and the Possible Types of Moral Enhancement. *AJOB Neuroscience* 3 (4): 3–14.
- Skinner, B.F., B.F. Skinner, E.U. Psychologue, B.F. Skinner, and B.F. Skinner. 1972. *Beyond Freedom and Dignity (No. 04; BF319. 5. O6, S5.)*. New York: Bantam Books.
- Somerville, L.H., T.F. Heatherton, and W.M. Kelley. 2006. Anterior Cingulate Cortex Responds Differentially to Expectancy Violation and Social Rejection. *Nature Neuroscience* 9 (8): 1007–1008.
- Synofzik, M., T.E. Schlaepfer, and J.J. Fins. 2012. How Happy Is Too Happy? Euphoria, Neuroethics, and Deep Brain Stimulation of the Nucleus Accumbens. *AJOB Neuroscience* 3 (1): 30–36.
- Tangney, J.P., R.F. Baumeister, and A.L. Boone. 2004. High Self-Control Predicts Good Adjustment, Less Pathology, Better Grades, and Interpersonal Success. *Journal of Personality* 72 (2): 271–324.
- Tanner, C., and M. Christen. 2013. Moral Intelligence: A Framework for Understanding Moral Competences. In *Empirically Informed Ethics: Morality Between Facts and Norms*, ed. M. Christen, J. Fischer, M. Huppenbauer, C. Tanner, and C. van Schaik, 119–136. Berlin: Springer.
- Temel, Y., A. Blokland, H.W. Steinbusch, and V. Visser-Vandewalle. 2005. The Functional Role of the Subthalamic Nucleus in Cognitive and Limbic Circuits. *Progress in Neurobiology* 76 (6): 393–413.
- Thomson, J.J. 1985. The Trolley Problem. *The Yale Law Journal* 94 (6): 1395–1415.
- Tremblay, L., Y. Worbe, S. Thobois, V. Sgambato-Faure, and J. Féger. 2015. Selective Dysfunction of Basal Ganglia Subterritories: From Movement to Behavioral Disorders. *Movement Disorders* 30 (9): 1155–1170.
- Valenstein, E.S. 1986. *Great and Desperate Cures: The Rise and Decline of Psychosurgery and Other Radical Treatments for Mental Illness*. Basic Books.
- Wallace, R.J. 1994. *Responsibility and the Moral Sentiments*. Harvard University Press.
- Weaver, F.M., K. Follett, M. Stern, K. Hur, C. Harris, W.J. Marks, J. Rothlind, O. Sagher, D. Reda, C.S. Moy, and R. Pahwa. 2009. Bilateral Deep Brain Stimulation vs Best Medical Therapy for Patients With Advanced Parkinson Disease: A Randomized Controlled Trial. *JAMA* 301 (1): 63–73.

Welt, L. 2009. Über Charakterveränderungen des Menschen infolge von Läsionen des Stirnhirns.	836
<i>Deutsches Archiv für klinische Medizin</i> 42: 339–390.	837
Wikström, P.O.H. 2005. The Social Origins of Pathways in Crime: Towards a Developmental	838
Ecological Action Theory of Crime Involvement and Its Changes. <i>Integrated Developmental</i>	839
<i>and Life-Course Theories of Offending</i> 14: 211–245.	840
Young, Liane, and James Dungan. 2012. Where in the Brain Is Morality? Everywhere and Maybe	841
Nowhere. <i>Social Neuroscience</i> 7 (1): 1–10.	842
Youngerman, Brett E., Andrew K. Chan, Charles B. Mikell, Guy M. McKhann, and Sameer	843
A. Sheth. 2016. A Decade of Emerging Indications: Deep Brain Stimulation in the United	844
States. <i>Journal of Neurosurgery</i> 125 (2): 461–471.	845

Uncorrected Proof

Author Queries

Chapter No.: 9 0004913897

Queries	Details Required	Author's Response
AU1	The citation "Young et al., 2012" has been changed to "Young and Dungan 2012" to match the author name/date in the reference list. Please check if the change is fine in this occurrence and modify the subsequent occurrences, if necessary.	
AU2	The citation "Skinner, 1972" has been changed to "Skinner et al. 1972" to match the author name/date in the reference list. Please check if the change is fine in this occurrence and modify the subsequent occurrences, if necessary.	
AU3	The citation "Gehlen, 1966" has been changed to "Gehlen and Mensch 1950" to match the author name/date in the reference list. Please check if the change is fine in this occurrence and modify the subsequent occurrences, if necessary.	
AU4	The citation "Lambert et al., 2012" has been changed to "Lambert et al. 2015" to match the author name/date in the reference list. Please check if the change is fine in this occurrence and modify the subsequent occurrences, if necessary.	
AU5	The citation "Callon et al., 1986" has been changed to "Callon 1986" to match the author name/date in the reference list. Please check if the change is fine in this occurrence and modify the subsequent occurrences, if necessary.	