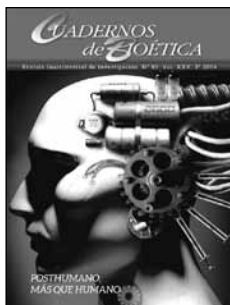


**POST-HUMAN ETHICS AND ESTHETICS**

**ÉTICA Y ESTÉTICA POSTHUMANA**





## IDENTITY AND CONFLICTS IN THE ETHICS OF NEURAL IMPLANTS

### IDENTIDAD Y CONFLICTO EN LA ÉTICA DE LOS IMPLANTES NEUROPROSTÉTICOS

L. E. ECHARTE AND M. GARCÍA-VALDECASAS

*Instituto Cultura y Sociedad, Universidad de Navarra*

*31009 Pamplona (Spain)*

*lecharte@unav.es; garciaval@unav.es*

#### ABSTRACT:

##### Keywords:

Extended mind theory, neural prosthetics, personal identity, neuroethics.

The development of neuroprosthetics has given rise to significant theoretical and practical challenges concerning personal identity. The Extended Mind Theory (EMT) attempts to provide an answer to these challenges by arguing that the mind and the external world are co-extensive to the point that both can make a seamless unified entity. The EMT also proposes that physical states determine the nature of mental states. Here, we propose a non-deterministic and less locationist view of mental states that we will call iEMT. The iEMT articulates, firstly, that the co-extensivity of the mind and the world does not justify the dissolution of the mind in the objects of the external world with which the mind interacts. Consequently, the agent's mind is still part of his unique personal identity. Secondly, neural implants cannot be regarded as mere replacement parts in the context of a weak concept of personal identity. Thirdly, there are no compelling reasons to believe or to fear that neuroprosthetics can alter personal identity at the profound level.

#### RESUMEN:

##### Palabras clave:

Teoría de la mente extendida, prótesis neural, identidad personal, neuroética.

El desarrollo de la tecnología neuroprostética está generando importantes problemas teóricos y prácticos relacionados con la identidad personal. En este contexto, la Extended Mind Theory (EMT) es una teoría que da una respuesta a dichos problemas señalando que la mente es coextensa con el mundo, y que como tal, la mente y el mundo se entremezclan hasta forman una única entidad. La EMT también presupone que los estados físicos determina la naturaleza de los estados mentales. En este artículo, proponemos una versión no determinista y aún menos locacionista de los estados mentales que la de la EMT a la que llamaremos iEMT. Defendemos, primero, que la coextensión de mente y mundo no justifica la disolución de la mente, y como consecuencia, de la identidad humana en el medio; segundo, que no está justificado pensar que los neuro-implantes constituyan simples piezas de reemplazo en el contexto de una identidad que llamamos débil; y tercero, que no hay razones de peso para creer, y en esa medida, para temer, que el uso de tecnología neuroprostética pueda alterar la identidad personal.

Recibido: 18/03/2013

Aceptado: 03/06/2014

## 1. The arrival of neural implants

For many years, the biotechnology industry has been researching into and developing cognitive neural implants (CNIs). CNIs are sophisticated prostheses that can be successfully installed by surgery into the central nervous system to improve cognitive functions. They are a crucial landmark in the medical treatment of physical impairments. Amid high life expectancy and enhanced quality of life, industrial efforts to leverage CNIs to further the limits of intellectual abilities in healthy subjects have also yielded promising results. Hence, if the use of CNIs with such end succeeds, neuroprosthetics may become a means of providing body modifications on demand. Given the potential of neuroprosthetics, the prospect of a more widespread use of this technology raises a number of ethical and philosophical questions. For one thing, CNIs may fuel the belief that humans are a mere collection of cells. But can we really act and understand ourselves in this way? Is man a kind of machine that can undergo the most daring of modifications? Are there ultimate limits to the manipulation and transformation of the body?

In a sense, this controversy is nothing new. Medicine has endeavored to replace damaged body parts with mechanical devices since the progress of surgery allowed it by implanting artificial legs, pacemakers, cochlear implants and stents. At the same time, although transplanting is easier than reproducing or repairing some organ, transplants have proved to be an effective and reliable means to recover biological functions. Interestingly, as the number of transplants increased, since the very earliest studies of organ transplantation (particularly kidney and heart), scientific literature has discussed problems relating to self-perception.

In particular, scientific literature has generally discovered three kinds of conflicts of identity in cases of organ transplant: a) conflicts of authenticity, e.g., "after surgery, the beating heart in my chest is not really mine, this is a foreign body"; b) identity conflicts: "my old heart is gone, I am no longer myself, I'm someone else"; and c) alienation conflicts: "with this new heart I will be perceived as a machine, I no longer control my

own life"<sup>1</sup>. When the replaced organ is, or is perceived to be by the patient closely associated with his identity, the conflict is even more intense. It is only natural then that patients that receive face transplants are prone to struggles of alienation, where conflicts are more intense and are well documented.

Conflicts of the kinds described were reported in patients with CNIs, particularly in those undergoing deep brain stimulation (DBS). Given the strong link between the brain and our self-perception it is no trivial matter that many patients complain that the CNI has altered their identity. There is a big difference between transplanting a kidney and modifying, even slightly, a patient's brain. The intense neuromodulation produced by CNIs frequently affects mental states, transforming beliefs, emotions, memory and attention. As expected, the patient's personality is also affected, giving rise to conflicts of authenticity, alienation and identity to a greater extent than conflicts associated, for example, with a kidney transplant.

Many patients claim that with a CNI, their strongest held opinions about life and their self are now controlled by a switch<sup>2</sup>. This perception provokes profound reflections about the identity of the self and its place in its environment. For example, when can the patient be said to be himself, when the implant is on or off? And above all, was there ever a real me? In the absence of any prior reflections, the patient with the implant is naturally driven to believe that the parts of his current body somehow exclusively determine his demeanor. Further, implants have been reported to largely unsettle long-established habits and personal relations, changes that, paradoxically, in some cases the patient is able to reverse<sup>3</sup>. As some of them report that their new self seems to be random or short of any lasting significance, the patient's reaction to his modified personality becomes understandable.

1 House, R.M. Thompson, T.L. «Psychiatric aspects of Organ Transplantation». *JAMA* 260(4), (1998), 535-539.

2 Halpern, C.H., Rick, J.H., Danish, S.F., Grossman, M., Baltuch, G.H. «Cognition following bilateral deep brain stimulation surgery of the subthalamic nucleus for Parkinson's disease». *Int J Geriatr Psychiatry* 24, (2009), 443-51.

3 Schüpbach, M., Agid, Y. «Psychosocial adjustment after deep brain stimulation in Parkinson's disease». *Nature Reviews Neurology* 4, (2008), 4: 58-59.

Amid wider technology use, CNIs are updating and reformulating the classical debate about human identity. To better contextualize this debate, we must ask ourselves, what is human identity? If CNIs not only modify tissues and organs but also affect self-perception, is there any reason to stick with the organs that we have and the way in which we perceive ourselves? If the answer to this question is no, human identity should not be a big issue and no further discussion should ensue. But if the answer to this question is yes, CNIs may seriously jeopardize personal identity. And here, we propose a view of personal identity that addresses this challenge.

The view that conscious states mark the identity of rational creatures is old. In his sixth *Meditation*, Descartes contemplated the possibility that God could destroy our body and leave our mind intact, concluding that he certainly could. To make sense of them, analysis should separate both mind and body. Consciousness, which in itself is not a kind of physical phenomenon, became the defining criterion of the mental with Descartes. Its phenomena included understanding, willing, and every form of sensation. As a result, he assumed that self-conscious rational substances bear a distinctive identity that contrasts with the identity of mere physical entities, a view subsequently endorsed by others with more or less ambitious rationalistic projects. By the end of the 20<sup>th</sup> Century, though, physicalists fully rejected Descartes's dualism and replaced it with monistic views according to which every mental property ultimately supervenes on the physical or is necessitated by it.

In a physicalist context, we encounter one of the most recent theories of the mind, the extended mind theory (EMT). Born in the current philosophy of mind, the EMT expands the boundaries of the mental and its physical constituents beyond the brain-and-skull barrier into the external environment. The theory is flexible and deep enough to have a number of possible applications in neuroethics; i.e., the EMT can be used to promote the man-machine paradigm, or the view that an individual with a CNI ultimately is a complex machine made up of disposable physical or biological parts. As this applica-

tion of the EMT seems fully compatible with the view of Clark and Chalmers, who are the fathers of the EMT, it is inspiring ethical controversy.

## 2. The proposal of the EMT

For decades, in line with physicalism, philosophy and neuroscience gradually came to accept that the mind is a neural structure in the brain, where it resides or operates as a complex function. But according to Clark and Chalmers<sup>4</sup>, who developed the EMT, this neuro-centric approach is insufficient to account for the true complexity of the simplest of mental phenomena. The EMT argues that the mind is not limited to that which lies beneath the skin and skull but extends to objects that interact with the brain. Depending on the degree of integration of these elements in the cognitive process, these objects can or should be deemed as much a "part of the mind" as the brain. For all elements that are causally involved in cognitive processes should be considered every bit as "mental" as thinking.

In this way, the EMT presents a concept of the mind that is co-extensive with the world. We can conceive of this co-extensivity through an example. Consider an elementary cognitive task, such as ordering books on a shelf according to size. If the advocates of the EMT are correct, the physical properties of books that the eye detects and which physically interact with the brain are part of the cognitive process as real causes of such interaction. The size of the books guides the movements of the subject as he proceeds. Indeed, the movements of his hands would be very different if the books were of different sizes. Therefore, the size of the book is not accidental or merely contingent on the cognitive process that precedes the movement of the hand; they are actually the cause of the process. Put in different words, (i) since the size of the book shapes the electromagnetic waves that interact with the eye, (ii) cognitive processes "learn" to convert the information that travels through these electromagnetic waves into the accurate hand movements that put each book

<sup>4</sup> Clarke, A, Chalmers D. «The extended mind». *Analysis* 58(1), (1998), 7-19.

in its place. Therefore, we may conclude that if (ii) is causally dependent on (i), (ii) is undeniably a causal part of the activity.

Although the newly acquired role played by CNI in cognition does not form part of the initial hypothesis of the EMT, whose original context is philosophical, the theory can be applied to CNIs as much as they are involved in cognitive processes. CNIs are designed to replace dysfunctional neurons and enable the stimulation of neural circuitry. Thus, if a patient with a CNI is capable of producing neuronal waves that restore or increase his ability to perform a calculation or recall old memories, the CNI, as well as the mental states that the CNI brings about, should be fully considered a part of the neural system. As the proponents of the EMT usually put it, epistemic action requires the extension of epistemic credit beyond the mind. If it is true that all that is directly or indirectly involved in the production of cognitive acts can be regarded as part of the mind, the CNI clearly meets these requirements so as to be considered an element of cognition.

In neuroprosthetics, some interpreters of the EMT have come to argue that we must still draw the consequences of a more externalist view of the mind. They think that the existential struggle of patients with neural implants ultimately stems from some misperception what their mind does. Some of these patients may not know that when their mind works with its environment it becomes a same thing with it. In that blend, patients should stop wondering what they essentially are.

### 3. Breakdown of the functional unit

An inevitable consequence of the EMT is that the human being ceases to be a biological and functional unit. In recent decades, philosophers have consistently questioned or denied the existence of a unified self. According to Parfit, the ability to replicate brain states involves the ability to replicate mental states, thereby creating beings that are psychologically continuous. Certainly, the idea that mental states can be replicated has spurred much controversy in the past. For psychological continuity—ideas, experiences or memories that are directly or indirectly perceived by the individual

over time—was thought to form the basis of the concept of personal identity. But according to Parfit, the relevant constituents of personal identity are those that can be found in the brain, the whole organism and the physical events that relate the two<sup>5</sup>. In this way, psychological continuity ceases to be a relevant criterion in the philosophical debate on personal identity.

The EMT holds that the brain and the body are not the sole sites of cognition, as the beliefs of an individual and his surrounding environment form a causal continuum with his mental states. According to Clark and Chalmers, “beliefs can be partly constituted by features of the environment”<sup>6</sup>. While the parts involved in the cognitive process receive epistemic credit, these parts are also their mental states. Accordingly, the relevant elements of personal identity are not confined in the limits of the brain-and-skull barrier. Hence, if the individual is actually a causal chain of interrelated and unitary physical or biological events, the psychological or conscious unit, now displaced outwards, forcibly loses its privileged position.

In the continuum that human beings form with the environment, human beings are unwittingly or purposely left without any significant quality to distinguish them. In fact, in *Natural Born Cyborgs*, Clark suggests that ultimately, humans do not differ from cyborgs, that is, living beings enhanced by mechanical body parts. “The more seriously we take the notion of the brain-environment engagement as crucial, the less sense it makes to wonder about the relative size of each of the two contributions. What really matters is the complex reciprocal dance in which the brain tailors its activity to a technological and sociocultural environment, which—in concert with other brains—it simultaneously alters and amends.”<sup>7</sup> In Clark’s view, unlinking the original biological body parts from the notion of being human offers us significant advantages, one of which is to awaken in us an intimate and open attitude towards technology.

5 Parfit, D. *Reasons and Persons*, Oxford, Oxford University Press, 1984, 210-211.

6 Clark et al, *op cit.* 12.

7 Clark, A. *Natural-Born Cyborgs: Minds, Technologies, and the Future of Human Intelligence*, New York, Oxford University Press, 2003, 86-87.

Other authors elaborated the idea that we are our body and everything that interacts with it. The particular resonance of this idea in bioethics is of great practical importance, given the influence of bioethics on other disciplines and in the public sphere. N. Levy, one of the most famous proponents of the EMT in neuroethics, has blended this theory with transhumanism, the view that promotes the unlimited transformation of the human condition through technology. In line with Clark, Levy contemplates the individual as a set of conveniently replaceable extensional parts that looks forward to immortality. If technology can satisfy its desire for immortality in a safe and controlled way, why be afraid of change?<sup>8</sup>

Levy is not the only one who takes this views. With similar arguments, Hayles argues in *Wrestling with Transhumanism* that technology no longer represents a chance tool that serves to repair bodily damage. Technology is inherent and a co-constitutive part of human nature. "This assumption, known as technogenesis, seems to me compelling and indeed virtually irrefutable, applying not only to contemporary humans but to Homo Sapiens across the eons, shaping the species biologically, psychologically, socially and economically."<sup>9</sup> Such reasons led I. Richardson to coin the neologism *mobile technosoma* for the body. In *Itinerant Media Devices*, Richardson contends: "we have to think through other ontologies, other ways of being-in-the-world, and in a Heideggerian sense, of being-with-equipment."<sup>10</sup> Although not cited, the connection of Richardson's idea with Clark's is plain. A. Miah goes even further than many others in radically proposing that the term "human body" should be abandoned and replaced by "somatechnics". According to Miah, only by enforcing this substitution can we eradicate the common mistake of locationism, in which the human body is portrayed as the biological tissue that is beneath

the skin. Miah also blames locationism for the unjustifiable misgivings of many towards biotechnology<sup>11</sup>.

The belief that the body can be blended with technology marks an anthropological turn around. While the consequences of this turn are many, some of them are highly problematic. Let us look at those that concern the way in which the EMT addresses the first two identity conflicts in the context of neuroprosthetics. If a patient understands Clark's arguments, if not emotionally then at least rationally, he may end up believing that, given the beneficial effects of neural implants, these implants have become a full part of his identity. Nothing seems to have changed much with the patient, who remains fundamentally the same individual. However, neural implants may often alter his self-perception. And because this alteration dramatically reshapes the boundaries of personal identity, inasmuch as the patient no longer knows what kind of identity he has, Clark's arguments puzzle him.

Then, are neural implants like genuine parts of our own body? Treating them like genuine, how much do our views on personal identity need to change to fully accommodate these implants? The following sections address this uncertain issue.

#### 4. A non-deterministic interpretation of the EMT

Although lacking a unified view of consciousness, Descartes emphasized the role of consciousness in marking the boundaries of the *res cogitans*. To the extent that there must be characteristic acts of a thinking substance, some inspired by Descartes have posited that awareness and psychological continuity are the criteria that enable to us identify a rational creature. After all, self-perception, dispensed in discrete and successive states of consciousness, seems to fuel the sense of personal identity, especially when memory can store such perceptions. For example, Locke singularly stressed the link between identity and memories in this way: "As far as [a] consciousness can be extended backwards to any

8 Levy, N. *What Makes Us Moral: Crossing the boundaries of Biology*, Oxford, One World Publication, 2004.

9 Hayles, K. «Wrestling with transhumanism». In Hansell, G., Grassie, W. (eds), *Transhumanism and its critics*, Philadelphia, Metanexus Institute, 2011, 216.

10 Richardson, I. «Mobile technosoma: Some phenomenological reflections in itinerant media devices». *The Fibreculture Journal*, (2005), 6.

11 Andy Miah presented this idea on 19 April 2007 at a meeting organized by him in Sydney. The title of the meeting was *Somatechnics: Bodily (Trans)formations*.

past action or thought, so far reaches the identity of that person; it is the same self now as it was then; and it is by the same self with this present one that now reflects on it, that that action was done"<sup>12</sup>.

Considering the criterion of psychological continuity, can the EMT be compatible with the concept of personal identity that relies on psychological continuity? In light of the arguments of Clark, Levy and Miah, the answer appears to be no.

The EMT holds that mental states cannot any longer hold the key of personal identity. If mental and brain states are connected by a deterministic relation, then mental phenomena are bound to be, in some way or other, phenomena that supervene on the physical or are essentially physical. For Clarke and Chalmers, the EMT is "consistent with the view that real mental states (experiences, beliefs, desires, emotions, etc) are determined by brain states"<sup>13</sup>. Thus, any brain state X must correspond to a mental state X' that is caused or physically determined by X.

To the extent that brain states determine mental states (X determines X'), Clark and Chalmers' theory is both deterministic and locationist. Cognition takes place somewhere between the subject and the environment, not just in the brain. Thus the deterministic relation between brain states and mental states should in fact be a deterministic relation between causal chains of physical states and mental states. To argue that the cause of a particular mental state lies somewhere between the brain and the external environment (in the way in which the size of a book causes its perception) means that in principle, cognition can be physically supervenient on physical states, or rather a rare and complex kind of physical state. But in contrast with this view, we will argue that the main premises of the EMT can be read without determinism. In fact, we think that there is room for an interpretation of the EMT that is non-deterministic. And to distinguish it from the original EMT, we will call it the iEMT and present the case for a more nuanced view of cognition.

<sup>12</sup> Locke, J. *An essay concerning human understanding*, New York, Dover, 1959, 449.

<sup>13</sup> Clark et al., *op. cit.*, p. 12.

The iEMT starts by taking multiple realizability from functionalism. Multiple realizability articulates that a same mental property, state, or event can be implemented by properties, states or events that are physically different. While other accounts have drawn attention to the causal anomaly of mental states using piecemeal arguments, we will invoke multiple realizability of mental states—without endorsing a functionalist agenda—to review determinism. Suppose that neuroscientific evidence indicates that a brain state X is causally connected to a mental state X', Y with Y', Z with Z', and so on. In principle, this connection involves only that Z causes Z'. But contrary to this, evidence usually reveals that in practice, mental states can be multiply realized. Brain states X or Y could equally be linked with Z', or Z' could independently obtain in distinct subjects with varied brain states. Scenarios might be also contemplated where mental states can be linked to the activation of non-organic devices such as CNIs. These scenarios may help us to expand the physical or biological origin of mental states. Yet by broadening the causal origin of mental states, the belief that these states have a common and identifiable neural basis surely meets obstacles. The more heterogeneous the physical state with which a mental state can be linked, the harder it becomes to identify such a source so as to successfully prove, following the EMT, that the relation Z-Z' is deterministic. In fact, to broaden the causal sources of the mental suggests that any series of physical states—A, B, C to Z—will be inherently insufficient to explain Z' deterministically, because the analysis of the physical causes that take place in cognition are virtually inexhaustible. In this way, if cognition is disseminated in a large web of causes and effects, the direct connection between Z and Z' may not work as old physicalists expected it.

## 5. The contents of the mind are independent of their platform

In our view, the possibility of a multiple realization of mental states invites us to reconsider the connection between mental contents and their overarching physical expression. The EMT points out that the expression of



mental content is not limited by the brain-and-skull barrier, but seamlessly extends to environmental objects. In what we could call the “expansive” property of mental realizability, the theory holds that, if cognition can be extracted from many physical sources, mental content then seems to adapt to the multiple physical vehicles associated with cognition.

This is a familiar idea for us. The date of a possible meeting can be stored and expressed in quite heterogeneous physical vehicles; for example, it can be noted on paper, saved in an electronic device or stored in the digital cloud, thus showing that the same piece of information can be normally kept in varied physical vehicles. Strictly speaking, in the example given we should not say that the date is “contained” in any of the three physical states of the information, for this piece of information is equally stored in all of them. None of them exclusively holds the information. The date exhibits thus a one-to-many relation to them as long as the physical states endure. Further, the piece of information contained in them would be useless without a mind that can access and correctly interpret the information. Until interpreted, the piece of information is brute content that free-floats on a variety of physical vehicles, none of which exhausts the possibilities of expressing the same information in a new way, such as through the sounds of a meaningful phrase. Indeed, the only place where the date of an appointment can be said to be contained (and only metaphorically) is in the human mind inasmuch as it can make sense of it.

Note that the laws of Mendel or Beethoven’s Ninth Symphony are not found in any specific place in the world, regardless of how much they have been historically expressed on countless paper and musical scores. The possibility of expressing the laws of Mendel or Beethoven’s Ninth Symphony in apt vehicles in a one-to-many relation is virtually exhaustible. And in effect, the mind seems to have a one-to-many relation with its surrounding environment that is comparable, if not equal, to the relation between software and hardware. The same software, with identical properties and functionality, can be installed and be made to run in the same way

on computers that may have contrasting physical states. If this is true, such states will not determine *per se* or as physical the specific state of its software.

In a similar way to which the physical state of a computer seems inadequate to determine the state of its software, mental states are endemically indeterminate from a physical standpoint. The inductive link that exists between brain and mental states serves us to draw valid inferences about people’s mental states on the basis of their neural activity. While reasonable and heuristically fruitful, such predictions will fail in two ways. First, they will always be fallible. Secondly, they will be incomplete, because the mental state of a given individual cannot be limited to and be simply encapsulated in the physical state that justifies such an inference.

Thus, the iEMT is not at odds with the main tenet of the EMT: the idea that the mind is largely co-extensive with the world. This idea is still kept in the non-deterministic scenario. This way, we understand that if a book has a direct causal connection with the subject’s perception of its size, the book is just as much a part of the perceptual act as the visual cortex that enables the perception of the book. We see differences between both physical states, but they are inherently linked. However, we only see reasons for the indeterminacy of mental states. Let us see why. The states of the visual cortex that enable the perception of the book size cannot determine *per se* the content of the visual sensation, which will always remain inexhaustible as an effect of the one-to-many relation that characterizes the mental. As in F. Jackson’s knowledge argument<sup>14</sup>, with independence of the quality of our physical knowledge, the content of the subject’s perception will always be indeterminate with respect to its underlying physical embodiment.

## 6. Personal identity

If this hypothesis is true, in the discussion about the potential effects of CNI, consequences follow for the concept of personal identity. For clarity’s sake, these consequences are not related to the purpose of the CNI,

<sup>14</sup> Jackson, F., «What Mary didn’t know». *Philosophical Quarterly* 32, (1986), 127-136.

which can vary greatly, but rather, to possible mistakes in identifying what the CNI can and cannot change.

We will call weak identity (WI) the identity which can be completely modified through changes in physical states, and strong identity (SI) the identity which can only be partially changed through changes to physical states. Let us now illustrate this distinction. If the carriage of a train is so often repaired that the carriage does not keep any of its original pieces, surely it can no longer be considered the carriage that it was. This carriage has therefore a WI. What about the identity of the individual? From the perspective of the EMT, this question may have a straightforward answer. Since any individual is a functional unit made of replaceable parts, people must have a WI, and little else can be said about it.

Yet the view of the EMT is not the only possible view of human identity. Some may recall Locke's principle of psychological continuity to disprove WI. Locke asserted that the stability of one's memories could be a plausible criterion of personal identity, one that may resist the passage of time. For the subject that experiences changes to his physical and mental states over time also finds that his memories endure. Locke's criterion may invite us to accept SI instead. But again, neural implants have remarkably altered the situation described by Locke. If neural implants can modify or alter the self-perceptions, injury can mar recollections and memories to a point that is beyond repair. In this context, the criterion of psychological continuity may no longer be held, showing that WI is a better choice.

Things will look differently in the context of the iEMT. If states of mind are physically indeterminate and their expression in neural states incomplete, as we have argued, mental states cannot be identical to brain states. In other words, mental states are physically irreducible.

The idea of irreducibility has been around for years. Prominent philosophers like T. Nagel and scientists like M. Polanyi<sup>15</sup> advocated for irreducibility in their own fields. Nagel claimed that the first-person perspective is essentially irreplaceable, particularly by that of the third

person, because the objective features of the world cannot account for its subjective features. As science is done from the third-person perspective, the first-person will always fall beyond the scientific viewpoint<sup>16</sup>. In a different context, Polanyi argued that in biology, the mechanisms that generate, read and interpret DNA could not be reduced to the physical laws that predict the formation of its elements, even if such laws are accurate. If irreducibility is true, then, higher levels of reality can only be partially explained by analysis of their lower and more fundamental levels. Irreducibility operates at cross-levels involving a varied array of phenomena. There is no reason to discard the view that it applies to the psychophysical relation. In this way, we might say that physical and mental states will describe both irreducible and inseparable parts of reality, that is, parts that cannot be consistently accounted for by analysis of its lower parts.

A typical sign of this irreducibility that is often pointed out is the subjective features of the first-person perspective, which do not seem to be expressible in the brain states that research can associate with it. As Bennett and Hacker remarked, it is a mistake to attempt a reduction of the first-person perspective or the qualitative features of perception. For them, to describe neurons as acquiring or gaining consciousness when some individual awakes, even if this is merely meant as a figure of speech is a categorial mistake<sup>17</sup>. Neurons or systems responsible for awakening are a mere part of such awakening, rather than the awakening itself. Only the individual as a whole awakes, and hence can be the appropriate subject of this property. Awakening is then more than a physical act.

Mental or psychological states must then be irreducible to physical states. Their irreducibility suggests that sentient beings possess a SI. Despite the fact that sentient beings are affected by changes in the mental and the physical order, such changes cannot alter their whole identity. Fundamental reasons prevent the wholesale reduction of cognition, regardless of the way in

<sup>15</sup> Polanyi, M. «Life's irreducible structure». *Science* 160(3834), (1968), 1308-1312.

<sup>16</sup> Nagel, T. *The view from nowhere*, Oxford, Oxford University Press, 1986.

<sup>17</sup> Bennett, M.R., Hacker, P.M.S. *Philosophical foundations of neuroscience*, Oxford, Blackwell Publishing, 2003, 3

which cognition varies in synchrony with physical inputs. Of course, to say that neural change does not affect the whole identity of a sentient being, or that sentient beings have a SI means that the physical viewpoint does not exhaustively capture the nature of cognition or the fact that a subject is sentient as normally understood. If these intuitions are valid and we can be exploit them, consequences follow for the EMT. The external environment can ultimately result in changes to mental states, but changing mental states cannot form part of the personal identity of the individual in a strong sense. I am still the person who I am as I see a red spot and later on I see a black one. We could call such cognitive variations accidental to the individual because they do not fundamentally affect her. But if so, we can see no ground to argue, following Hayles, that technology is inherent and co-constitutive of nature, nor that this thesis is "virtually irrefutable"<sup>18</sup>. For the replacement of biological tissue with CNIs, no matter how disruptive, is an accidental change.

To be sure, a systematic defense of SI would require further elaboration and a more detailed discussion of its implications. For the time being, our defense of SI through irreducibility is a working hypothesis that supports common sense and has some advantages over its rivals: regardless of a subject's state of consciousness or unconsciousness, upholding SI involves that the patient with a CNI continues to be the person that he was prior to implantation. Implants are accidental to identity. Although a patient may feel the loss or impairment of cognitive functions, such a loss cannot destroy his identity, that is, in the sense that he cannot become a different person from the person he was. Any patient should rest assured about it. He will remain the same as before until the disintegration of his constitutive parts inflicted by death. Until then, endorsing SI renders impossible the change in identity.

## 7. Ethical advantages of iEMT

In the end, the iEMT can be considered a more ontologically conservative theory than the original EMT.

While the iEMT incorporates some valid intuitions of the EMT, such as the idea that the mind is co-extensive with the external world, the iEMT confronts Levy, Hayles and Miah in their attempt to dissolve the identity of the individual in the external environment and usher in what Levy calls the *web-extended mind*, because Levy's view sidelines important tenets about human identity.

Still, the iEMT has both greater therapeutic power and better deals with adverse effects. First, the iEMT preserves the patient's identity as conventionally understood: as limited, autonomous and real. Since nothing essential has changed after implantation, the patient with a CNI does not need to modify his self-understanding. For similar reasons, the subject is not obliged to think that the interplay between mind and world means that whenever he knows a different person he and the other person may become a single new causal or functional unity. Instead, the iEMT invites the patient to realize that his true identity is unaffected by the series of physical and psychological changes that a neural implant may bring about, let alone cognition. Consequently, the patient's lifestyle does not have to change either, for the relevant aspects of his world remain stable.

As we see, in arguing that human beings have a SI, the iEMT preserves their identity through change. With this, it provides a specific answer to the conflict of identity described in the first section. Surely, the so-called accidental changes that follow neural implants are no excuse to pretend that the identity of the patient is intact after therapy. Any slight change to the patient's self-perception or conduct often bring about traumatic changes of personality. But even in these cases, in which the patient may need the further support of therapy, the iEMT stresses that, while harmful or traumatic, such changes do not justify the belief that the CNI transforms the patient into an altogether different individual or device.

This should encourage a different understanding of the conflict of identity. Let us now turn to the conflict of alienation. We believe that iEMT may help the individual to address the conflict of alienation in two different ways. First, while the iEMT equates organs and

<sup>18</sup> Hayles., *op. cit.*, p. 216.

implants at the functional level, it does not equate them at the ontological level. Both organs and implants play certain functions in the organism, which may be crucial for survival and self-realization, but they do not change his identity. By the indeterminacy of the mental, there will always be features that a neural implant cannot change. Of course, this also entails that these features stay beyond our control. Ontologically, the individual stays unchanged. Secondly, the iEMT acknowledges that the patient has not lost his freedom. Here the iEMT aligns once again with common sense, which says that a neural implant maintains or increases the freedom of its user to the extent that it enhances the overall quality of his life. At this point, it is less relevant whether the patient's actions are carried out through natural or artificial means as long as he can carry them out. And even if the implant fails, disease cannot deprive patients of their inner freedom, in the most radical and profound sense of the term. The subject may lose his powers of autonomy, forget or even lose consciousness of his condition, but preserving the patient's identity we preserve the identity of a free being at a profound level.

Finally, how should we judge treatments such as DBS? DBS involves an extremely close co-existence with the implanted device and the person. As the implantation of two electrodes in the brain is quite an invasive therapy, few deny the trauma associated with some of the side effects of DBS, which includes damage to something as intimate as emotions or self-understanding. Marring them is surely worse than losing organs. However, if DBS can effectively treat incapacitating illnesses such as Parkinson's disease or severe depression, then it may be worth the loss. After all, the patient continues to be the same. And what about the lasting dependency on artificial devices? Here the issue should be less relevant. For it should be irrelevant whether one's behavior is determined by the normal functioning of the brain alone, or whether the brain regains functionality through a neuromodulatory device. What is most important is that, with the limits and fluctuations of DBS, the patient feels significantly better.

## 8. A final remark

The EMT and its common interpretation have serious theoretical and practical shortcomings. The least problematic of these is that, given that the theory predicts a deterministic relation between physical and mental states, the EMT fails to overcome a locationist approach despite Miah's claim to have overcome it. While we say nothing about the limits of the body, we reject that the mind and its states can be made sense of as physical and spatial phenomena. A more significant shortcoming of EMT is its defense of a weak model of human identity. In this model, the individual is lost within the network of interacting devices, an issue that is no less important than the several issues which the theory claims to solve.

As a result, the iEMT offers a middle-way on personal identity in an externalist context. However, unlike mainstream externalism, the iEMT does not present mental states as mere epiphenomena of a causal interplay between the subject and the environment. Rather, mental states are shown both as environment-dependent through cognition and irreducible in their own right. This independency ultimately stems from the integration of mental states in the strong identity of a sentient being, the central character that went lost in the evolving conflicts of neuroprosthetics.

## References

- Bennett, M.R., Hacker, P.M.S. *Philosophical foundations of neuroscience*, Oxford, Blackwell Publishing, 2003.
- Cash, M. «Extended cognition, personal responsibility, and relational autonomy». *Phenomenology and the Cognitive Sciences* 9 (4), (2010), 645-671.
- Clark, A. *Natural-Born Cyborgs: Minds, Technologies, and the Future of Human Intelligence*, New York, Oxford University Press, 2003.
- Clark, A, Chalmers D. «The extended mind». *Analysis* 58(1), (1998), 7-19.
- Halpern, C.H., Rick, J.H., Danish, S.F., Grossman, M., Bal-tuch, G.H. «Cognition following bilateral deep bra-

- in stimulation surgery of the subthalamic nucleus for Parkinson's disease». *Int J Geriatr Psychiatry* 24, (2009), 443-51.
- Hayles, K. «Wrestling with transhumanism». In Hansell, G., Grassie, W. (eds), *Transhumanism and its critics*, Philadelphia, Metanexus Institute, 2011.
- House, R.M. Thompson, T.L. «Psychiatric aspects of Organ Transplantation». *JAMA* 260(4), (1998), 535-539.
- Jackson, F., «What Mary didn't know». *Philosophical Quarterly* 32, (1986), 127-136.
- Levy, N. *What Makes Us Moral: Crossing the boundaries of Biology*, Oxford, One World Publication, 2004.
- Locke, J. *Essay Concerning Human Understanding, vol. 1*, Oxford, Clarendon Press, 1894.
- Locke, J. *An essay concerning human understanding*, New York, Dover, 1959.
- Nagel, T. *The view from nowhere*, Oxford, Oxford University Press, 1986.
- Parfit, D. *Reasons and Persons*, Oxford, Oxford University Press, 1984.
- Polanyi, M. «Life's irreducible structure». *Science* 160(3834), (1968), 1308-1312.
- Richardson, I. «Mobile technosoma: Some phenomenological reflections in itinerant media devices». *The Fibreculture Journal*, (2005), 6.
- Schüpbach, M., Agid, Y. «Psychosocial adjustment after deep brain stimulation in Parkinson's disease». *Nature Reviews Neurology* 4, (2008), 4: 58-59.

