FREE WILL AND THE BRAIN: ARE WE AUTOMATA?

Włodzisław Duch
Department of Informatics, Nicolaus Copernicus University, Toruń, Poland,

Google: W. Duch

1. Who am I?

Quis ego et qualis ego? Who am I and what kind of man am I? St. Augustine (400 AC) asks an ancient question, that echoes the inscription "Know thyself" over the Apollo temple in Delphi, and repeats the advice given by Solon, who wrote the first laws for Athenian democracy one thousand years earlier . Similar statements are found in the Mahabharata, the Tao Te Ching, and the Dhammapada, showing that the need for understanding oneself has been articulated from the beginning of philosophical inquiry. These timeless questions were formulated in ancient Greece in terms of *nous* and *psyche*, while modern preoccupation with self and consciousness started with Descartes in philosophy and entered the mainstream of science only near the end of the 20th century (Damasio, 1999). Although our modern conceptualization of the question may be different the essence remains the same.

2. How have such questions been answered in the past and how can we answer them now?

Early philosophy looked for simple explanations and focused on the difference between dead and living organisms. How can the growths of plants, the movement of animals, or the movement of thoughts and mental images be explained? No physical difference could be observed at the moment of death, so something immaterial must have left the body. Ancient Egyptian priests were obsessed with life after death and distinguished many elements that could explain the observed change. Essentially everything that moved (including heart, body, name, and even shadow) had separate metaphysical essences (Bunson, 1991). Many important Greek philosophers (Pythagoras, Plato, Democritus) studied in Egypt, and adopted this idea in a modified (and simplified) form. Souls had the power of originating movement (understood as any change), as Aristotle writes in his treatise De Anima (Aristotle, 350 BCE): "what has soul in it that differs from what has not, in that the former displays life … Living, that is, may mean thinking or perception or local movement and rest, or movement in the sense of nutrition, decay and growth." Aristotle distinguishes various souls: vegetative or plant soul (responsible for growth, nutrition and reproduction), an animal soul (capable of sensation, local movement, response), and rational, thinking soul, capable of knowing.

Aristotle understood matter as potentiality, because it may be formed into many actual structures, for example many pots may be formed from the same clay. For him "it is the soul which is the actuality of a certain kind of body", and it cannot be separated from the body. This conceptualization was developed further by St. Thomas in "Summa Theologica". This philosophical concept of soul is rather hard to translate into scientific terms: specific forms of organisms, determined by genetic code and adapted in the process of evolution to particular ecological niches might indeed be called “the soul” of these organisms, but this concept has no explanatory power. There is no advantage in calling complex genetic and metabolic processes of plant growth “vegetative soul”, or to summarize the patterns of animal behavior as “animal soul”.

Ancient metaphysics has no priority over modern science, because it has been based on conceptualizations derived from naïve observations. Aristotle’s reasoning was perhaps too subtle for folk psychology and religion, and thus his concept of soul become reified and used to answer the question “Who am I?” giving an illusion of understanding. I am a soul, controlling my body. The ancient view of the external world has been replaced by modern astronomy, infinitely more sophisticated. We do not believe in a flat Earth at the center of the Universe, although our direct experience favors such beliefs. While the modern view of the physical Universe is much more interesting than the ancient, the ancient view of a person has still not changed in folk psychology and in religion. The quest for a better understanding of human nature has just started and neuroscience is at its forefront. Michał Heller, a cosmologist and priest who was recently awarded the Tempelton prize for his writings on science and religion, said in an interview: “I predict that as there was a Galileo case, there is a Darwin case, sooner or later there will be a neuroscience case. If the Church does not prepare for it we shall have a crisis deeper than in Galileo’s times. It is time for the Church to train experts, otherwise theology shall remain in ancient times.” (Heller 2008, my translation).

In the last few decades cognitive neuroscience has made significant progress towards understanding the mind, and has thus become the branch of science that is best posed to answer the question “Who am I?” Francis Crick, who in his later years studied consciousness, writes in his book “Astonishing hypothesis” (1994) : ”You are nothing else but a bunch of neurons”. Joseph LeDoux, an expert in the neuroscience of emotions, makes a similar statement: “You are your synapses”. One common reaction to such simplistic statements is that if “I” am just neurons, or synapses, or a physical brain, then “I” do not really exist. “I” am then an automaton! There is no free will and our feeling of having control is only illusory. This seems to be quite contrary to our everyday experience, as Pascal in the 17^{th} century expressed it: “It is impossible that our rational part should be other than spiritual; and if any one maintains that we are simply corporeal, this would far more exclude us from the knowledge of things, there being nothing so inconceivable as to say that matter knows itself. It is impossible to imagine how it should know itself.” (Pascal, 2006).

An understanding of our own nature is the basis of ethics, politics, and education. It is ultimately the most important factor that guides human activity. How and why did modern science imagine what was impossible for Pascal? How did we come to believe that brains create minds, and how should we understand this relation? A brief history of the long struggle to understand the mind-body problem follows. The third section presents the latest developments in the neuroscience of free will. The final section discusses some of the problems solved and created by a scientific understanding of our human nature.

3. How did we come to believe that minds are products of brains?

The traditional view of who we are has been summarized by Steven Pinker in "The Blank Slate.The Modern Denial of Human Nature" (Pinker, 2002). The three main dogmas of that tradition have been identified as:

- mind is tabula rasa, or a blank slate, there are no innate traits, only environment matters, as John Locke argued and behaviorists believed;
- nature (including human nature) is good, the noble savage is corrupted by society, as J.J. Rousseau imagined;
- there is a ghost in the machine, a soul that makes free choices according to one's will, the idea expressed by Descartes.

For Descartes the bodies of animals as well as humans are just complex machines, their bones, muscles and body organs are component parts of that machine. However, since humans may use language to demonstrate that they are conscious, having mental states, humans have souls but animals do not. Animals do not feel pain, their cries are just squeaks of machinery. This doctrine created the problem of how the interaction between the soul and the body occurs, and the even more serious problem of infinite regress: if the soul provides the intention to act, where does this intention comes from? There is no intention to have intention, it just seems to pop up in our mind out of nowhere.

For hundreds of years, scientific speculation continued (Miller, 1995). Thomas Hobbes, in his treatise On Human Nature (1651) wrote: "For what is the heart but a spring; and the nerves but so many strings; and the joints but so many wheels, giving motion to the whole body". Unlike Descartes, he thought that mental functions could be explained in a purely mechanistic way. For him "will" is just a verbal label to describe attractions and aversions (motions towards and away from objects) experienced while interacting with the environment. Thinking was a motion of an internal substance in the brain, and feeling a motion of the heart. To explain visual or auditory perception at a distance he has postulated infinitely small motions of the nerves (called "endeavours") that carry the motions from senses to the brain. He tried to apply the mechanistic law of inertia to perception and imagination, viewed as a kind of meeting place for motions in the brain.

David Hartley, an English medical doctor and philosopher, founder of the associationist school of psychology, in his "Observations on Man" (1749), makes more precise observations: brain damage and neurological problems clearly lead to changes in perception and thinking. Sensory experience is caused by vibrations in the nerves, reaching the brain and causing vibrations in the "infinitesimal, medullary particles" which cause sensations and ideas. Moderate vibrations are interpreted as pleasure, while violent vibrations that damage nerves as pain. Memory, or "ideas of sensation", are traces of previous vibrations that leave in the brain tendencies to vibrate in a similar way in future. Voluntary action is explained in a deterministic way, connecting motor vibrations with vibrations coming from sensations or other mental states.

In the 19th century scientific study of the nervous system made progress and it was known that reflex automatic action involves the reflex arc in the spine that does not enter the brain, the organ of consciousness and seat of the soul. The brain, as the seat of soul, should not be associated with automata, as this could erode religious ideas, as well as public morality. Eventually, Thomas Laycock, English neurophysiologist, in the “Mind and Brain, Or, The Correlations of Consciousness and Organisation” (1860), had to admit that: “... the brain, although the organ of consciousness, was subject to the laws of reflex action, and that in this respect it did not differ from the other ganglia of the nervous system. I was led to this opinion by the general principle, that the ganglia within the cranium being a continuation of the spinal cord, must necessarily be regulated as to their reaction on external agencies by laws identical with those governing the functions of the spinal ganglia and their analogues in the lower animals.”

Soon after other researchers, come up with even stronger statements: Ivan Sechenov, in his “Brain Reflexes” (Refleksy golovnago mozga ,1866), claimed that all conscious and unconscious acts are reflexes in terms of their structure. This was considered subversive to public morals and social order, and Sechenov was accused by the tsarist government in St. Petersburg of spreading materialism and “debasing Christian morality", but he was not persecuted. Prevailing morality standards had a very strong influence on medical research. In 1873 Sir John Ericksen, Surgeon Extraordinary to Queen Victoria, stated: “The abdomen, the chest, and the brain will forever be shut from the intrusion of the wise and humane surgeon”.

Thomas Huxley in his “On the Hypothesis that Animals are Automata, and its History” (1874), wrote: “... the feeling we call volition is not the cause of a voluntary act, but the symbol of that state of the brain which is the immediate cause of that act. We are conscious automata ... “. Benjamin Carpenter (1874) analyzed the experiments of James Braid with hypnosis and concluded that the perceptual system almost completely operates outside of conscious awareness, and mechanisms of thought also operate largely outside awareness. These unconscious automatisms were largely forgotten until Freud came with his id, ego and superego ideas, that may be roughly mapped to the triune brain of MacLean (brain stem, limbic system, cortex).

At the end of 19^{th} and early 20th century William James, Wilhelm Wundt and other early psychologists studied conscious processes via introspection while Sigmund Freud and his followers studied unconscious conflicts they held responsible for psychological problems. The results were so confusing and hard to test empirically that psychology turned to behaviorism, concerned only with the study of observable behavior. The study of mental processes, internal representations and consciousness became taboo, because it was believed that as these concepts did not describe psychological mechanisms but only abstract constructs that were used in a similar way as soul and spirit, giving the illusion of explanation without referring to specific mechanisms (Skinner 1938).

How then should one try to understand the mind? British philosopher Gilbert Ryle (1949) agreed to some degree with the behaviorists, stressing that mental events are higher level processes that we interpret as bodily feelings or states. Mental processes are a series of acts and dispositions to act that are perceived as emotions, feelings, intentions or imagery. Mind is a process, a succession of brain states, not a thing that is inside the brain. There is no ghost in the machine.

The cognitive revolution in psychology - the shift from the study of the theories dealing only in observable phenomena to theories dealing with mental processes that were not directly observable - started with Noam Chomsky. He pointed out that behavioral principles are not sufficient to account for language, and in particular rapid language acquisition by children. Cognitive psychology postulated instead many internal mechanisms for behaviours (such as beliefs and goals) that could not be directly observed. With the development of neuroimaging techniques and cognitive neuroscience it turned out that many earlier psychological constructs were wrong. Conceptualizations of brain processes such as working memory or attention arise from the lower levels of complex neural network dynamics without the need to postulate specific mechanisms. It seems quite likely (although there is yet no agreement) that the phenomenon of consciousness has a similar explanation (Duch 2005).

The battle for a special status of the mind as non-physical entity has been entirely lost. One of the last attempts was by Karl Popper and John Eccles in "The Self and Its Brain" (1977), followed by Eccles' book "How the Self Controls Its Brain" (1994). Eccles' attempt to justify the Cartesian dualism has been clearly motivated by religious belief. This and other approaches that invoke exotic theories (including quantum mechanics) to talk about "mind" and "consciousness", without defining what exactly they try to explain (Stapp 1993; Penrose 1989, 1994), are not credible when challenged to be consistent with evidence from many related fields of research, such as brain imaging, psychological experiments, brain lesions and numerous neuropsychological syndromes. Notwithstanding the progress in the sciences of the mind, the illusion that there is a "ghost in the machine" or homunculus inside the head is very strong. The development of computational models of brain functions is contributing to the understanding of mental decision processes, giving inspiration to the new engineering disciplines of cognitive robotics and neurocognitive informatics, whose ambitious goal is the creation of neuromorphic electronic systems and computational models that can perform the same functions as mind, providing explanatory mechanisms for human and animal behavior.

4. Free will, or the illusion of conscious choice.

There is an obvious gap between psychology and brain science, between what is called the first person view (subjective experience) and the third person view. How can physical neural impulses be sufficient to create subjective mental events? This is a central theme in philosophy of mind (Chalmers, 1996, Dennett, 1996). Without convincing theory showing how the mind arises from brain activity it will be very hard to accept this as a fact. Cognitive science started as an incoherent mixture of various the branches of science that are related to cognition, and has not yet completed the search for central, unifying theory. Brain states are observed and analyzed and the transitions between them lead to finite mechanistic models of behavior, in translation losing the subjective mind. Is a satisfactory understanding of the mind from neuroscience perspective possible at all? At which level of description should we look for it? At the smallest scale neurons are simply counting impulses. Large groups of neurons (almost 100,000 in a square millimeter of cortex) form mesoscopic networks, implementing sensory and motor maps, creating self-organizing filters for analysis of sensory information (for example, phonemes), directing motor behavior using population coding. At the scale of several centimeters, neural networks in the cortex and subcortical structures perform quite complex functions, some quite specialized, such as the face fusiform area performing face recognition. In most cases, neural networks are involved in integration of information from various sensory modalities, with some contributing to many processes. At the whole brain level, goal-directed intentional behavior, language, thinking, reasoning and problem solving becomes possible.

Roger Shepard (1987) tried to deduce universal laws of generalization for the psychological sciences, looking for simplified models for the transformation of stimuli into perception. In his view: "What is required is not more data or more refined data but a different conception of the problem". The neuroscience credo is: mind is what the brain does, a potentially conscious subset of brain processes. How can my inner life result from the counting of impulses by neurons? How to relate brain states to mind states? This is what to some degree is done by the brain-computer interfaces research community: brain activity, usually derived from EEG signals, is correlated with mind-related signals: intentions to move in some direction, or to select a letter or an object on the screen. Can one read precise intentions from the activity of the brain? Frontal lobes create competing plans for various actions, and the final decision is made by networks in a winner-takes-all process, the winning plan inhibits the activity of networks that created alternative plans. In the experiment reported by Hayens et al. (2007) subjects were asked to decide whether they will add or subtract two numbers. After a variable delay during which they had to maintain their intention the two numbers appeared on the screen, followed by 4 numbers: the result of addition, subtraction and two incorrect answers. Subjects then pointed to the result they obtained by performing the chosen arithmetic operation, thereby revealing their intention. Activity of the medial and lateral prefrontal cortex during the delay period measured by fMRI consistently revealed their intentions. Similar brain imaging techniques may be used to detect lies or covert attitudes, raising concerns about the privacy of personal thoughts (Haynes, Rees, 2006).

Decisions result from plans that are created by the brain, therefore it is interesting to ask at which point do we become conscious of these decisions? It is rather obvious that a specific and relatively strong signal in the brain should appear before the feeling of intention and understanding of what actions will be taken may arise. In a long series of experiments started in the 1970s, Benjamin Libet asked subjects to act whenever they felt like it in a certain time frame, and then press a button or flex a finger. The precise moment of action was measured as well as the moment when the will to act was indicated by the subjects by noting the time. The time difference between action and intention was about 200 milliseconds. During these experiments EEG signals from the secondary motor cortex were measured and it was found that the intention to act could be reliably predicted to occur 300 ms before the feeling "I want to press a button" arises, and thus 500 ms before the actual movement (Libet, 2000). Thus consciousness of decision has been preceded by unconscious brain activity preparing for the action. An external observer of brain activity can know a fraction of second earlier when and what decision will be made than the internal brain processes of consciousness. Moreover, in a series of new experiments Lau and his colleagues (Lau and Passingham, 2007; Lau et al. 2007) showed that applying a magnetic stimulation (rTMS, repetitive Transcranial Magnetic Stimulation) to the pre-supplementary motor area (pre-SMA) <u>after</u> the execution of spontaneous movement shifts the perceived onset of the motor <u>intention</u> backward in time, and shifted the perceived time of action <u>execution</u> forward in time. "We conclude that the perceived onset of intention depends, at least in part, on neural activity that takes place after the execution of action." (Lau et al. 2007).

Initial plans for action are obviously not formed in the motor cortex, but earlier, in the prefrontal cortex. Soon et al. (2008) found that subjects were conscious of their decisions (pressing a button with left or right index finger) about one second before actual action. However, plans to act could be inferred from the activity of the prefrontal and parietal cortex up to 10 sec. before they become aware of their decisions. The neuroimaging techniques used in such experiments show only grossly averaged neural activity, and there is no doubt that more precise information could be obtained from electrodes placed directly on the brain surface or connected to neurons. These experiments leave no doubt that decisions result from unconscious processes that have to reach sufficient level of intensity in the brain to be noticed (and thus experienced) by other parts of the brain.

Daniel Wegner in “The Illusion of Conscious Will” (Wegner, 2002) analyzed various situations in which people’s perception of their actions and intentions does not coincide with reality. In some situations we may be acting but be convinced that we are not the acting force, for example pushing the ouija board, moving the dowsing rod searching for water, in hypnotism or facilitated communication. In other situations we are not acting, but think that we are: subjects may be induced to believe that they have performed some actions, or that their actions are achieving far more than they in fact are. Wegner concludes provocatively, that conscious acts of will are never the direct causes of our actions, instead, both conscious willing and action are the concomitant effects of a common unconscious cause. In experiments with magnetic stimulations of the prefrontal cortex subjects had to press a button with their left or right hand, and even if magnetic stimulation influenced them to select one side 80% of the time they were convinced that their choices are free. This finding raises concern that people could in principle be controlled remotely without even noticing that something has influence on their will.

In summary, will seems to be the feeling resulting from attention to the state of several cortical areas that are active when the brain is ready to act. “Attention” here simply means highly synchronized activity that dominates the competition for access to the working memory. Awareness of intentional acts correlates with activity of specific cortical areas (Farrer and Frith, 2002, Jackson et al, 2005). Moreover, some cortical areas react to the pain experienced by others because the same neurons receive multi-modal inputs from internal motor as well as external auditory and visual subsystems. This is the basis for empathy, compassion, and thus natural morality.

Many more experiments have been done in recent years, leading to a more coherent picture of the processes that the brain is using to make decisions that result in feelings of empathy, compassion and natural morality. Neuroscience leaves no doubt that the ancient picture of the ghost in the machine is untenable. What we can now say, at least in a metaphorical sense, is that mind is but a shadow of neurodynamics. It is in this sense, that very few brain processes are sufficiently strong to cast a shadow upon those higher level areas of the brain that can add linguistic comments to brain states, explaining why it is so hard to describe feelings. In time, computer simulations may provide a way to test increasingly complex models of the dynamics of brain states to form a more complete picture of mind events (Duch, 2005). Given what we now know, what are the societal implications of this emergent understanding of the mind?

5. Implications.

The Earth seems to be flat, and yet we know that it is round; matter seems to be solid and yet we know that it is mostly empty space. The mind seems to be in control, and yet we know now that this is only an illusion. One reason why this illusion has been so dear to us is that it has been the basis for personal responsibility. Thomas Jefferson wrote the “Declaration of Independence” (1776), starting with: “We hold these truths to be self-evident, that all men are created equal …”. This statement was meant as a negation of the divine rights of kings and aristocrats, but it has been greatly misunderstood. Obviously men are not equal in the physical sense, nor in the sense of their talents, and also not in their moral sense. Empathy and the ability to feel compassion is a result of complex brain processes, and in their absence due to the developmental problems or brain damage antisocial or even psychopathic behavior may result.

Over half a million people were murdered in the year 2000 around the globe[1]. How do we explain violent behaviour on this scale? In the extreme case people are not held responsible for their actions because they are insane. The first such legal case dates to 1843, when Daniel M'Naghten attempted to assassinate the British prime minister and was found to be insane. In the USA the insanity plea requires that the accused does not know, because of mental illness, that he did wrong. Can people be simply divided into normal and insane, or is it a matter of degree? Violent behavior may result from genetic mutations that lead to abnormal brains, and thus predispose people to murder. An overactive amygdala is correlated with increased aggression, and underactive frontal lobes with the reduced ability to inhibit aggression. Do such people have sufficiently strong will to control themselves? What if their brain initiates action before conscious recognition of their intentions occurs? Do they recognize what they intend to do and are they able to stop it? We all do things that we regret later, and it is not hard to imagine that .

The prefrontal cortex makes us moral and rational; damage to this part of the brain leads to acquired sociopathy, manifested by impulsive criminal behaviour. Damage to the amygdala leads to poor empathy and low fear, typical of psychopathic emotionless criminals. It is estimated that about 25% of all imprisoned in the USA fall in these two categories, frequently due to birth complication, trauma or brain injuries and lesions. Should the courts take genetic/brain anatomy information as excuses? These are some of the questions that neurolaw experts are considering. Michael Gazzaniga, a cognitive neuroscience pioneer, in “The Ethical Brain” (1998) book writes that since the physical world is determined brains must also be determined. The interpreter function in the dominant temporal lobe creates a narrative comment on our perceptions and actions, making an illusion that we are in charge.

Gazzaniga sees brains as automata, automatic, rule-governed, determined devices but tries to reconcile it with the personal responsibility. People are free to make decisions and personally responsible because personal responsibility is a public concept, something that exists outside of the brain, in our relationships and interactions with other automatic brains. This is a very strange solution – what kinds of brain are able to obey rules and thus be responsible? The Law and Neuroscience Project[2] sponsored by the MacArthur Foundation, tries to address complex issues that neuroscience raises for legal systems. Experts working on this interdisciplinary project investigate the use of neuroscience-based arguments in le-

---

[1] http://en.wikipedia.org/wiki/Murder#Demographics

[2] http://www.lawandneuroscienceproject.org/

gal reasoning, psychopathy, drug addiction, prediction of criminal behavior and individual responsibility. This research may be considered as a part of general bioethics, and more precisely neuroethics, focused on issues related to moral reasoning and personal responsibility (Mobbs et. al, 2007). Some results appear in the journal “Neuroethics”, and the activity of the project has been summarized by Gazzaniga (2008).

The “nature versus nurture” discussion and the development of behavioral genetics has increased our understanding of the complex factors responsible for human development and behavior. Genetic determinacy has been widely discussed and although in most cases it leaves considerable room for freedom of choice it is also clear that dysfunctional genes may be responsible for serious mental problems that impair the ability for self-control and moral reasoning. The notion of neural determinism should also be considered: even if the brain has normal structure thinking processes which are guided by neural pathways that determine behavioral responses. Brain-washing leads to real physical changes in the way the brain processes information. Popular culture glorifies bad characters that are able to reject their evil ways, able to show empathy (Lord Vader in Star Wars). People with acquired sociopathy, because their natural empathy is only masked by brainwashing still have a chance to empathize. However, it would be unrealistic to expect empathy from congenital psychopaths. This is obviously not a binary, yes/no condition, but a continuous multidimensional spectrum, from zero to some maximum, in respect to the ability for self-control, experiencing empathy, or strong violent desire. As a result some people commit serious crimes without any hesitation, and other people take a long time deliberating on trivial decisions. Moral behavior may be easy for some brains, but quite impossible for others.

How then should personal responsibility be understood in view of genetic and neural determinism? Are we really capable of free actions? The traditional view cannot be maintained: there is no scientific evidence for the concept of “self”, or ghost in the machine pulling the strings. Some skeptics still hope that the problem will simply go away, that neuroscience is immature and may be wrong, that better understanding of the brain will bring back conscious control, show how mind controls matter. There is fanciful speculation about chaotic processes combined with quantum mechanics. Chaotic automata are determined but unpredictable; the brain in many states may indeed be quite unstable against small perturbations. Could it then be controlled by immaterial mind? No progress has been made in supporting these ideas in several decades. The perception of self is one of many processes that the brain is implementing; consciousness is experienced for only a small subset of brain processes, and the brain’s plans for actions may become conscious only in their final stage. The final selection process that we may become conscious of is the result of neural competition between several patterns of excitations.

If “I”, or self, is not in control of our choices then how can it be held responsible? The brain is a physical device, so a separate “I” does not exist, it is only an illusion. Some brain activities simply manifest themselves as “my decisions”. “I” am not really responsible, my brain made me do it! As in the case of the “alien hand syndrome”, that does things that “I” cannot control, in some brain states the ability to control one’s action may be reduced (as in hypnosis), and in some dissociation disorders the feeling of control may be completely gone. Is this feeling too only an illusion? To what degree and in what sense can we speak of free choices and decisions, and be held responsible for making them? Some lawyers motivated by the deterministic understanding of behavior have started to question the foundation of the whole legal system (Gazzaniga, 2008). How should we understand relations between the mind ie my subjective sense of myself and the brain?

A basic understanding of these relations was intuitively captured by Aristotle in his notions of potentiality and actuality, or matter and form. We can likewise say the brain provides only the potentiality, that its neural matter is only the substrate of all mental processes. Mind, then is a specific form of the actuality that results from the interactions within this matter. Our brains contain whole evolutionary history (phylogenesis) as well as personal histories that have shaped individual neural pathways. Minds recreate the relations between past events and have the power to create novel formulations, but memories and relations are abstractions and do not belong to the material world. Patterns of brain activity that arise when memories are recalled are only pointers to past events and follow the logic of these events. The stream of brain activations has sense only at this abstract, mind level.

Brains are responsible for all decisions and actions, and in doing so take into account the complex goals and needs that each brain has. The flow of information within the brain is not perfect and thus not all actions may be imagined and evaluated internally. Those that can create brain states that can be reported (both internally and in some form externally), may allow these states (or thoughts) to be reflected upon in a conscious way. In many cases studied by neuropsychology this lack of integration of brain information flow is evident, for example in the split brain patients, split personality cases, or neglect syndromes. In everyday life it is seen in "stronger than me" drives, or the need to "learn about oneself".

The human brain adapted quickly to relatively recent cultural inventions such as reading, writing and arithmetic, with specific cortical areas being devoted to these functions (Dehaene and Cohen, 2007). A change of the information flow in the brain was required for reading (the mapping of graphemes to phonemes), and even to understand text without reading it aloud. Reading aloud, hearing and understanding requires relatively straightforward mapping, but understanding without hearing it spoken took many thousands of years. Silent reading was enabled by the use of spaces between words that developed between 12-14th century (Saenger, 1997). However, just when self-consciousness emerged as a trait of the brain is in debate. Julian Jaynes, a psychologist at Princeton, proposed "the bicameral mind" hypothesis claiming that until about 3000 years ago, people had no subjective consciousness, the capacity for self-awareness that we experience as consciousness. They were not conscious of their own thought processes. His argument is based on the lack of evidence in such early texts as the Iliad, Odyssey and Bible of introspection or self awareness. Thirteenth century medieval illuminations still show ancient authors being dictated to by divines. Although Jaynes idea of consciousness as a neurological adaptation to social complexity is still speculative there is no doubt that we are not aware of most high-level processes, and can even find ourselves suddenly humming on whistling, or acting and trying to rationalize that behaviour later. Learning about ourselves means learning about the whole brain, a task that cannot be fully accomplished because much thinking involves weakly activated patterns that may contribute to the final, conscious outcome, but cannot be directly experienced. Metaphorically speaking, an ocean of mist extends between the peaks of conscious experiences.

The solution to the personal responsibility problem may be rather simple. We should identify with the whole person, not just the selected part of the brain that provides an imperfect model of "I", or the ego. To give a trivial example, I am supposed to make a phone call at a specific time. What if I am too busy and other processes in the brain prevent the recall? Are the limitations of the brain a good excuse for my bad behavior? No. If the matter is really important I should foresee possible distractions and arrange appropriate actions to remind me at a proper time of what should be done. I am responsible for the error and should be punished to help me to avoid it in future. In some cultures punishment for relatively small errors has been very severe (sumo wrestling judges still carry a dagger, a reminder that they should be ready to commit suicide if they make an error). Unless I am totally absent-minded, irresponsible and not able to control my actions the rules of socially acceptable behavior help to correct it. Most people have inflated opinions about themselves, but we hold them responsible for their behavior, not for what they think about themselves, or for their good intentions.

6. Conclusions.

Understanding ourselves is in a sense the final frontier, the dream of humanity that took superficial, pseudo-answers for granted since the beginning of civilization. As modern astronomy has completely changed and expanded the ancient world view so does neuroscience now. The change is even more radical, as it concerns our deeply entrenched illusions. That “I” am in charge and “I” direct all body activity is as obvious as the fact that the Earth is flat. Any attempt to dethrone the I has to meet with great fear. Yet there is no doubt that this I has access to only a small subset of brain processes and does not always know what is going on, contributing to the ineffectiveness of various forms of psychotherapy. Developmental psychology shows how the ideas about our own mind and mind’s of others develop and how babies learn to define themselves. “I” has an important function, it is just a model that the brain has created to help our total being function in a complex social and cultural environment. The self is a complex subset of all mental functions.

So, are we automata unaware that we are controlled by deterministic brains? Not quite. Brains are far too complex to be predictable. Being capable of self-modification the brain adapts constantly, changing its structure when exposed to an open and unpredictable environment. Typical automata work according to some program, and are thus more instruction-driven than data driven. Brains are just material substrates upon which the mind functions. We should not mistake the experience of a great movie or a computer game with the arrangement of atoms in plastic on a DVD disk. Physical substrate may restrict the capacity and the quality of the recording, but the game itself, expressed in a proper environment, will interact with its users activating its own characters and their behaviors. If it could learn and modify itself through interactions with the players, the game will become a unique reflection of these players and of their environment. Although we may someday build such brain-like systems they will significantly differ from automata that we have today and should not be treated as such. Even now an Aibo dog or other pet robot that has learned through months of interactions with humans acquires a special value to its owner. Erasing the program that controls such a robot will be felt as a great loss, as this program reflects human intentions and efforts. A baby robot, with much greater capacity to learn than the pet robots we have today, will have much greater value to people involved in its development.

The knowledge gained by all of our cognitive sciences requires longer philosophical reflection before we shall learn to understand it in a proper way. In fact it is full of paradoxes: although free will and the “I” may be seen as illusion, giving up these illusions means we shall gain more than we lose. There are real physical processes in the brain, body and environment that create these illusions, so in a sense the whole brain, and the whole world, can contribute to an enlarged self, containing part of everyone I have ever interacted with. Buddhist tradition has seen ego as an illusion since its beginning 2500 years ago. My egoistic I is a set of ideas about myself and the world I know, but my enlarged I has no boundaries, as everything has direct or indirect connection that forms it.

Deeper understanding of who we are brings more responsibility, not less. We are responsible for who we become, for our own development, especially for the development of our minds, and our talents. The mind has to educate itself and to "know oneself" better. We are responsible for the development of our children, setting out examples and role models for them. And we are responsible for the environment they will live in. It is this understanding that will hopefully change the world.

Self development is our moral obligation. Albert Einstein, in a letter to a rabbi in 1950 (quoted in H. Eves, Mathematical Circles Adieu. Boston 1977), expressed it in a beautiful way: "A human being is a part of the whole called by us 'the universe,' a part limited in time and space. He experiences himself, his thoughts and feelings, as something separate from the rest – a kind of optical illusion of consciousness. This delusion is a kind of prison for us, restricting us to our personal desires and affection for a few persons nearest to us. Our task must be to free ourselves from this prison by widening the circle of understanding and compassion to embrace all living creatures and the whole of nature in its beauty. Nobody is able to achieve this completely, but the striving for such achievement is in itself a part of the liberation and a foundation for inner security."

Acknowledgement: I am most grateful to Ed Brezina for critical reading and numerous improvements of the initial version of this paper.

References

Aristotle, De Anima (350 B.C), Translated by J. A. Smith, the Internet Classics Archive.

Bunson M.R. (2002). *Encyclopedia of Ancient Egypt*, Facts on File, New York.

Chalmers, D. J. (1996). *The conscious mind: In search of a fundamental theory*. Oxford: Oxford University Press.

Dehaene S, Cohen L. (2007). Cultural recycling of cortical maps. Neuron, 56(2):384-398.

Damasio A.S. (2000). *The feeling of what happens*. Harvest Books.

Dennett, D. C. (1991). *Consciousness explained*. Boston: Little-Brown.

Duch W. (1999). *Soul and spirit, or prehistory of cognitive science* (in Polish). Kognitywistyka i Media w Edukacji 1, 7-38.

Duch, W. (1997). Platonic model of mind as an approximation to neurodynamics. In: S. Amari and N. Kasabov (Eds.), *Brain-like computing and intelligent information systems* (pp. 491–512). Singapore: Springer.

Duch, W. (2005). Brain-inspired conscious computing architecture. Journal of Mind and Behavior, Vol. 26(1-2), 1-22.

Eccles, J. C. (1994). *How the Self Controls Its Brain*. Berlin/New York: Springer.

Farrer & Frith, Experiencing Oneself vs Another Person as Being the Cause of an Action: The Neural Correlates of the Experience of Agency, Neuroimage 15, 596, 2002.

Gazzaniga M, The Law and Neuroscience. Neuron 60(3), 412-415, 2008. Jackson et .al, How do we perceive the pain of others? (2005)

Haynes J.-D. & Rees, G. Decoding mental states from brain activity in humans. Nature Rev. Neurosci. 7, 523–534 (2006)

Haynes, J.-D., Sakai, K., Rees, G., Gilbert, S., Frith, C., & Passingham, R. E. (2007). Reading hidden intentions in the human brain. Current Biology, 17(4), 323-328.

Heller M. (2008), Interview (in Polish), Tygodnik Powszechny, 8.07.2008.

Jaynes, J. (1976) The origin of consciousness in the breakdown of the bicameral mind. Boston: Houghton Mifflin.

Lau HC, Passingham RE (2007) Unconscious Activation of the Cognitive Control System in the Human Prefrontal Cortex Journal of Neuroscience, 27:5805–5811

Lau HC, Rogers RD, Passingham RE (2007) Manipulating the Experienced Onset of Intention after Action. Execution Journal of Cognitive Neuroscience, 19: 81-90

Libet B. et al. The Volitional Brain: Towards a Neuroscience of Free Will (2000).

Miller, J. (1995). Going unconscious. In R.B. Silvers (Ed.), *Hidden histories of science*. New York: New York Review.

Mobbs D, Lau HC, Jones OD, Frith CD, Law, Responsibility, and the Brain. PLoS Biol 5(4): e103 (2007) Gazzaniga M, The Ethical Brain (1998).

Pascal B (2006). *Pensées*, Project Gutenberg EBook, http://www.gutenberg.org/etext/18269

Penrose R. (1989). The Emperor's new mind (Oxford Univ. Press 1989);

Penrose R. (1994). In the Shadow of the Mind (Oxford Univ. Press 1994)

Pinker, S. (2002). The Blank Slate: The Modern Denial of Human Nature, Penguin Putnam

Popper K, Eccles J.C. (1977). The Self and Its Brain, Berlin: Springer.

Ryle G. (1949). The concept of mind, Univ. of Chicago Press.

Saenger P, Space Between Words: The Origins of Silent Reading. Stanford Uni. Press, 1997.

Shepard, R. N. (1987). Toward a Universal Law of Generalization for Psychological Science. Science, 237:1317-23.

Skinner B.F. (1938). The Behavior of Organisms: An Experimental Analysis Of Behavior.

Soon, C.S, Brass M, Hans-Jochen H, Haynes J-D, Unconscious determinants of free decisions in the human brain. Nature Neuroscience 11, 543-545 (2008)

Stapp H.P, (1993). Mind, matter and quantum mechanics. Springer Verlag, Heidelberg.

Velmans, M. 1991. Is human information-processing conscious? Behavioral and Brain Sciences 14:651-69

Wegner DM, The illusion of conscious will. MIT Press (2002).

Book Chapter, quote as:

W. Duch, Free Will and the Brain: Are we Automata? Ethics and Science for the Environment, 3rd European Forum. Ed. M.Jaskula and B.Buszewski, Societas Humboldtiana Polonorum, 2011, pp. 155-170.