Motivated Emotional Mind

fragments of the book by Wiesław Galus and Janusz Starzyk

Reductive Model of the Conscious Mind

Table of contents

Can we understand the wealth of manifestations of our spiritual life, explain it through the brain processes, and even claim that it is very simple? Let's accept this challenge together. We aren't saying that we have a ready explanation of all processes that are taking place in our brain. Many details are still covered by the impenetrable darkness of our ignorance and the fog obscuring the unprecedented complexity of the brain structures. However, the basic blocks that built our consciousness already exist. The model presented in this book allows us to grasp what has seemed inconceivable so far. We still will be unable to fix the brain or cure mental disorders and neurological damage. We still can't build an artificial brain that exhibits consciousness. But we should understand how the mind can work and how an extraordinary sense of presence in the world can arise—the awareness of the events around us and that we have a place in this world. We should understand that there is nothing mystical about the action of the mind that could prevent us from exploring and imitating it. The effective theoretical model presented in this book may be the beginning of attempts to build a material model imitating mental functions, which is equivalent to building an artificial brain that is the seat of the conscious mind.

In this book, we will attempt to explain how the most important mental states can arise and how our own personal, subjective sense of consciousness can be created. Based on the latest achievements of neuroscience, biophysics, and cognitive science, we want to show that the brain processes that have already been discovered and researched, can lead to conscious feelings, conscious actions, and self-consciousness.

Part I Motivated Emotional Mind

The Eternal Question: What Is Consciousness?

Janusz recalls his early thinking about the essence of self: "I remember that as a young boy, I was thinking about my existence, and I remember the terror that seized me at the time when I was thinking that I would die one day. It seemed inconceivable to me that I would not be here anymore. It didn't agree with the feeling that I existed independently of my material body. Somehow it didn't disturb me that I was not always here, or that I was practically disappearing when I was sleeping. After all, I wake up and I am still myself. But death seemed something terrifying—that I would disappear. But what does it mean? Where will I be? It was so terrifying that for many years I didn't allow myself such thoughts. Much later and for a different reason, I began to think about consciousness and mechanisms of its formation."

Most of us have feelings all the time of consciousness, this special state of awareness of and participation during surrounding events. We have the ability to feel the world with our senses, but also the ability to feel and experience our own internal mental states. People have wondered what this amazing property is that allows us to feel, think, analyze, perceive, and understand what is happening to us and what surrounds us. Many of us are satisfied with the simple, intuitive explanation that in addition to the material body, skin, organs, body fluids, and the brain, we have a mind in which psychological processes, also called mental processes, occur. We know that the brain is the habitat of these processes. We are fascinated by its cosmic complexity and sophisticated construction and by the subtleties of our mental states. We are learning more and more about the structure and functions of the brain thanks to the fantastic advances in brain-imaging techniques. However, many specialists studying the brain—neurologists, psychologists, and psychiatrists, as well as philosophers and biophysicists—are convinced that when it comes to how the material brain creates mental states that reflect the phenomena surrounding us, how it creates our "thoughts" and how in turn our thoughts can move any matter, we can't say much.

First, consciousness may concern the awareness of what we feel. In the state of consciousness, we feel many emotions as a reaction to our experiences. Our bodies and minds have many needs that we must meet if we want to continue to exist. Unmet needs cause feelings and emotions that strengthen the reactions of the body. We also have innate feelings and emotions that have evolved in response to constant cravings and threats. Our bodies provide us with many feelings, as well as signals of pain and pleasure. We can feel pain as a result of the pricking, cutting, or tearing of tissues. However, this feeling of pain is obviously linked with our consciousness, because when we lose consciousness due to blackouts, a strong hit to the head, brain hypoxia, or the effects of alcohol, narcotics, or psychotropic drugs—we don't feel the slightest pain. Even in a dream, the feeling of pain is subdued. If the intensity of pain gradually increases while we are dreaming, at first we don't feel it at all. Only when strong pain awakens us do we start to feel it. Similarly, we can perceive all other sensory impressions only when we are conscious and aware.

Secondly, consciousness concerns what we perceive with the help of our senses. In the state of consciousness, our senses are flooded with a steady stream of sensory impressions, not all of which are realized. Looking at the surroundings and the scene of images in front of our eyes, we observe only some of them, and only these few are consciously perceived. This type of consciousness of feeling sensory stimuli is called sensual consciousness, also called **perceptual consciousness** (from the word perceive). How can we detect it in animals or artificially built mechanisms? Of course we will see it easily when we observe an affective reaction to stimuli. The vast majority of animals exhibit this reaction. Yet it is hard to claim that jellyfish, night crawlers, or primitive worms have conscious states, although even the latter two have the germ of a brain.

Let the nematode *Caenorhabditis elegans* be an example. It is a small worm that feeds on bacteria and lives up to twenty days. Its body consists of exactly 959 cells, of which 302 cells are neurons (383 neurons in males). The brain of *C. elegans*, arranged in a ring around the throat of the nematode, contains eight thousand synapses. (In comparison, the human brain has 86 billion neurons and roughly 100 trillion synapses.) Despite such a simple structure, the worm's brain contains neurons responsible for touch, smell, taste, and temperature. This worm has many genes very similar to human genes, as evidenced by fact that it functions perfectly when a fragment of its own DNA is replaced with human's. We managed to reproduce the exact map of nerve cell connections in the "elegant" brain. Thanks to this and full knowledge of *C. elegans*'s genetic code, one can predict exactly how the cells of its body, including neurons, will change in the course of their personal development, and how it will respond to any stimuli. One can see that it responds to stimuli quite automatically, although it has the ability to "learn" to a minimum extent. Without any major problems, one can build an artificial nematode brain with the same characteristics and abilities. Does the nematode have conscious feelings? And if yes, will its mechanical copies also sense touch and smell? It seems that perceptual consciousness requires something more. The automatic reactions of jellyfish, earthworms, and nematodes or existing artificial machines don't allow them to be granted the status of conscious beings. Why? We will write about it in later chapters.

The world of higher-organized beings is more complicated. People and a large number of animals have an idea of the space where they move. They judge not only distances but also spatial relationships between objects in this space. This corresponds to the creation of an environmental model in the mind. Primitive creatures don't create this model. A moth will circle the candle hundreds of times before it dies in its flame. The moth determines the direction of its flight at a constant angle in relation to the light source. This way of moving is effective when the distant moon is the light source. But the same rule applied to a candle or the bulb of a nearby lamppost inevitably leads to a tragic end.

Similarly, a fly unsuccessfully attacks the glass of a half-opened window, not supposing that other half of the window might be open. Meanwhile, cats and dogs flawlessly find their way to the spare entrance or window of a building when the main entrance door is closed. The wolf pack collaborates well, hunting for animals and dividing the roles to cut off the path of the running deer and prevent its escape. The falcon can fly through a flat obstacle like a wall made of boards or a thicket of branches, even when the hole in the wall or space between the branches has dimensions coinciding exactly with the cross-sectional area of the falcon's body in flight and with folded wings. The bird develop speed of up to fifty kilometers per hour, which, with the slightest error, can lead to tragic consequences. In these behaviors, a two- or three-dimensional plan of the space in which the tasks are executed is very useful. These animals and people must have and use such a plan, whether it's created on the spot or remembered from previous observations. They must be aware of space around them and have at least the seeds of consciousness of the scene geometry of activities and objects in it. Therefore, this specific type of consciousness is sometimes called geometrical or **spatial consciousness** (Matei et al. 2007).

A much bigger problem is the perception of own person and one's role in the world, and the perception of the effects of one's own actions, preferably a perception that predicts these effects and draws constructive conclusions that allow for making optimal or even suboptimal decisions. Creating not only a geometric plan of the environment but a plan of action requires the use of a wide range of abstract concepts to define objects and describe planned activities. This in turn requires experience acquired during learning and a significant amount of knowledge. A bigger brain isn't enough. Therefore, even very intelligent animals and people begin to perceive their own person, see their dependency on parents and tribesmen or members of the herd, and plan and predict the effects of their own actions, after a certain period of development, education, and upbringing. Only then do they perceive the similarity of beings and objects to those previously known, and learn classify them not as single instances but as groups—for example, blocks, balls, apples, and stools. The names of categories of objects gradually generalized to bigger groups like toys, food, and furniture are more and more abstract concepts symbolizing whole categories of objects. The kind of language that contains abstract concepts in symbolic form is also useful. By creating names, symbols, and rules for their application, language is gradually born in the mind. Concepts can be symbolized by gestures, contractual signs, spoken words, and all other symbolic ways of communicating or preserving concepts in memory. With such tools, it is easier to describe your specific location to yourself and to others. Whoever can do this possesses a higher level of consciousness called **self-consciousness**. The uniqueness of this state lies in the fact that besides its subjective feeling, we have ability to analyze and describe it with various levels of detail. We deny this ability to inanimate objects, such as machines and computers, even if they have the ability to process information and communicate with us. We also deny it to most of the animals that accompany us in our earthly fate. We are ready to admit some degree of self-consciousness only to a few mammals due to the similarity of their behaviors in situations analogous to those in which we find ourselves. These animals include apes, elephants, some cetaceans, and sometimes dogs, cats, and pigs. Researchers even attribute the ability to use a primitive language to animals like chimpanzees, gorillas, dolphins, and elephants. Recent studies of animal behavior have extended the circle of self-conscious beings to include some birds (corvids) and even mollusks. Results of these studies are so significant that in some countries, painful experiments carried out on octopuses and squid without anesthesia are forbidden. The new directive of the European Union includes also cephalopods. Research results have justified the inclusion of animals like squid and octopuses in the scope of European Union's Directive 2010/63/EU due to the degree of development of the animals' nervous systems and thus their intelligence and ability to feel pain (Pietrzykowski 2011).

We now understand how difficult it is to detect self-consciousness in other people or animals. Even harder is to find it in a machine, not to mention beings from others planets and other worlds. A simple test of self-consciousness in animals, with which we can't communicate using verbal language, is known as a mirror test. In this test, we stick an attractive fruit on the head of a chimpanzee who is standing in front of a mirror, so that he can't see it without the mirror. If the chimpanzee, wanting to reach the fruit, reaches toward the mirror, he probably doesn't have self-consciousness. If, however, after seeing his reflection, he looks for the fruit on his own head, then it is highly likely that he is conscious of his own self in front of the mirror. This test puts high demands on tested subjects, and only very few two- or three-year-old human children pass it successfully. A similar test involves the researcher pointing toward an object. A significant proportion of dogs and almost all cats fail to understand the symbol of pointing. If the researcher indicates a direction, the majority of dogs and cats just gape at the finger. Only the most clever ones look in the direction indicated.

Let us note that the area of reality expands into the next level of consciousness, which includes the environmental model built in the mind. Consciousness moves from individual and more complex sensory inputs received here and now, through the geometry of environment, and to the model of the world, which is created by a collection of experiences and knowledge acquired in the course of one's own life or from others. However, self-consciousness can reach not only to the stars but may have the ability to analyze itself. Some people can analyze their own and other people's mental states. They can split them into the smallest elements through the prism of their knowledge and can critically evaluate them, detecting sensual illusions, suggestive images that don't correspond to reality, and all deviations of the mind. The level of consciousness including not only the external world but also the inner psychic world, enabling analysis of one's own psyche, is called the **reflective consciousness**. People who have it also have a kind of "theory of mind," which allows them to interpret the feelings and predict the behaviors of other people. This kind of consciousness definitely requires the use of symbolic language, and preferably also the skills of registering or recording the observed phenomena—for example in the form of writing. This allows an individual to follow step by step the order of phenomena and events at larger time intervals and then return to them in order to draw deeper conclusions. This level of reflection allows us to assign a moral value discovered in our own mind's mental abilities. This level of consciousness isn't manifested by any animal. It isn't certain whether all people have such an ability. Our mental life includes not only images of reality but also how we interpret what happens to us when we detach our thoughts from reality. These metaphysical aspects of consciousness are in the center of interest of religion, Christian and Islamic theology, in the Buddhist idea (Zen ideas), Taoism and Hindu ideas of oneness, or New Age philosophy, etc. (Newberg 2010, Newberg & Waldman 2010), Blackmore 2009; Nevzat Tarhan 2019; *Beverly 2009; Hanegraaff 1996; Ladd, Anesko, Phillips, Meyers 2010).*

The types of consciousness presented above referred to knowledge of perceived, remembered or imagined facts, ideas, beliefs, desires, intentions, all of which constitute the content of consciousness. That is why we call them **cognitive consciousness**. This knowledge reflects the attitude toward all objects, events and facts of the material or the imagined world and can be described by sentences in symbolic language, which is why this type of consciousness is also called intentional or propositional consciousness (Kim 1996,1998) and access consciousness in the Block's categories (Block 2007). Most of one's sensual experiences is the **access consciousness** because it makes them accessible to one's language production system.

We are aware that this kind of content of consciousness does not exhaust the richness of sensations, feelings, and emotions that a conscious being feels. Direct sensory impressions are seen from the first-person perspective as a subjective experience and their awareness is called **phenomenal consciousness**. Phenomenal consciousness is also the "what it is likeness" of a mental state. What is it like to be a being reliving those subjective feelings and emotions? The distinction between access consciousness and phenomenal consciousness was most famously pointed out by Block (1995); but see also Natsoulas (1978, 1992, 1994) and Nagel (1974).

The disclosure of self-consciousness in one's mind doesn't exclude simultaneous perceptual consciousness of sensory impressions and consciousness of the geometry of the space in which a person is located. The sharp distinctions of types of consciousness in the world of living beings never occur. In different situations, in different individuals, one or another type of consciousness dominates in different proportions, which the mind uses in variety of ways. The same person can simultaneously ride a bicycle, using the memory of pedaling and balancing, while steering the wheel reflexively, remembering the right road to the destination, thinking about the upcoming meeting, and admiring the blue sky and the scent of the flowering meadow. These process take place at different levels of consciousness either simultaneously or though frequent switching of attention.

What is this "attention"? What brain process is responsible for it? How can we deliberately guide it, and how often does it accidentally switch itself? And if we can direct it, do we have a free will? If our consciousness is a function of the brain, and the brain is a material jelly in which only biochemical and electrical process occur, then the result of these processes is strictly determined by the laws of physics. Our behavior is determined up to the end of our existence by the laws of physics and initial conditions under which all our particles are at a given moment. Indeed, external conditions can still influence the reactions of the organism controlled by our mind. But these reactions must in such cases be strictly defined and automatic. The same action should always bring the same result. Philosophers call this cause and effect determinism. If so, where is the place for "free will"? We have already seen that in the case of animals and primitive machines, reactions can be automatic and predictable. Simultaneously, unlike worms, we have the conviction that we can do what we want. We are free in our decision making. We almost always have a choice, and we feel that we can counteract the pressure of our environment by following our own, sometimes contradictory, will. The explanation of consciousness, as promised in the title, must also explain these paradoxes.

In searching for answers to these questions and problems, we can rely little on existing, known models. First of all, we must reject all dualistic models, because they assume the occurrence of some nonphysical factor as an independent explanation of phenomenal consciousness or in cooperation with known material-based psychological processes. Nonphysical, nonmaterial causes are contrary to our assumption of seeking a reductionist explanation. Accepting such explanations that go beyond the physics of the material world is a manifestation of the helplessness of science in the face of phenomenal consciousness, which David Chalmers explicitly postulates in his book *The Conscious Mind: In Search of a Fundamental Theory* (1996). He states there that we are not aware of any indications that any physical facts are involved in the formation of feelings, sensations, and experiences that are the elements of phenomenal consciousness.

%%

Self-consciousness, similar to the other levels of consciousness, is strictly personal. We can't say whether other beings are conscious, feel something, or think about something.

The main basis for believing that we aren't the only thinking beings in this world is the observed similarity of behaviors. When I hit my hand with a hammer, I feel pain. I frown or writhe in pain. If a being similar to me also frowns or writhes in pain after a similar experience, I can suspect that she also has similar feelings. This also applies to behavior of animals. From this we can know for sure that they also feel pain.

We have already written about pain as the most elemental sensory feeling. Yet pain and suffering accompany us even when our body isn't being tortured. We experience the pain of failure, of parting from a loved one, of losing the world we love, of discomfort, even of frustration when the fridge is empty and the specter of hunger is hanging in front of us. We intuitively know that this is a completely different pain. It's psychic pain, and it's sometimes worse than physical. For example, some might say, "I would cut off my hand to avoid being a slave." What inside us causes this feeling, the desire to avoid psychic pain to the point where we would consider cutting off members of our own body? To answer this question, we need to learn the different features of our thinking, which we will describe in the next chapters.

We can quite accurately describe the content of consciousness, especially when we are thinking about abstract technical problems, ethical or social issues, or any other sciences. We can then rationally argue that our mind objectively has just such thoughts and no others. However, this only corresponds to cognitive consciousness. It is in contradiction with the conviction that the feeling that we are awake and conscious has a personal, subjective character and may relate to hard-to-define feelings and fleeting sensations. We call these states phenomenal consciousness. We feel the state of consciousness clearly when we pay attention to it. We know when we are conscious and when we lose consciousness. By drinking alcohol at a slow pace, we can even see how we gradually lose consciousness. So, are we aware objectively or subjectively? We will try to explain this contradiction further.

We attribute the ability to think to intelligent beings. But does every intelligent system think? Many natural systems have intelligence, or at least are able to behave and perform activities that people do using intelligence. Intelligent systems, such as most animals and some artificial machines, can recognize their surroundings and perform effective actions beneficial to the system. After all, intelligence is the ability to learn how to survive in a hostile environment. So, how one can recognize that they think? Are they conscious of the stimuli? Do they feel pain and satisfaction? Do they understand who they are, why they exist, and whether they have self-consciousness? And if they don't have thoughts, how are they different from those of us who do? Why don't they have thoughts? What are they missing? These are fundamental, important, and urgent questions.

Thought—the essence of thinking—is the ability to evoke the feelings we experienced when we were learning to perceive the world, our surroundings, and our loved ones, but also the feelings we had when we perceived and felt our own distinct mental states. These states could refer to both the simplest, most directly sensual feelings (so-called qualia) as well as states corresponding to objects that we identified in our mind and perhaps also named. This is the ability to imagine new experiences and new objects in a completely new environment. It is also the ability to imagine the effects of one's own actions, thus planning and choosing the way of conduct, which means the ability to make decisions. Finally, it is the ability to imagine one's own impact on the perceived objects, on transforming them and influencing their different properties. But how is it possible? How can it be done? To find the answer to these secrets, we must enter the world of imagination.

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

Representing the World

Creating a reductive model of consciousness seems like mission impossible. "Reductive or reductionist model of mind" sounds like an epithet, and admitting to reductionist intentions is considered to be a simplified, primitive, or naive view. For many decades, physicalists and reductionists have not confessed their views and have hidden them under other names, like naturalism or connectionism. Despite this, being aware of the difficulty of the task, we try to create a model for reducing consciousness, including phenomenal consciousness. We claim that modern knowledge of neuronal processes is sufficient to understand how consciousness can arise in natural brains. In the following chapters of this book, we indicate that this model must take into account the existence of the body or the housing in which the modeled mind is located, the needs and feelings that motivate it to act, and the emotions that arise when these needs, drives and instincts remain unsatisfied. That is why we will call such a model Motivated Emotional Mind (MEM).

The patterns of object features and their representations show a certain degree of similarity between individuals raised in the same environment, regardless of whether these individuals belong to the same or different species. Of course, the quite different senses that animals may have will make a lot of difference. Even small changes in eye sensitivity in different spectral ranges result in differences in perception. For example, dogs' eyes have a narrower spectral range, making them less sensitive to the violet color corresponding to the shortest light waves and to the red color corresponding to the longest wavelengths. Thus, they see the world mainly in yellow, blue, and shades of gray. Bees' eyes, on the other hand, consist of thousands of miniature lenses, each of which only sees a part of the image. However, these fragments are folded in the brain into a complete picture. Good ultraviolet vision makes it easier for them to distinguish between pollen and flower petals. Birds also see well in the ultraviolet range. Sensitivity to the magnetic field allows them to "see" the magnetic poles. In addition, birds of prey, such as falcons and eagles, have a very wide field of vision, allowing them to create a panoramic stereoscopic image and accurately estimate distance. Sharks, in turn, see perfectly in the water—over ten times better than man. However, they are color blind and do not distinguish colors. An interesting example of building the representation of a mental environment is creating an image of the environment based on the received, reflected ultrasound signals in a bat's brain. These signals in no way reflect the spatial shape of the surrounding environment. And yet, from the effectiveness and efficiency of the bat's movement in its environment, one can infer that such a picture is created in its brain (Nagel 1974). A similar, though slightly simpler, scheme of action leads to the creation of a spatial acoustic scene in the human imagination. Thanks to binaural hearing, we are able to distinguish the arrangement of the instruments in a symphonic orchestra. However, we are far from locating a moth flying in front of the stage, although this is not a problem for a bat.

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

An even more interesting question concerns our desires, intentions, and beliefs. Can they be the reason for action? It seems that psychology clearly answers this question. After all, we operate on the basis of our beliefs and strive to fulfill our dreams and desires. It is clear that if we want to drink, we pour water into a glass and drink to our heart's content if there is water available. We will feel relief that the desire has been quenched. At the same time, however, our mind-guided body is capable of performing another cycle of activities. When dehydration occurs, resulting from reduced water supply or fluid loss in an amount exceeding the body's compensatory capacity, then appropriate peripheral sensors will register this as hypovolemia. There will be a stimulation of the thirst center in the hypothalamus, followed by the transmission of signals to the frontal lobes to stimulate the motor system that initiates movements to fill the glass with water, raise it, and start drinking. If we do this, we will achieve the desired psychical effect: the feeling of thirst being quenched.

This second explanation is not based on mental processes but only on biochemical and neuronal processes. At the biophysical level, it is possible to indicate a causal chain in the performance of individual activities. It can be considered that giving these reasons explains the behavior of the body responding to water deficiency. However, the question arises, What really caused the drinking of water? The essence of the problem is that we have two causes for both effects: the physical effect (we had a drink of water) and the mental effect (we felt satisfaction from quenching our thirst). Did the signals the nervous system brought to the motor fields control the movement of the hand? Should we rather consider the mental feeling of thirst as an explanation for the operation of drinking water? Can the mental state cause physical movements, or is it the beginning of a chain of reasons causing a change in the final mental state? Or is it maybe the stimulation of neural fields in the brain, based on the principles of biophysics, that steers the body to achieve homeostasis, and both the mental state of thirst and the final state of hydration are caused by the states that the body attains during this process? Jaegwon Kim formulates this problem on the basis of metaphysics and tries to find the answer to this question (1998). He gives an even more condensed example: "Doreen winced because she felt a sharp pain in her elbow. What causes Doreen's face curvature?" The answer to this question focuses on the main problems of the philosophy of mind and *de facto* determines the choice of the philosophical system within which we move.

The reductive model we are looking for in this book is inextricably linked to physicalism and materialism. Physicalism or materialism presupposes that everything that exists in the world is physical or material. The obvious problem concerns the mind: is the mind physical? Are subjectivity and purposefulness physical? Some proposed solutions try to reduce such concepts to brain conditions or to simply eliminate them, an idea we call eliminativism. We do not expect restoration of any natural brain, and we are ready to settle for the functional model of the brain that leads us to functionalism. Like many other researchers, we will identify functionalism and physicalism as interchangeable concepts (Block 1980). Both approaches assume that the mind is a function that can be accomplished in any physical system—in the biological brain, in electronic or optical systems, and so forth. The essence of functionalism and physicalism lies in the function of the system. If the system functions as if it were equipped with a human brain, then this system also behaves like a human and we can say that it is conscious. Input data, internal system structure, causation, and effects are important. Thus, the functional model of the brain will be considered the sought-after reductive model.

And what? Is that it? Is this supposed to be a model explaining the impact of the mind on the body, answering one of the most difficult questions formulated by a humanity wishing to distinguish itself from the rest of the animal world? First of all, we see that this is not a mathematical reduction model consisting in deriving mathematical transformations that transform biophysical phenomena into mental phenomena, as can be done in the sciences, like transforming the Kepler model of planetary motion around the sun by introducing Newton's equations into it, or supplementing a model of an internal combustion engine with thermodynamic equations.

Similarly, our model should prove that an organism equipped with a neural network forming an associative memory capable of creating hierarchical representations subjected to experiencing phenomena in its environment will reveal mental phenomena and if not, it will subjectively feel them. Let's consider whether in the brain described by our model this is really happening and what premises indicate that all aspects of the mind-body relationship can be explained on this basis. Can we explain why Doreen has a grimace of pain on her face?

%%%

Embodiment

The previous chapter made us realize how far we are from the sought-after model of a motivated emotional mind (MEM). The layered structures of the hierarchical associative memory recording the neural representations of perceptions, sensations, and images of the environment, described in the previous chapter, do not resemble natural brains that are known to us. They lack emotions and feelings; they also lack the motivation to act and many higher mental functions that we associate with the features of conscious beings. Creating feelings requires emotional perceptions, memories, plans, beliefs, and intentions. How can this be achieved? How are perceived objects and events given values referring to their significance for the fate of the conscious system? Are these events favorable or unfavorable? Does the system perceive them as nice and pleasant or rather painful and threatening? Do they meet the various needs of the system, both material and mental?

To answer these questions let us more systematically distinguish mental states and events. There are two broad categories of mental property. Mental states, such as thoughts and desires, contain content that can be described by referring to something that we perceive or expect, that we believe. The central axis of this content is “something.” For example, you might have the thought of wanting to meet a friend. These states have intentionality.

The second category concerns sensory feelings that do not contain intentional content but instead have different qualitative properties. Qualitative features of objects are perceived as perceptions, impressions, feelings, and sensations (Kripke 1980).

How to make a being with a brain have phenomenal awareness? The path to resolving this dilemma leads through embodiment. In chapter 1 we show that relating perception to the body-environment relationship enriches the repertoire of created abstract concepts and ideas. It turns out that the impact of the embodiment on the degree of development of the mind can be extremely comprehensive and sophisticated. The suspicion that the body of living organism is a source of sensations shaping its feelings and emotions has appeared many times in the work of neurologists, psychologists, and philosophers of the mind. The inseparable relationship between the categories of thinking and feeling with the body was postulated by Francisco Varela and others (1992). The authors pointed to the body as the subject of impressions and the habitat of cognitive processes. Antonio Damasio devoted much space to the role of feelings and emotions related to the body in his works and well-known books (2010, 2018). Although this idea does not have many opponents, no one has indicated how experienced emotions and feelings in the psychological sphere can shape the structure of the brain so that these mental states result in full cognitive and phenomenal consciousness.

In this book, we would like to propose four domains in which the body or the embodiment of an artificial system can affect the brain to create and expand consciousness. These domains are:

1. The homeostatic system. The body or housing may contain sensors informing the brain about the internal conditions of the body. The signals from these sensors can complement the information coming from the external senses.

2. The motor system. The housing and body, together with the motor system, allow an individual to manipulate objects in the environment and its own body in the environment. The effects of these manipulations can broaden the experience and allow for their evaluation;

3. Participatory analysis. The body or housing can be used to analyze and plan activities by making calculations through a physical process.

4. The global states of the organism. Internal power supply parameters, information-processing speed, dynamics of operation, and sensitivity thresholds for internal and external sensors can affect performance, the results of evaluation of sensations, and the shape of neural representations.

We will discuss sequentially which processes can occur in these domains, beginning with the idea of embodiment. The essence of the embodiment is to place the brain controlling the system in a body with senses and motor systems that can perform actions under the influence of commands flowing from the brain. This composition is typical for most living things. Their brain gains new opportunities and new chances for a deeper understanding of the surroundings and objects in the vicinity. The embodied being can manipulate objects and its own position in the environment. In this way it can check the effects of these manipulations. The embodied being may have internal sensors controlling its homeostatic state. Such a being may feel a change in the parameters of the functioning of all organs of its body. Finally, it can use the body to learn how to perform complex tasks that require complicated interactions with the environment.

Ad. 1. Homeostatic System

A growing group of psychologists and cognitive scientists has no doubt that the sense of consciousness is associated with the emotions and feelings flowing from the inside of the body, directly from our intestines. According to Antonio Damasio, we perceive the external world that reaches us from the senses through the image of our own body (2010, 2018). An internal sense of compliance or deviation from homeostasis allows us to assess the perception of the environment. The interior of the body of animals and people is packed with various types of sensors that, by sending signals to selected parts of the brain, inform us about the state of the body and internal organs. The receptors for sensation of the body's own state include deep-sensation receptors, so-called proprioreceptors, located in muscles, tendons, and joints. They react to muscle tension, pressure, tendon and muscle stretching, movements of body parts, and their relative position. Nociceptors, by sending signals to the brain, make it possible to feel pain. Internal interoceptive sensations provide signals from interoreceptors located in the viscera, internal organs, and blood vessels. Interoreceptors also analyze information from endocrine glands. The cluster of receptors is located in the membranous labyrinth regulating the sense of balance and motor-visual coordination.

Now an embodied organism can integrate the effects of its action with signals from within the body. The network of associations will thicken considerably, because the fixed patterns of objects and phenomena will be accompanied by internal sensations previously generated when similar phenomena occurred when the body came into contact with them for the first time and had the opportunity to feel their effects.

Can the mere observation of the effects of the action and the recording of changes in the state of internal organs account for the feeling that the accompanying observation of the object or perception of the phenomenon is carried out consciously? In any case, only electrochemical pulses reach the brain. How can the brain convert these impulses into a subjective feeling? In addition, each of these pulses has a similar amplitude determined by its own energy of action potential generated by the neuron. How should they stand out so that the brain can consider them as subjective, first-person, conscious impressions?

Of course, a single impulse does not create consciousness. Its association with the dynamics of the process extracted from episodic memory will allow us to identify what is happening. The simultaneous sensation of cut skin evokes association with similar events from the past, when we suffered other injuries, and immediately provokes the reflex of hand withdrawal and an attempt to stop further damage, involving procedural memory. At the same time, a stimulation signal will flow from numerous nociceptors, which our brain will interpret as pain. Please note, no signal of pain will flow through the nerve fibers that provide information about nociceptors. There is no pain at the site of injury. Normal ionic currents flow through these fibers, the same as in other axons in the neighborhood, in the brain, and throughout the body. Due to the associations with the entire described complex, this signal will be very extensive and will dominate others received parallel stimuli. It will dominate over the constant background of other signals and as a result it will have the nature of an alarm interrupting the normal functioning of the mind. This elementary signal of feeling, which is our genetic heritage, causes such an alarm to be associated with negative effects and is therefore perceived as a signal for an immediate radical reaction of the body to reduce damage. Strong stimulation of a large number of pain receptors can dominate all perceptions. An entity experiencing this will thus not be able to perform any physical or mental activities other than trying to avoid this overwhelming feeling of pain.

An artificial system devoid of this genetic heritage has yet to learn to feel pain. If signals from specialized receptors are associated with negative effects, then the embodied agent will not “like” this feeling. Life experiences can strengthen these associations. We are then sensitive to pain. We can also try to embed artificial pain into artificial brains, and such artificial pain will then be the legacy of the act of creating the motivated emotional system.

Wiesław recalls, “My own experience introduced me to the opposite phenomenon. As a young sailor, I used pliers to loosen my shackles. An attempt to unscrew the pin without pliers caused severe finger pain. Sailing with an experienced colleague, I noticed that he was unscrewing the pin with his fingers. I asked, ‘How do you do it that you feel no pain?’ He replied that he just took the pin and unscrewed it. At the time I tried it with determination. And strangely enough, the shackle gave way, and I, feeling the same pressure as before, did not feel any pain at all. I did it later hundreds of times with the same result. So pain perception is subjective. To a large extent, it depends on your mental attitude. There are examples of seriously injured people who remain in traumatic shock and do not feel pain.”

It should be emphasized that nociceptors do not transmit pain signals. Like other skin receptors, such as thermoreceptors, they do not convey a feeling of cold or hot. A sensation of pain, cold, or hot arises in our brain. Mainly, the psychical effect is regulated by signals coming from the upper layers in the up-down direction. These signals regulate the sensitivity of receptors, that is, the pain threshold. Even extensive, numerous receptor stimulations can be suppressed by mental processes and then can be ignored thanks to learned associations that assure the mind that the alarm is exaggerated.

The opposite course of these processes will result in the sensation of acute pain. The intensity of experiencing pain is the result of neurological filtering. When the pain threshold is exceeded, consciousness receives an aggregate, comprehensive impression of pain, often localized, contextually assessed in terms of danger and relevance to the body.

The brain constructs the impression of pain, just as it constructs the scene of the surrounding reality from a multitude of visual signals. This does not mean that the pain is apparent, that the tissues or important organs have not been seriously damaged. The construction of pain by the brain is not a volitional process. It does not depend on our conscious belief in how serious the damage is. This impression is independent of consciousness and self-awareness. Although this is a subjective impression, we are usually unable to remove it from consciousness, just as we cannot control other emotional reactions, such as a blush of shame or the hairs of your neck standing up because of fear (Thernstrom 2010).

Continued stimulation of receptors can develop associations that build pain sensations. They become more and more efficient in detecting threats and reporting them with pain signals. The brain trains to feel pain. This also can result in the construction of higher-order pains. The development of chronic pain is even more dangerous, much more severe, and more difficult to treat (Woolf 2010).

The other senses provide further examples of contextual perception. Wine drunk straight from a bottle does not provide the same sensations as classic degustation. The symphony concert taking place in the background of a pub brawl will not arouse admiration. What we feel depends on the memories of past experiences, associations, and accompanying circumstances.

The associations of memories and perceptions with past experiences allow us to feel their value and meaning. When we determine that taking some actions—for example, acquiring specific items—will bring measurable, well-defined benefits, it is not about evaluating their real value. It is about feeling that it is good, pleasant, and brings good memories.

The last term is both figurative and literal. This contextual perception combined with assessment creates an emotional relationship to the perceived trait. We like some colors, smells, and sounds, and others don't. If we have positive associations, we can admire them. We are not delighted that the pink piglet reminds us of a rose. We admire the very color pink in the painting because its view can cause a state of pleasure that we experienced when we admired a rosebud. The state of stimulation of neurons throughout the brain is associated, not just the knowledge of color and situation. Therefore, the source of our tastes can be a mystery to us if we do not remember how they were shaped. We call these direct sensual impressions qualia. Qualia, as we will explain in this book, constitute individual cases of subjective, conscious experience.

Emotions also appear as a result of unmet needs and a different course of events than expected. As a result of this process, priority of activity is set. In this way, perceptions, plans, beliefs and desires are evaluated. They acquire emotional value. Our attitude toward them is no longer neutral but becomes involved—we care about something. And again, we can see the release of a whole range of feelings depending on the level at which this happens. Different priorities and values will appear in the long or short term, and they will differ in relation to the activities and effects of the activities in which the system has involved various amounts of time and energy.

Of course, emotional feelings may be accompanied by epistemic analysis of the effects of planned or ongoing activities. As a result of this epistemic correction, emotional correction may occur. However, feelings of fear or anxiety or hope and joy about these actions are qualia, as opposed to a rational assessment of expected beneficial or adverse effects. How do mental states tie into their emotional values? Again, co-occurrence decides. Of course, as it happens, the life story and experience gained decide.

Ad. 2. Motor System

The embodied system with senses and actuators can perform actions under commands flowing from its brain. He can manipulate objects and his position in the environment. Simultaneously, he can check the effects of these manipulations. By viewing from different angles, exerting forces and pressures, touching, stroking, tasting, smelling, and performing similar operations, he can gain experience inaccessible during passive recording of signals received by the senses. By associating repeated impressions with performed manipulations, he can assess the usefulness of recognized object properties for the effectiveness of achieving its intended goal.

For example, a person walking on an icy surface may assess the adhesion of the soles of his shoes. He can assess the reliability of his grip on a rough knife handle by gripping with the hand. He can obtain and develop the impression of roughness or slipperiness by touching different surfaces with different textures. Will it be a silk handkerchief, a wet bar of soap, a stone, or sandpaper? By associating different kinds of experiences, an embodied being can evaluate such feelings. Feelings with positive effects can be subjectively perceived as nice and pleasant. Feelings that impede the achievement of intended goals can be perceived as unpleasant and causing impatience, anger, or other types of discomfort.

By testing the range of possibilities in achieving goals, we have transformed simple sensual perceptions (qualia) into emotionally marked sensations. The objective states of the observed objects have acquired a subjective character, because during attempts to implement the planned tasks, they were associated with emotions through obtained results. The objects were associated with the effects of actions, impressions, and emotions related to the role that these objects played in the life history of the organism or the embodied robot. The cognitive system gains the significance of these objects because they are now embedded in qualia.

This feature is not available to systems that algorithmically process information about abstractly defined objects and the logical relationships between them. We must realize that ideas of abstract concepts cannot be directly felt, because we cannot present them to our senses. We can only feel secondary emotions caused by their recall. Likewise, we will encounter a problem in trying to define our direct sensual feelings (i.e., qualia), such as the perception of colors, taste of wine, smell of violets, and so on. They do not have objective characteristics that would allow them to be defined. That is why we can only speak about them in the language of poetry, using metaphors, comparisons, hyperbole, and synesthesia.

Ad. 3. Participatory Analysis

Most animals are active in their environment. The drive of activity is the pursuit of homeostasis and adaptation to changes in the environment. This is a prerequisite for evolution. The ability to learn how to survive in a hostile environment is synonymous with intelligent behavior. Some animals with more developed brains have the ability to plan purposeful activities to achieve the intended effect. The level of intelligence that enables analytical thinking has ensured success for humans and some apes.

It seems that analytically controlled, intelligent, and planned behavior would always dominate. Despite this, animals deprived of high intelligence function extremely effectively in their natural environment. Their bodies adapted to the environment in the course of evolution, and they themselves learned to make the most of these adaptations for survival.

A good example of using the body to perform complex operations is to hit a target. When a chameleon is aiming at an insect, it doesn't try to "calculate" the angle of its tongue. It has a fixed image of the point where its tongue hits when it thrusts it out in a practiced way. Then it sets its body and its head, equipped with perfectly functioning eyes, in such a way that the point of impact coincides with the image of the prey, and the chameleon performs routine execution with its sticky tongue. This is a completely different procedure than an artillery computer controlling a cannon. If an ancient hunter had had to make similar calculations when aiming a spear at an antelope, he would inevitably have starved. A trained hunter initiates a spear-throw as soon as he sees an animal. During the throw, he observes the direction indicated by the spear blade and corrects the thrust direction so that it coincides with the planned hit location. Of course, the matter of practice is to take into account corrections for the gravitational curve of the flight path and the parallax error resulting from the shift of the observation point. A slightly different procedure is used by a ball catcher. He also does not calculate the point of fall but is heading for the predicted hit. In the final phase, he corrects his position and modifies his precise hand movements so that the approaching ball is in his central field of view when his head is turned toward it. There is motor control with automatic correction of the gripping place. Often, when looking for an approaching object, the catcher does not notice any obstacles that may be at the point of fall. Efficiency is at risk.

Ad. 4. The global states of the organism
Often the stimulation of mental processes is caused by unconscious stimuli. This happens when they are masked by other mental processes that are tracking processes in a state of focused attention, planning, and conscious analysis of problems. It is believed that many of these types of stimuli and processes can be analyzed and processed at the subconscious level. The effectiveness of this process can be mitigated by the parameters of the neural network. These are parameters such as axonal and dendritic conductivity; triggering thresholds; potentials on synaptic membranes; concentrations of calcium, sodium, and potassium ions in synaptic clefts; neurotransmitter and hormone concentrations; the brain's own oscillation activity; and the like. Let living organisms serve as an example of the occurrence of these interactions.

The processes of association and transfer of stimuli are largely determined by the instantaneous values of these and many other parameters determined by the endocrine system and the dosage of hormones spread throughout the body through the circulatory system. The proper functioning of the circulatory system ensures the correct distribution of nutrients and prevents brain hypoxia. The functioning of many areas of the brain depends on the level of neurotransmitters and neuromodulators. Specific body reactions are controlled by the levels of adrenaline, testosterone, oxytocin, or vasopressin and other hormones and endogenous neuromodulators, as well as by those administered from the outside, by injection or with food or drink. All these factors affect brain function, reaction speed, and all other mental processes, sensory sensitivity, range and richness of association, and stimulation of imagination, as well as generation of visions, illusions, delusions, and many other symptoms known to psychiatry and psychology. Changes in mental states caused by global chemical and electrical influences often result in the direct creation or indirect induction of emotional states through a different perception of the environment or its modified feeling. An unrealistic, changed picture of the world can cause fear, surprise, or euphoria.

We can imagine many global processes, including the entire embodied system, even if this system is not a living organism. For example, a robot may have variable dynamics of operation if the power-supply conditions change due to battery depletion or a change in the sensitivity of receptors, response thresholds of neural network cells, characteristics of filters transmitting control impulses, and so on. Many processes in the brains and living organisms that lead to emotional effects are described in part II, and ways to control global parameters of artificial brains are discussed in part III.

The impact of global interactions on mental states is associated with the modification of signal propagation between the sensory, motor, and brain systems affected by these states. Because the secretion of many endocrine glands is controlled by signals from the brain, it is possible to form feedback that stabilizes the mental state or, if necessary, mobilizes the body to aggressive actions or rest.

Impressions, Feelings, and Qualia

All combinations of these four domains can generate an extraordinary wealth of mental states, both intentional and emotional, as exemplified by the human psyche. This will happen regardless of whether we are talking about living organisms or artificial systems. Of course, when considering the impact of the body on natural minds, we must take into account the evolutionary conditions affecting these interactions. We must be content with evolutionarily developed senses, reactions of natural centers of pain and pleasure, or the body shape of individuals. Their evolutionary origin common to all species causes some similarity in psychological and physiological responses. The richness of the psyche created by artificial minds in artificial bodies seems unimaginable.

However, we do not know if, once it is possible to isolate topologically specific neural fields trained in feeling pain or pleasure, these fields will not be duplicated by constructors in subsequent constructions creating standard centers of these sensations. We do not know whether, for example, the alarm associated with a decrease in power felt as energy hunger will be a common solution in all models. This will result in a certain similarity in the psyche of subsequent generations of robots. Functionally, our model allows creating a mind accumulating knowledge and feelings, showing cognitive and phenomenal awareness. The real adventure will start when we start building operational models.

Psychology provides many descriptions of experiences and feelings that we can try to interpret on the basis of the presented model. Let's look for the sources of their multitude and diversity. Sensory stimuli reaching the receptors can cause quite different effects. The effect of stimulations reaching the memory structures depends on the degree of development of these structures. This is particularly evident in the case of living beings. The reactions of animals endowed with very primitive brains are different than those of animals showing even slight signs of consciousness. The latter have much more complex brains, with more neural fields and layers of memory cells that make up these fields. A conscious being, recognizing the surroundings and the effects of its action, will have a specific property that is not intentionally programmed but is acquired during experiments. Well, such a creature, animal or human, having the ability to recognize the elementary features of objects by feeling the direct sensory impressions they cause, will be able to comprehend their meaning—for example, their suitability for planned activities. Man, feeling the hardness of the nail, will be able to assess whether it can be hammered into a slightly softer wooden board. Walking along the icy path, he will expect to slip, but given the roughness of the soles of his shoes, he will feel the stability of the steps, preventing a fall. By the brightness of a fruit's colors, he will be able to assess its ripeness, and by the smell of meat he intends to eat, he can assess its freshness. The examples given show how important qualia can be for effective functioning in any environment.

How can the sensory stimuli that reach the brain change into sensory impressions that imprints the mind of a conscious being? It is bodily reactions that can give value to individual feelings, as described above. Feeling hardness or smoothness, assessing the attractiveness of smells, judging the importance of sounds, and evaluating the favor of the environment based on images all go beyond the direct response of the senses. Reactions of the being conscious of these stimulations are combined with an assessment of the possibility of achieving goals, an assessment of the necessary effort, and the sufficiency of resources or threats, and this generates an emotional response to these stimulations. In animals and humans, these sensations are enriched with evolutionarily developed physiological reactions associated with the stimulation of pleasure centers and the release of hormones and neurotransmitters affecting the mental state of the individual. We can see that the perception of stimuli reaching the senses can be associated with the mental state of the creature subjected to these stimuli.

Let's think about the variety of sensual sensations and feelings depending on the features of the environment with which the system interacts: what impressions can be felt, how they can be recorded. Regardless of the differences in the perception of objects and their individual features, the feelings of beings exploiting a given environment may vary depending on the way in which the properties of the environment are experienced. The arising feelings depend on the assessment of their significance for a creature experiencing specific characteristics of the environment. Values added to the features are the basis for reactions and actions taken. A man looking at a fat, hairy caterpillar can feel disgust and reflexively move away from an unpleasant "worm." A bird sitting next to it, seeing a similar image, can experience a sense of the irresistible charm of creeping food. Even if we put special filters on our eyes to make the caterpillar look the same to us as it does to a hungry chickadee, we would not have the bird's feelings. Our impressions will be different, specific to our species and our experiences. There is an insurmountable barrier between the change of perception through the use of any filters, adapters, and simulators and the change of feelings and, subjective impressions induced by these perceptions. We will never be able to overcome this barrier, regardless of technology used.

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

Qualia is one of the so-called "hard" problems of sensory perception, in which it seems impossible to translate electrochemical signals occurring in sensory cells and synaptic connections of neurons to the mental feeling of cold, roughness, or color. The enigmatic nature of qualia, supposedly impossible-to-explain personal and subjective "sensory impressions," is associated with a lack of understanding of how sensations are associated with motor activities that close the flow of information between perceptions, actions, and observation of the effects of these actions. Animals, people, and other autonomous agents need the brain to move.

For this reason, an intelligent system cannot reach states of consciousness unless it is able to move and observe the effects of its operations and its environment, or to adequately respond to events in its environment. When making a move, the conscious system must be able to predict the effects of this move. Rubbing a finger across a rough surface, we stimulate touch receptors; surface irregularities determine how often and with what strength these receptors are stimulated. Predicting the characteristics of these stimuli is synonymous with recognizing the sensory input signals, and the compatibility of these stimuli with the experience of similar stimuli in the past creates an impression of surface roughness. The body feels the sensation of roughness, but it will similarly experience the feeling of softness, warmth, cold, itching, burning, or pain. It is this association of the effects of the action with the perceived signal of roughness, coldness, stickiness, or smoothness of the surface being touched that leads to a reproducible sensation of the feeling of related qualia in similar situations.

One of the most important feelings determining the behavior of living beings is the feeling of pain. The sense of touch protects the body against the harmful effects of the environment. When we feel pain, there is an automatic, unconditional movement of the body to avoid contact with the factor causing the pain. In natural brains, the mechanisms of pleasure and pain have a very diverse, evolved structure. As a result of neurological and psychological research conducted in recent years, the theory of pain and pleasure is quite well developed (see, e.g., Wyller 2005). A sense of pain (and other bodily sensations such as tickling, itching, and orgasm) is a local response to stimulation of the body area similar to actual or imaginary tissue damage (Pitcher 1970). In particular, strong stimuli exceeding the acceptability threshold of receptor systems may be felt as pain (Pain 2005). The meaning we attribute to the feeling of pain is related to the fact that it is a consciousness-related experience. It is one of the fundamental feelings formed at an early stage of evolution.

Generation of pain usually begins with the stimulation of special pain receptors (nociceptors) in the peripheral parts of the nervous system, responding to stimuli potentially damaging the tissue. They are transmitted to the central nervous system through specially dedicated nerve fibers, through a spinal cord system modulating information into the midbrain and thalamus. In the thalamus, specialized, somatotopically organized fields process these stimulations and direct them to higher brain fields. From this point of view, the mechanism of pain sensation can be perceptual. To explain the phenomenon of pain, it is necessary to study the relationship between the perception of pain (which requires categorization), its experience (sense), the identification and location of the source of pain, and the adequacy of the created mental representations (Aydede 2009). In people, even in the early stages of development, we observe sensitivity to pain. The ability to feel pain seems to be an innate trait with low plasticity, which is contrary to the postulated need to first develop the concept of pain and the ability to feel it.

Why is the stimulation of specific nerve fibers perceived as pain? In the case of injury from numerous nociceptors, an excitation signal will flow to specialized centers in the brain that our brain should interpret as pain. They are evolutionarily formed centers. Due to the associations with the entire described complex, this signal will be very extensive and will dominate other stimuli received in parallel. The more so, that the constant background of other signals leads to the habituation of other receptors. This signal will have the nature of an alarm interrupting the normal functioning of the mind and hindering the performance of other functions, especially mental ones. The genetic heritage means that such an alarm will be associated with negative effects and therefore will be perceived as a signal for immediate radical action of the body to reduce damage. We learn to recognize a threat when we experience injuries. We perceive panic in the system as an unfavorable value, disrupting homeostasis and requiring immediate response. This feeling escalates from "uh oh, no good" to "wow, it hurts!" A repetitive pain stimulus causes sensitization to this stimulus, sensitization, and ultimately even hypersensitivity. Even slight local tissue damage will evoke memories of disturbing sensations flowing from different areas of the body, blocking pleasant ones and revealing new disturbing ones. These unpleasant feelings will cause us to attempt to avoid pain, withdraw body parts from the area of exposure, escape, and cover ourselves. Then there may be a reaction expressing panic and an attempt to stop adverse effects, such as by cooling the burned area, stopping blood flowing from the wound, or a making a facial expression of dissatisfaction known as a grimace of pain.

Primordial pain is a complex phenomenon, involving both its simple perception and sensation at a high conceptual and emotional level. The apparent "physicality" of pain is a consequence of the intentional interference of the state focused on the part of the body in which the person feels pain, with the pain being a quale (Crane 2003).

Sensations caused by motor stimulation of other senses have a similar genesis. Although they are triggered by other sensors and cooperate with other motor functions, they give the body the opportunity to recognize threats and to detect and avoid pain but also to find a reward, recognize favorable conditions for the body to survive, obtain food or shelter, or acquire a sexual partner. For example, in sniffing the air, we provide chemical signals informing us about the state of the environment (e.g., about the freshness of a piece of meat). The olfactory receptor proteins react to the presence of certain chemical compounds in the olfactory receptor cell environment and stimulate sensory nerve cells. The previously remembered states of aroma are used to recognize whether, for example, the food we've found is suitable for eating or not. The stimulated sensations of taste or smell are integrally connected with the basic function of acquiring food or a partner.

There was a reason why, in the course of evolution, taste buds dedicated to specific purposes were developed, and their sensitivity to various ingredients is closely correlated with what the body needs. Real sensations were received by the body, even before specialized sensors were developed to respond to physical stimuli useful for the body. The organism sought such stimuli in the environment, and when it found them, it received a reward. The award was specific to the organism's need, and this need determines the perception of the perceived sensory stimulus. Once this real sensation arose, specialized sensors developed to detect the stimuli the body was looking for. Finally, based on sensory stimulation, the nervous system receives information about the activation of a specific sensor. In this way, the signal that the brain receives from the stimulated sensor is correlated with the original sensation that the body had when it first encountered specific stimuli during previous activity. It then could find out that malodorous meat had unpleasant effects after being eaten.

Social beings equipped with a language of social communication can associate smells with the warnings of their guardians or other friendly people whom such a creature trusts (like a trained dog). They can also associate their feelings and experiences with knowledge acquired from an encyclopedia, the internet, or school. Therefore, the perception of an object usually combines the intentional sphere with the feeling of the accompanying quale.

The body, actively using its sensors, recognizes the quality of perceived impressions in terms of their usefulness or threats. The way of perceiving the world thus formed leads to complex compositions of impressions both useful and harmful—that is, those that need to be reacted to but also those that can be ignored. Similarly, the role of interactions with the environment can be formulated to actively receive auditory or visual sensations. It is clear that these sensations are interpretations of stimuli reaching sensors that respond to excitations based on the laws of physics associated with moving the tympanic membrane by sound waves or photoreceptors activated by light. This corresponds to the creation of qualia by the motor system.

Sometimes it is believed that qualia are the most elementary impressions of a homogeneous character, such as the color red, the sound of a certain pitch, or one of the basic flavors (sour, salty, or sweet). We believe that impressions with a more complex structure can also be called qualia. We rarely see homogeneous red in front of our eyes, like in a darkroom. More often it is a specific-colored shape, like a red circle, or a set of colors, like a rainbow. These objects are already more abstract. For example, a rainbow forms an arch in the sky and may have more or less bright colors. A circle can have a specific diameter and be on a diverse background, making its perceived color distinct in contrast. These features can be described, dimensioned, and defined in detail, and in this respect they assume the character of abstract objects. However, we can perceive the colorful circle as a whole and not think about its diameter. Similarly, we can only see the rainbow colors splashed in raindrops blurring against the sky, and practically, at the first impression, we cannot say anything more about this color mix. Also, when we hear a chord, we do not think about the consonance of sounds that create it. Nor can we distinguish the composition of the fragrances that create subtle perfumes, although we can admire their composition.

Qualia can also complement complex, abstract concepts. In this case, our perceptions would combine two components, both the propositional (or functional) component and the phenomenal component. For example, when thinking about the bike we learned to ride in childhood, we can remember its colors, the coolness of the frame, and the softness of the seat. Our generalized, abstract concept of a bicycle is then supplemented with qualia giving it individual, specific features that do not allow it to be mistaken for some other bicycle. Defining it a bit more strictly, qualia lead to the consolidation of symbols of real objects by using their physical properties, relating various sensory feelings associated with the observed object. The red bicycle frame is not only shiny and smooth but also cool and hard to the touch, unlike the seat, which seems soft and warm. Therefore, symbol grounding contains elements of object recognition based on perceived features such as shape, color, or size, regardless of the angle at which the object is perceived, in what lighting, or from what distance.

Qualia are a sensual feeling of these traits coming from different senses and also undergoing modifications, such as the loss of intensity of the smell of a scented handkerchief, the faded color of a dress, or the hard, worn seat of a bicycle. However, the key process is a direct manipulation of the environment with the help of effectors and observation of the effects in the form of sensory impressions. This can be done by a conscious being. If the living biological entities that we also are can convert physical sensory stimuli into psychic sensations, it means that we have the capacity of an imaginary conscious agent, and that agent's consciousness becomes understandable to us.

Until now, neurocognitive scientists studying the phenomenon of sensory perception have expressed the belief that there is nothing to be said about the subjective perception of these sensations by other people. We mentioned a language of poetic metaphors that could be relevant here and that boil down to the fact that we transfer experience from one field, from one modality, to another field (modality). In particular, as previously thought, we are not able to describe the subjective internal states from the perspective of another person standing next to us (from a third-person perspective) using concepts objectively describing the subjective states of the first person. The thesis about the similarity of qualia radically changes the way of looking at this issue. As we noted above, differences in perception caused by the use of other senses are not a serious obstacle to imagining what image the senses create in the minds of alien beings.

In Nagel's book, cited above, the author convinces us that beings belonging to an alien civilization—for example, Martians—without sight will never be able to imagine how Earthlings would see the Martian landscape.

Nagel sees even greater difficulties in imagining how a bat "sees" the world with the help of its echolocation devices. The author's worry seems to be completely pointless. If the Martians have sufficient physical knowledge, they will easily examine what part of the spectrum of sunlight or artificial light people perceive with the help of their eyes. Knowing the reflection characteristics of the entire spectrum of light waves, they can reproduce the image that people will have in front of their eyes, if they have any eyes at all. If they use the sense of smell instead of the eyes, it will be difficult for them to imagine how we perceive the world through our eyes. A blind person will not see our world, and if he is also deaf, his possibilities of understanding our way of perceiving the world will be even more limited.

Even more difficult is to imagine what image the world shows a bat from the reflections of ultrasonic squeaks it sends. The issue of propagation and reflection of sound waves in connection with the subtle properties of the bat's ears is a difficult issue for our knowledge and ability to simulate sound images. Not knowing in detail the frequency characteristics of the sounds sent and how they are processed after reflection from objects of different sizes and textures, even modern supercomputers, which are accustomed to assembling layered images from magnetic resonance scans, may have difficulties. We do not know how important angular distortions, phase shifts, and time delays of received signals are for a bat's hearing apparatus. But all these difficulties are a description from a third-person perspective! Our lack of knowledge as non-bats makes it difficult for us to empathize with the bat's experience. Any strange distortion of the image of the surroundings by the bat's mind is not a problem, because such distortion cannot occur. Of course, the properties of the bat's senses do not exclude that some objects will not be highlighted or that others will seem clearer to him. It does not matter. We can detect and describe these differences objectively. Observing the bat's effective pursuit of a moth, its ability to bypass all obstacles, and its ease of maneuvering in a cloud of other bats, we have no doubt that the right qualia are formed in the bat's mind such that it not only perfectly recognizes the objects but also probably has a spatial plan of the environment. Similarly, we, meeting Martians, could examine what they see. Then, assuming that they are able, thanks to their eyes, to model the surface of the Martian planet as well as their and our place on this surface, we could assume that their view of the Martian world is consistent with ours.

We often have to predict what kind of images we can perceive by changing the characteristics of the senses. This happens when we use vision-distorting tools for observation. Aircraft controllers, observers of radio and echolocations, ultrasound camera diagnosticians, astronomers, and users of thermal-imaging devices see a completely different world on their screens than the one we are used to. When they become skilled, they develop new qualia that most people don't have. They see much more than the untrained eye. However, anyone who learns the objective, physical principles of these devices can imagine what world they subjectively perceive.

And finally, let's cite the simplest example. Many of us have been in a room where the only light source is a red bulb that emits red light in a narrow spectral range. Is it so hard to imagine that we could be suddenly in a room lit by a green bulb? Having a little knowledge about the characteristics of absorption and reflection of colors by the surfaces of various objects inside the room, you can imagine which of them we will see better and which will disappear from view because of too-low contrast.

The question is often asked, why do we perceive qualia in this and not another way? Why is the contrast between red, yellow, and blue so large? After all, the difference in the length of received waves is insignificant, but the difference in perception is enormous. It is an achievement of the evolutionary process, to which we all were simultaneously subjected together with the accompanying fauna and flora. Now that we understand that emotions, depend on how we assess the value of the perceived stimulus, it is easy to understand that, from evolutionary perspective, distinguishing even slight differences in the emitted spectrum of electromagnetic waves or sound, while searching for food and avoiding danger was beneficial. Co-evolving organisms sought to exploit these growing discriminatory abilities by adopting colors and patterns that stimulate co-participants to respond in the way favorable for them. Especially if it concerned flowers and fruits, or partners during the mating season, which "wanted" to be noticed or eaten, and predators, who wanted to remain hidden from the senses of threatened individuals. Therefore, even a small difference in the distribution of colors could arouse huge emotions and be associated with something nice and pleasant or scary, or even something unusual, arousing interest. These associations and related emotions resulted in perception of a colorful and vibrant world around.

Consider the case of Mary, who knows everything about the physics and biophysics of receiving red but does not know what red looks like because she has never seen it. Can her case be explained now? Yes, of course. We write about it throughout. The condition for the correct feeling of red is the experience of associating the perception of red objects through experience. If Mary experiences the feelings associated with the view of a beautiful red dress, a red apple, a worker's red banner, and red blood, then the quale of red will find its mental representation (Harman 1999, 44). Michael Tye first drew attention to this, writing, "Phenomenal concepts are the concepts that are utilized when a person introspects his phenomenal state and forms a conception of what it is like for him at that time…Intuitively, possessing the phenomenal concept 'red' requires one to have experienced red and one to have acquired the ability to tell in the appropriate circumstances which things are red directly on the basis of one's experiences" (1995: 162, :167; 2000).

The structure of complex representations of the world can be seen through an autopsy. One of the authors, Wiesław, recalls awakening from a very deep and pleasant dream. The scenery of the dream was palace interiors filled with nice though unknown people. Wiesław remembers that when he woke up gradually, he tried to stay asleep and stay with these people for as long as possible in such a nice scene. Unfortunately, the image of the palace room gradually fell apart, as if it were made of fragments representing walls, windows, tables, paintings, and chandeliers. There were also fewer people. The person with whom he been conversing was the last in sight. Wiesław tried to answer her last question, but in a moment he couldn't remember what the question was. Disappointed, he tried to evoke the disappearing scenes, but the signals from the real environment soon dominated the perception and turned the content of sleep into a pale memory. Sensual arousals from the real world began to compete effectively with the patterns inherent in the subconscious mind and evoked by the sequence of stimulations in a dream. Winning this competition, they removed the sleepy representations from working memory, creating the subconscious. Only a small part of them found a permanent place there. Attempts to consciously maintain sleeping representations allowed the consolidation of some of these representations in short-term memory, which could dominate working memory again and prolong contact with sleepy matter. The accompanying electrical states may have deformed the proteins forming the basis of the neural processes, and the representation structure could have been modified due to co-occurring synaptic couplings. In this way, the whole dream could be remembered and re-created by evoking it from the depths of the subconscious through the process of reminding.

The possibility of at least partial feeling of qualia in a dream is debatable. Many people talk about colorful dreams. The subjective perception of colors is considered a typical example of qualia. At the same time, many people report that the colors in their dreams are blurred as if faded. Other qualia, sound, smell, or touch, are also very indistinct in a dream, devoid of the clarity and detail characteristic of waking. Qualia in a dream resemble memories—mostly images, sometimes words. Often, awakening occurs at the time of expected touch or sensation associated with movement. However, nothing prevents one from remembering dreams or past events to evoke the same emotions that accompanied them. If in recalling memories or dreams we expect to arouse identical emotional states through the global states of the body, then we may be severely disappointed, because we cannot count on the fact that recalling the saved scenes will release to the body a similar sequence of hormones, neurotransmitters, and neuromodulators, causing similar emotional states as when the episode was remembered. This is rather uncommon. The levels of adrenaline, serotonin, and oxytocin we experience when recalling past episodes or dreams are not as high as during the original events. Anger and fear are milder, and admiration and love are less intense.

We have previously asked the question why *Caenorhabditis elegans* cannot feel qualia. Now the answer is simple. This primitive worm does not have mental sensory feelings, because its automatic responses do not relate to the wider model of the environment. It has no psyche, because this needs contexts for its activity. Qualia cannot be integrated into the *C. elegans* model of the world because it does not have it. Its "idea" of reality is only those perceived physical stimulations of sensory cells.

So are qualia identical to the psyche? Unfortunately not. People, animals with a high degree of consciousness, and hypothetical conscious robots must exhibit other characteristics that we have not discussed so far. First of all, they must want to think. They must have motivations to think and to act.

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

Living organisms are constantly in contact with motivational pressure in the form of feelings of pain or pleasure. Pain sensation and the emotions associated with it are their most common feeling, absolutely punishing behaviors that do not sufficiently adapt their nervous system to the environment.

Of course, for the most primitive animals (e.g., insects or worms), the ability to feel emotions may be questioned. One may wonder whether instincts or automatic reactions should be discussed here. For many activities for which humans use intelligence, animals are motivated by threat of punishment or expectation of reward, equated with pain or pleasure. Errors in the use of reflexes, as a rule, are punished with pain in the broad sense of the word, but we are aware that the philosophy of pain uses a definition and analysis referring mainly to bodily pain, which relates to tissue damage. It seems that higher animals, especially social ones, experience more complex forms of pain. There is also the mental pain associated with the loss of members of the family or herd. The higher the level of intelligence, the greater the role of pain motivation and behaviors learned under the influence of this motivation. On the other hand, the role of instinctive and innate behaviors is decreasing. The mental nature of feelings of pain and pleasure requires at least some degree of consciousness. The reactions of organisms lacking a central nervous system and communication abilities are purely automatic. They can demonstrate behaviors with a certain degree of intelligence, ranging from simple tropisms to the ability to group, modify physicochemical functions, and so forth. They can even show some learning ability. However, they do not have the ability to plan, predict the effects of their own actions, build a sufficiently complex model of the environment, and perform other activities that show manifestations of consciousness. Without consciousness, pain cannot be felt, so it cannot fulfill its motivational function. Hence the suspicion that the most primitive animals with limited intelligence are not able to feel pain. Their behavior is automatic and does not require any motivation.

As we have written many times, consciousness, the ability to feel sensations, is a gradual feature. It may appear even in animals in the early stages of evolution, those to which we generally deny consciousness of the world and certainly self-consciousness. It is difficult to clearly define the boundaries of psychological phenomena. Mollusk consciousness studies suggest that the beginning of the psyche is the central nervous system. That is why psychic abilities and the conscious feeling of primitive stimuli should definitely be granted to vertebrates. Amphibians and reptiles, and even fish, clearly demonstrate this, which forces us to respect their right to protection against torment and unnecessary pain.

Knowing how a sense of pain can arise in the mind, we immediately realize how commonly, how often, and in what circumstances it can occur. Let’s trace the sources of pain. To this end, it is worth realizing that an organism operating in the environment has many needs that go beyond conscious perception, planning, and implementation of activities leading to the set goals. The body must maintain the efficiency of all organs, obtain energy to operate, maintain the right temperature and, to this end, maintain metabolism and a location that guarantees safety while avoiding exposure to extreme overloads, shocks, electromagnetic, sound, and even ionizing radiation.

Many systems inside the body warn against the malfunctioning of important body systems. This is done by so-called deep (proprioceptive) sensation, which transmits to the brain signals from special sensors called proprioreceptors, located in the muscles, tendons, and labyrinth, which allow determination of muscle tone and the position of the limbs and, in the case of the labyrinth, also recognizing the position of the body in space and a sense of balance. It is accompanied by visceral (interoceptive) sensing transmitting signals about the state of internal organs from special cells called interoreceptors to brain centers located in the brain stem and subcortical centers. Interoreceptors are located in internal organs and blood vessels. The integration of information takes place mainly in the brain stem and subcortical centers. Deep and visceral sensation is often vague. For this reason, it is usually difficult for us to localize the source of pain inside the body.

Bodies of living organisms commonly feel the need for metabolism, which is a condition of the body’s functioning. The process of metabolism provides the body with the energy necessary for life. The main needs are liquids, food, and oxygen. The deficiency of these factors is associated with the feeling of thirst, hunger, and suffocation when there is insufficient oxygen. The desire is manifested in the urge to drink the right amount of water. Thirst increases as the overall fluid volume decreases or as the amount of some substances, such as salts, increases. If the water level falls below a certain threshold, or the concentration of salt ions reaches a specific concentration level, then the osmoreceptors located in the front of the hypothalamus enable this part of the brain to increase thirst and efforts to satisfy it. The basis of the feeling of hunger is a shortage of food resulting in a decrease in the concentration of such nutrients as proteins, fats, sugars, vitamins, and mineral salts. For example, drops in glucose are recorded by hunger and satiety centers in a part of the brain called the hypothalamus.

In turn, other sensors, the so-called central chemoreceptors (in the brain stem), are sensitive to the concentration of hydrogen ions in the blood and cerebrospinal fluid, as well as blood pressure and carbon dioxide partial pressure. In contrast, peripheral chemoreceptors located in the cervical and aortic glomeruli respond to oxygen partial pressure. They allow us to react dramatically to oxygen deficiency, giving us a chance to defend against suffocation. Perception in these receptors is based on transmembrane and membrane receptor proteins that recognize individual chemical compounds. Activated cascade reactions lead to stimulation of the chemoreceptor cell, which transmits the signal farther, through sensory nerve cells to the hypothalamus, from where the signals are taken to the subcortical layers and the cerebral cortex. An imbalance in the supply of these feeding substances, especially their deficiency, results in an immediate feeling of growing discomfort, which turns into psychological pressure to obtain the lack of air, drink, or food. This is associated with unpleasant feelings that dominate the course of thinking and action. Emotions generated in this way can be treated as pain, in which we can distinguish both the physical component (reactions of internal organs, weakness, symptoms of suffocation) and the mental one (fear, despair, a feeling of resignation to suffocation). The inability to meet these elementary needs often leads to apathy, giving up further struggle to meet them, and ultimately to the death of the body.

However, these elementary emotions, felt directly with internal senses, do not exhaust the range of psychological responses to threats related to scarcity of resources. Before we feel hunger or thirst symptoms, along with the passing of time without replenishment, we can realize this threat, predict that the feeling of insufficiency will soon appear, and imagine the dramatic effects of maintaining such a situation. We may thus feel discomfort, forcing us to taking preventive action. This deliberate counteraction to the long-term effects of scarcity of necessary resources and long-term threats reminds us of the hierarchy of pains discussed in part I, when we analyzed the motivations for the action of an embodied system equipped with a conscious mind. We called them abstract pains, unlike primordial, primitive pains. Now we understand that each type of pain has a phenomenal component associated with direct feeling and an intentional component associated with understanding the wrong situation and predicting further possibly fatal consequences. The hierarchy of pain is related to the proportion of these components and their etiology. In primitive pain, the intentional component, for instance, is the knowledge about the location of pain. In pain with a high degree of abstraction, the phenomenal component is often limited to tightness of the chest, a “sinking feeling,” or anxiety about a difficult challenge. A description of tragic consequences in the event of an unsuccessful exam, loss of important documents, or missing the goal is an intentional component of higher-order pain. Of course, this description may be in the form of internal speech or may be a form of expiation. A similar hierarchy may relate to feelings of nice, pleasant sensations and satisfaction for the conscious mind. The whole hierarchy of pain and pleasure is used to motivate organisms and can be useful in motivating artificial minds to perform mental functions of a higher order, especially in the case of motivated learning described by Starzyk and colleagues (2012).

Pattern Recognition

Let's analyze whether we can use the available registration and information-processing systems as part of a brain capable of pattern recognition. Can an ordinary camera with a large memory recognize a quale or feature? Definitely not. The optical system of a camera can reflect the image in front of its lens on a photosensitive sensor. The image processed into electrical impulses can be stored in the memory, especially if said memory has sufficient capacity. However, there are no subsequent layers of processing of these impulses that allow for categorization of objects filmed by the camera. The system is not able to recognize objects and react to them. There is also no possibility of recalling them and taking action under the influence of memories or imagination. At this point the idea may arise that the camera could be connected not to the memory chips but to a computer, which will analyze signals received from the optical system and processed into impulses. Well, yes, but then we would have to establish the rules for this processing. Here, the complexity of the world around us is an obstacle. Imagine that our camera has only one hundred lines of one hundred pixels each, which corresponds to a total of ten thousand pixels. If we remember the content of each pixel and then find an image that closely matches the stored pattern, the probability of finding a pattern is equal to the inverse of the number of combinations that individual pixels can take in the image, and the number of different combinations of darkness of individual pixels reaches 100^{100}. This number is huge and exceeds the number of all elementary particles in the universe. A further increase in the number of pixels analyzed leads to an absurd number of combinations, making it impossible to even imagine that computers will ever be able to process such a large amount of information. Therefore, this situation has been called a barrier of combinatorial complexity. And yet our visual system provides us with information from over twenty million pixels. It is clear that such a system would be extremely ineffective, and therefore it cannot operate on this principle. Fortunately, computer scientists working on artificial intelligence systems invented rules that are much more effective for comparing and recognizing patterns.

Could it be that pattern recognition systems could not work? They can, and quite effectively, in fact. After all, we are familiar with industrial and domestic robots that perform their tasks splendidly. They move smoothly indoors. When they reach a wall or encounter an obstacle, they turn back or bypass it. They search for harmful substances and characteristic sounds, recognize and segregate objects on factory lines, and process and subject those objects to technological processes. Intelligent rockets reach their destination. More and more often, we interact with a whole army of intelligent robots. And yet they lack something that we would expect from them: their intelligence is very limited. We know that although they work well, they don't understand what they do. Although they perceive their immediate surroundings, they are not aware of them. They act like the most primitive animals: jellyfish, earthworms, and nematodes. They do not build in their minds a model of the surrounding reality and do not create abstract notions generalizing the recurring characteristics of a given class of objects.

They do not carry out categorization at higher levels of generalization. But they do not recognize qualia, because they do not create general notions and ideas embedded in the experience of manipulating the environment. They have no feelings related to the result of their actions. Their ability to recognize the features of objects and use them to perform simple activities is an effect of knowledge embedded in their computers by the constructors who created them. Their creators were those who equipped them with algorithms deciding their behavior. Unfortunately, designers are not able to predict all the situations in which an automaton may operate in a rich, natural environment, where other systems performing different missions may meet. Of course, such systems can be equipped with a certain learning ability, such as *Caenorhabditis elegans* has. Algorithms can be adaptive and can dynamically adapt to operating conditions. However, their behavior is predictable, and their reaction repertoire is limited. Even if they use knowledge of environmental geometry, their environmental model is extremely primitive and does not go beyond the reach of actuators, manipulators, or machining tools. We would never expect that awareness of such systems could go beyond perceptual awareness. The more primitive the environmental model created in their processors, the lower the level of awareness we are ready to grant them. The barrier of combinatorial complexity described above does not apply to these simple robots. They do not need an extensive environmental model and are not very capable of creating it. The number of parameters they control is very limited. This is sufficient for them to work, because they are very narrowly specialized. The greater the degree of specialization, the greater the limitations in action in the choice of ways to carry out tasks, and the smaller the need to get to know the environment.

It is widely believed that research on classical artificial intelligence (AI) is experiencing a crisis. Last century's great expectations of successes in the construction of intelligent and autonomous systems—that is, those operating independently—have not been met. Intelligent bullets, robots, and even self-driving cars can not satisfy us. But such a statement raises an obvious protest. After all, there are excellent systems to search for any information on the internet. The amazing DeepFace facial recognition system achieves 97 percent recognition efficiency, and the newest FaceNet reaches up to 99.6 percent facial recognition, even when the face is partially obscured and slightly rotated. There are systems that successfully recognize speech, license plates, and military objects. Unfortunately, the successes we know do not change the essence of things. All these systems require a preliminary limitation of the class of recognized objects. Their narrow specialization is still necessary. Military systems do not distinguish between faces, and facial recognition systems do not detect tanks or missiles. Although the FaceNet system can capture the human face in the picture, it cannot recognize any other objects. Worse, it cannot recognize a face that's in profile, distorted, unusually illuminated, caricatured, or in many other situations that natural minds manage without difficulty. Attempts to increase the universality of systems inevitably lead to an avalanche in the number of formulated hypotheses and necessary assumptions, the number of rules, and, as a consequence, the number of necessary logical operations.

Google's autonomous car performs well on the road using road signs, painted lines at the edge of the road, or markings of intersections and roundabouts. It also benefits from the fact that the vehicles encountered on the road and their behavior are largely predictable. However, as of now, the car cannot cope on streets and motorways that turn unexpectedly, and it will leave the road and head into a different area. Even worse, when it encounters unforeseen obstacles—sand, pits and hills, or a herd of irritated rhinoceroses—it does not know how to respond. But let us mention its achievements. The car's control system undoubtedly has an environmental model. It has the foundations of geometric consciousness, because it must know distances, angles, and speeds in relation to the surrounding objects. It must also anticipate the route of its movement and other vehicles in its surroundings. Does it not seem like a falcon planning a flight in a crowded space? Can we say that the bird has a geometrical consciousness? Well…it's a matter of definition. How can we determine its characteristics? Let us list the differences between the Google car and the falcon.

• The falcon has the ability to learn how to avoid obstacles. The instinctive reflexes built into his brain play a minor role in his behavior. The car mainly consists of built-in algorithms, and its ability to modify its reaction through learning is extremely limited.

• The falcon uses learned knowledge and its own experience. The car uses the knowledge instilled in it by designers and the knowledge from other intelligent beings providing data through satellite navigation, radio warnings, and so on.

• The falcon has a model of closer and farther surroundings and locates perceived objects on the map of the nearest operation and its neighborhood. This model is developed as it learns. The car has a partial model only of its immediate surroundings based on immediate "sensory" impressions. The model of the farther area results from the algorithms of using an artificial satellite navigation system.

• The falcon senses and understands its situation in space, comparing observation data with the current operation model and the neighborhood model. The car has no senses that allow it to feel anything, and without a model for a further plan, it does not understand its current situation.

• The falcon has motivation to perform the task. The car does not have its own motivations and uses the designers' motivations when executing the assigned task.

• The falcon has its own will to act. The car does not.

It seems justified that the lack of an environmental model that prevents understanding of the arriving stimuli and the lack of intrinsic motivation to act do not allow us to award the Google car, or many other similar machines, any degree of self-consciousness. The car can remember information about spatial relations, the direction, and the route. It can be trained to smoothly perform typical operations. For this purpose, it can compare the current situation on the road with the reaction patterns learned during the training. It recognizes features much as it can learn to recognize and create new patterns. However, it cannot create ideas, because it does not categorize these patterns. It does not generalize them, analyzing various characteristics of an object or an event in numerous situations. It cannot independently create the concept of "danger on the road" based on various experiences encountered while driving. Even if the engineers equip it with a warning system that signals danger, such as standard hazards like excessive or unacceptable speed or too little distance from approaching vehicles, that does not mean that the car computer will have the concept of "danger." This is because such a concept would not result from its personal experience. We would remind the system of all the constituent concepts, impressions, and feelings associated with this concept. What is the "recollection" of the processor in the car? Maybe just the path of warning-signal activation. A car without consciousness will have no reminiscence.

We have a slightly different situation in the case of more universal systems operating in a much richer environment. Let the facial recognition system be an example. The facial recognition task is related to the analysis of a much larger number of different objects than those encountered in standard situations on the road. The system, observing the faces presented to it, learns to distinguish more and more subtle details. The FaceNet system has been trained through the example of 260 million photographs presented to it. No man has a chance to get acquainted with even a small part of that collection. Undoubtedly, the distinguishing features allowed FaceNet to categorize facial features in many respects. Created categories are very specific concepts: a man's face, a woman's face, a child's face, a slim or obese face, a wide or narrow nose, a scalp covered with lush hair, a bald head, and so on and so on. Ironically, the system's performance is due to a full selection of faces and a richness of environment, as well as the system's enormous versatility, because no defect, no pimple, escapes its attention. It identifies the whole world of the faces. But do we accept this model of the world? This model is too narrow, too one sided, to suggest that the system is aware. This brings to mind an obsessed savant's model of the world. These types of minds are not aware of the real world and are locked in their own world.

Recall that the greater the degree of consciousness, the more complex the model created in the mind, the greater the wealth of details, and the more these details reflect reality.

We are talking about the adequacy of the model to the modeled reality. We are not ready to grant consciousness to a tapeworm, because although it lives in a very complex environment, its world is completely different from ours. Maybe we do not care how much the tapeworm is aware of its world. We will never recognize it as a conscious being, because we are interested in the consciousness of this world we live in, with all its wealth, to the boundaries of the known cosmos. Similarly, artificial facial recognition systems that have the ability to recognize many distinctive features cannot be considered to be intentionally conscious, although they identify and discover all necessary features, learn from them, and use them to categorize and recognize faces. However, this model of the world, though filled with the ideas of various faces, is too one sided, too detached from the reality we know, for us to consider it a conscious system capable of creating complex models of actual reality. Let us now consider the most advanced creations of classical artificial intelligence (AI), systems showing efficiencies exceeding human abilities, systems aspiring to be called intelligent because they are able to talk with people and because people may have difficulty recognizing that they're talking to a robot. Systems have been built that operate the most complex abstract concepts, prove mathematical theorems, and simulate the human psyche. There are programs that can play games like chess and go. Information centers use chatbots to conduct conversations in natural language and provide information to clients inquiring about the range of services. Expert systems support doctors, lawyers, and engineers, and intelligent enterprise management systems manage everything from financial investments to the battlefield. The British mathematician Alan Turing proposed testing for intelligence by having a conversation with a person through a computer monitor without disclosing whether it is a human or a program answering questions. The latest versions of systems capable of talking to people and using natural human language, or chatbots, seemingly pass the Turing test intended to show that the computer has intelligence indistinguishable from a human's. However, this only applies to amateurs who are not used to conversations with the machine. Any AI specialist can easily detect the tricks of these specialized robots.

A well-known computer program that has taken on Turing's challenge is CleverBot, an AI internet application that talks to people. Its capacity for natural conversation was tested by having it lead over a dozen or so thousands of conversations in the first decade of its operation. Since it was made available on the internet in 1997, the number of calls has exceeded 160 million. Everyone interested has access to the program on the internet and can see that the chatbot's answers and questions are very convincing. CleverBot constantly "learns" by collecting more data and perhaps raising the degree of "intelligence" that it seems to have. CleverBot answers are not programmed. The system "learns" based on data and answers obtained from people. The software is also successively upgraded. Competing with it is chatbot ALICE (Artificial Linguistic Internet Computer Entity). ALICE uses heuristic rules to match patterns to process natural language. Its ability to learn independently is very limited, because the system must have preprogrammed pattern-matching rules. Although ALICE is one of the best programs of this type, it cannot pass the Turing test successfully, because even an ordinary interlocutor can determine in a relatively short time that he is talking to the machine.

Another chatbot program installed in the IBM supercomputer named Watson won a million-dollar prize in the *Jeopardy!* game show, defeating the show's two most successful human players. The last episodes of the game show were watched by record numbers of Americans. *Jeopardy!* has no limit on scope of knowledge. Players must answer questions from any field. IBM programmers put the vast knowledge contained in encyclopedias, dictionaries, and literary works in Watson's memory. The DeepQA supercomputer they used to store and process this information is the size of several refrigerators and is based on the IBM Blue Gene machine. It has 2,880 cores and fifteen terabytes of operating memory. A serious limitation in Watson finding the right answers was that the response time was limited to three seconds. However, the supercomputer had a processing speed of 80 trillion operations per second, so it coped extremely well with this problem. We also remember perfectly when grandmasters of chess and masters of the game go lost games with computers. The groundbreaking moment was world chess champion Garry Kasparov's match with the Deep Blue computer in 1997, after which Kasparov acknowledged the superiority of the computer. In 2017 go master Ke Jie lost to the AlphaGo program, developed by DeepMind, and stated that the gap between humans and computers was becoming too great and that the program was improving too fast for him to beat it.

Does this mean that we are already able to build machines with human intelligence and having consciousness, perhaps even self-consciousness?

The artificial speech recognition and synthesis systems built so far do not meet the expectations of users or specialists. Highly specialized algorithms of known chatbots allow the program to recognize a significant number of words of a specific language, and other algorithms select answers and very often also questions that imitate a conversation in natural language. The expert systems integrated with the chatbot even allow useful information to be obtained as a result of such conversations. However, we realize that this is just an imitation of the conversation, because the chatbot with whom we are conversing does not understand its own statements. It is not taught except in a designated area. It does not have the creativity to give an answer that goes beyond the deterministically programmed repertoire of reactions to the questions asked. Often, the anthropomorphization of the chatbot, which prevents it from repeating in a strictly reproducible manner the answers to the same questions and thus more closely resembles a human, is attained by the introduction of randomized controlled elements of a random or chaotic nature to mask the algorithmically determined nature of its response. But chatbot conversations are just imitations of conversations. The Turing test proved to be completely useless for the detection of consciousness. The difficulties faced by designers of chatbot systems significantly exceed technical difficulties related to the speed of information processing, the capacity of the required memory, access to knowledge resources, or the ability to process huge data sets. The basic difficulty lies in the fact that understanding the language requires a grounding of concepts that are necessary to formulate answers in the broad, multidimensional context of general knowledge, and this requires consciousness. It is necessary to have a model of surrounding reality and sometimes the model of the whole world that surrounds us. What is needed is the ability to learn and, as a result of learning, the ability to constantly expand and deepen this model. In addition, the machine needs the ability to constantly expand the model as a result of acquiring new knowledge during the conversation. So far, no concepts have emerged that can explain how to create an algorithm that would give the system consciousness and the ability to talk intelligently.

Attention

In our mind, a mechanism is needed that chooses which in a group of ideas are significant at this moment. The criterion of significance (that is, the choice of what is important for the system and what is not) depends on the type of system and tasks it performs. Different criteria are set by the human's mind, by animals at a higher level of mental development, by simple animals, and finally by specialized robots designed by engineers to perform specialized tasks. The above-mentioned mechanism of selection of significant ideas is the mechanism of attention. It describes how the mind selects the contents of consciousness and what mechanism causes our thought to focus on one thread or object or to jump from topic to topic, looking for what is most important at a given moment. Receptor fields consisting of visual cells respond to changes in stimulus intensity. In a single pixel, these changes can be caused by a change in the intensity or color of the light. However, the greatest dynamics of change are related to the movement of the object. To the higher layers of the memory structure, excitations are sent from a huge number of visual cells representing numerous dynamically variable objects. However, those that come from the fields where the changes are the most dynamic have the greatest strength. That is why the retina of the eye is so very sensitive to movement in the field of view.

If the reader is involved in following the explanatory thread of an interesting theory, it seems impossible for him to give it up. In the next moment, the state of stimulation of the senses may change. The reader may be called to attention by the bang of a gunshot or a shout from a partner. The impact of the environment is a major impediment in maintaining focus on a selected mental thread. Companies use this to advertise in spots that promote their interest. When we open a new website looking for the information we need, targeted ads draw our attention. At some point, the body may begin to send intense signals of fatigue or physiological needs, such as eating or drinking something. Even then, our attention can fight for continuation by associating the planned break with the effects anticipated in such case. Maybe we associate the disturbance with the close date of an exam in a related subject? Maybe we will realize our ambitions to impress our company with knowledge, or we will remember other motivations guiding our intention? Maybe ordinary curiosity will prevail? Ultimately, attention switching is conditioned by competition in our brains. These stimulations have a complex dynamics, enriched by the processes of habituation, which is the extinction of sensitivity to long-lasting, monotonous stimuli and resensitization if these stimuli disappear. This is where our rich, often unpredictable reactions to stimuli that we are constantly subjected to come from.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

We must recognize that the entire brain is involved in the creation of a conscious mind, along with all of the sensory fields, the network that controls the movement (motor cortex), and all other brain structures. Consciousness arises as a phenomenon within the brain. We cannot see this. We cannot measure it in any way. This is an emergent phenomenon that is not reflected in any part of the brain's apparatus. There, at the level of individual neurons, we can detect some activity, some flowing currents or streams of ions. But none of these phenomena will tell us what thought is currently flowing through a particular memory cell, axon, or synapse, because such thoughts are not there. However, if this brain is equipped with effectors that can perform actions stimulated by signals from fields containing associations of ways to achieve a goal with specific intentions, then the system will achieve these goals. If, moreover, other effectors can generate speech and learn the language of symbols that the mind enmeshed in this brain uses, then we can find out what the mind of the system perceives with the help of its senses. An independently functioning system does not need a complex symbolic language. It is satisfied with labels that symbolize a set of memory traces associated with experiencing sensations caused by a given object. The language is useful when it comes to interacting with other systems and exchanging or supplementing the accumulated knowledge of an abstract nature. If the system has already developed such a language, it can tell us what its plans are and what it thinks about us and the world in which it had to exist, act, and be influenced. It will be able to talk about its experiences and emotional states. It will be able to tell us what its vision of the world is like, both of the immediate surroundings and of the farthest corners of its universe. It will also insist that it thinks, wants to pursue its goals, and is conscious.

Motivations for Action

It is not enough to have a mind. You still need to be able and willing to use it. An autonomous agent capable of intentional action must still want to do something. He must have motivation. Motivation to act in the animal world had to be created before the first effectors that allowed movement of the body were created. Previously, simple unicellular animals could only move with the current of water, grabbing food on the way. The motivation to obtain this food was a factor of selection in natural development, giving an evolutionary advantage to those single-celled individuals who were able to move toward the source of light, abundance of food, or a higher, more favorable temperature. It was only after the development of motivation-driven motor skills that the brain that could control movement could be created. Without motivation to act, it is impossible to talk about intelligence.

In a robotics joke, the creator of a conscious robot reports his first contact with the intelligent machine, saying, "She said she could think, but for now she has decided not to." In order to achieve consciousness, one must have motivations. It is widely believed that the main motivation for the operation of intelligent systems, including systems governed by natural, animal, and AI brains, is to avoid pain and seek pleasure. The most common source of pain is the scarcity of resources that could meet the generally understood needs. In a natural environment, there is always a shortage of resources necessary for expansion and evolutionary success. And these resources include, among others, available space, available energy sources, available sexual partners, and achievable security. In mathematical modeling using methods of artificial intelligence, this issue can be simplified by assuming that pain and pleasure are mathematical parameters, variously defined depending on the physical parameters describing the system's adaptation to the environment. Such generalized parametric pain sometimes comes down to one parameter, where the lack of pleasure is called a pain. It is to be expected that autonomous agents with a higher degree of consciousness will have more complex motivations. One of the situations when such more complex motivations may appear is the case when the agent feels uncomfortable. It can be said that it "worries," predicting problems in satisfying its needs in the near or distant future.

Janusz Starzyk (Starzyk 2008; Starzyk and Prasad 2011) presented an interesting concept to find higher-level motivations. The concept he presented recognizes reduction of generalized pain as a need requiring fulfillment or satisfaction. Therefore, it is a fundamental motivation for the operation of an intelligent agent—creating foundation for motivated embodied intelligence. By interacting with the mechanism of attention and working memory, pain-driven motivation allows the agent to formulate goals and plan smart actions that lead to satisfaction from pain reduction. However, direct actions to avoid pain can be impossible for various reasons. An intelligent agent can learn to predict that although the pain does not currently occur, it may occur in the future. Complex situations that make it impossible to avoid direct harm, which the author calls primary (lower-order) pain, and anticipation of discomfort in the future lead to creating abstract (higher-order) pain. According to the example given by the author, seeing an empty refrigerator may cause discomfort in predicting hunger. Similarly, looking at an empty wallet may lead to the painful idea that the fridge will remain empty.

This process creates a hierarchy of pains from the original, sensually felt pain to abstract pains, which are subject to logical analysis and reach the conceptual-symbolic layer of our consciousness (Starzyk 2011a; Starzyk et al. 2012). Similarly, a hierarchy of psychological needs may arise due to the lack of resources necessary for survival. Guided by these abstract motivations, we should generate a wealth of behaviors appropriate for highly intelligent, conscious beings. The above mentioned motivation system is designed to give a “sense of existence” to autonomous intelligent robots. And indeed, pain avoiding behavior is observed in intelligent, autonomous systems (Haikonen 2007; Nuxoll and Laird 2004, 2012). Well, a skeptical reader may ask, but from where does a full or even empty fridge arise in the mind? We must assume that a system with a relatively high degree of consciousness has different means of satisfying the needs at his disposal—those needs that require immediate use of available resources, and those that involve planning for a longer or shorter time and anticipating the effects of long-term action. Therefore, the environment must be sufficiently rich and complex. If we do not provide an arsenal of sophisticated tools for complex interactions with the environment, then achieving a high degree of consciousness will not be possible. The model of the environment that the agent creates will be primitive, inadequate to the real challenges created by the dynamics of processes in this environment. Let’s be direct. In every environment with limited resources in which independent systems/beings operate, beings that want to exist and develop, there is a fight for their survival and a fight for resources. A primitive, intellectually limited mind has no chance in this fight. Even if our computer had a cognitive brain, if its tasks are limited to the role of a typewriter, it will not reach consciousness or surprise us with its intellectual level.

We have written a lot about the numerous motivations imaginable in an imagined artificial intelligent system. Living organisms have even more motivations, drives, and instincts that emphasize action, that there is an unlimited palette of emotions affecting an individual’s reaction and the dynamics of operation. Living organisms, especially people, are emotional individuals. Without emotions, they would cease to be animals or people! When autonomous intelligent agents, which currently are living organisms formed during the course of evolution, act intentionally, they must satisfy their needs. Any shortage of resources or lack of ability to meet needs leads to strong emotional states that drive the energetic action.

But could not it be simpler? Can’t you just write a simple algorithm that would force a lazy agent to act to meet its own needs? Or maybe it would still be good to program it to do something useful for us, its constructors—its creators. Hmm…we have a problem. We cannot do this. There is no processor in the brain that we could program. What we can do is to construct a network of connections between neurons in such a way that, in critical situations of threats or deficiencies, they change their characteristics—speed up the operation, increase or decrease the thresholds of sensitivity to stimuli, and so forth. We can also block access to some processes to increase the processing power of those that are more important at a given moment. We can also increase or decrease the range of associations to focus on a specific problem or to look for more complex solutions. This type of network connection manipulation is far reaching, sometimes including the entire system (we call them the “global” interactions). They appear as a tendency of the autonomous system toward a modified action, which can be easily identified as a specific drive or instinct, and we can interpret the psychological reactions associated with this as artificial emotions. Our system will become an emotional being. This is an important step towards the sought after model of the motivated emotional mind (MEM), which is to settle the brain of an embodied, autonomous agent.

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

If bodily experiences are to give meaning to what an individual perceives and remembers, then he must have the ability to evaluate information—that is, to assess the value of knowledge for the goal that the system is to achieve in the short and long term. However, logical analysis of scenarios using available sources of knowledge is not enough for a sense of consciousness. We must make a strong distinction here between information processing and feeling. While the processing of information may be unconscious, the feeling is always conscious. It is the essence of “feeling.”

Phenomenological considerations about the role of feeling in mental life allow us to define the role of feeling as a sense of the significance of information processed for the system that processes this information. Let us note that this is not about *understanding* what this meaning is but about *feeling* this meaning. It is about knowing that what we are feeling is absorbing that something important is happening and that we are participating in it. One of the most important feelings is the feeling that we know and understand something.

In general, we recognize when we know something for certain. We know that the sun shines, that we are warm, that we are in our home or at the park. We know this because we are doing it and feeling it. This was noticed by the neurologist Robert Burton and described in his book *On Being Certain* (2008), where he writes that the feeling of certainty that we know something is a mental feeling rather than a logical proof of the fact's existence.

Much evidence using brain imaging indicates that the confidence of knowledge arises in the deep, primitive structures of the brain, not in the areas of the cortex responsible for logical reasoning. This feeling is largely independent of conscious reflection, contemplation, and analysis of facts. Rather, it results from associating the current state with previous experiences, indicating that the current situation is safe and that there is no time or situation pressure. This indicates that the agent recognizes and controls the situation well. Such a sense requires gradual education in the course of many life experiences. It usually doesn't appear when the situation is new, dynamically changing, and stressful. This feeling must involve large areas of the brain containing memories of many dramatic or pleasant experiences. Feeling is a process in which significance is attributed to information through the global impact on the information-processing system in the various ways mentioned above. Such global long-range impacts affect the structure and dynamics of the entire system. It is thanks to them that we can consciously experience emotions, feel qualia, and gain a state of understanding.

Free Will

Most people think that they have free will. In particular, we have the unwavering feeling that we ourselves personally have free will. Sometimes we hear from people that "something" seemed to direct them, and they did something they didn't really want to do. However, this happens extremely rarely. There is a widespread belief that we are able to decide what we want to do or what we want to think about. We think that in all circumstances we can freely decide what we consider moral and immoral, what values we have and follow in life, or what goals we want to achieve on a given day and how we plan to do it. People treat free will as inseparable from their person. Even if they are in captivity, they will claim that their mind is free, that they can decide what they are thinking and what they are planning to do.

Criminal law assumes that we all have free will and act in accordance with it. All legal systems assume that a person is guilty if he has violated the law of his own free will. Without his free will, there is no guilt, punishment, or restitution. Criminal law considers reduced liability for a crime if it was committed by a natural person whose free will was weakened by medication, insanity or strong emotions. Free will suggests that, depending on our will, we are able to reflect on the effects of our actions, which in turn is associated with consciousness. Belief is not the same as feeling. We may have a "free will," but someone could argue persuasively that this subjective feeling of freedom is an illusion, that we don't really have free will. How is it really?

Free will means that we can act according to our own desires and preferences. We feel full control over our own mind. If someone were to convince us that this was an illusion, they would have to point to any physical, social, or psychological restrictions on freedom of action. The development of psychology, research on the biophysics of brain processes, and the development of social and economic sciences actually find serious reasons to question the freedom of decision making and activity consistent with what we want.

Let us ignore here unscientific views that come from ancient religious messages convincing us that our fate rests in the hands of the gods, who control our actions or predestine us to fulfill the missions for which we were created or appointed. The great monotheistic world religions admit that God has given people free will but maintain that he can intervene and change human destiny at any time, especially if he listens to the individual or collective requests of the faithful. Free will is in this case the result of negotiations with a divine being, the result of which we cannot predict. We cannot influence what calamities or favors a divine being will send us.

Dualists take a different position from religious believers. Their opinion that we have free will is connected with the duality of soul and body. According to this view, our immaterial spirit is free from the restrictions imposed on it by our limited body and can emerge victorious from any oppression, preserve the morality of actions, and make the right decisions. From this, in contrast, it also follows that a machine without a soul cannot make an independent decision but only responds mechanically to the stimulation and memory content. Its actions are predictable and deterministically decided. All one has to do is to know the state of its neural network and memory structure to predict what it will do or think about. Once again, the (human) mind triumphs over matter.

But if so, then the machine cannot be conscious. What are we actually trying to show here? Adopting a dualistic position completely undermines the sense of studying brain processes as a substrate for the mind. Contemporary cognitive science, which is the study of the brain and how the mind explores the world, claims that thinking, planning, and making decisions are the result of biophysical processes that occur in neurons and other brain material. This is at odds with the age-old belief that thinking, feelings, and beliefs are the result of the action of an immaterial soul that saturates our passive bodies with thought. According to these traditional views, it is the soul that gives us limited free will, which allows us to be responsible for our worldly deeds. This makes us aware and moral people. In chapter 9, we will try to reveal the source of this type of mystical belief while retaining the scientific materialistic explanation.

The most dangerous attack on free will was made by supporters of physical determinism. They also include many cognitive scientists studying brain processes. The deterministic position of cognitive scientists causes serious worry to philosophers. This worry is not due to the scientific position contradicting religious beliefs. The reason is more fundamental, and it is only materialistic. If thinking is the result of the biophysical processes of the matter the brain is made of, it has dramatic consequences. After all, every particle of matter is subject to strict physical laws that determine precisely what the effects of each physical process will be. The same causes must lead to the same results. And one physical event passes smoothly into the next, because all material particles are subject to constant motion. *Panta rhei*—everything flows, according to Heraclitus. Therefore, we have entire cause-and-effect chains, in which the final state results from what we found at the beginning. So, if we know the initial conditions of chemical particles in synapses connecting individual neurons and in the entire brain matter, and if their state is associated with some thought, then from this initial state and the laws governing matter, all the next thoughts that will ever pass through our mind arise. And because thoughts direct our body, all our behavior will be strictly defined. It will be determined by the laws of physics, boundary conditions, and the initial state at the beginning of our existence. According to British physicist Stephen Hawking (2010), "our understanding of the molecular basis of biology shows that biological processes are governed by the laws of physics and chemistry and therefore are as determined as the orbits of the planets… ., So, it seems that we are no more than biological machines and that free will is just an illusion."

This means that we cannot do what we want. Or rather, we want what has already been established in advance by the laws of physics. We cannot avoid what is intended for us.

Does fate, destiny, thus obtain a scientific explanation? Are the sentences of fate irreversible? That would be a dramatic conclusion for materialistic thinking. What is worse, the above conclusion does not result from blind faith in destiny, fate, or an inability to avoid a quirk of fate. This belief results from the consistent application of the laws of physics and seems to be scientifically justified. It seems that rejecting the determinism of physical phenomena contradicts all our scientific knowledge. So if we have no choice, it means no free will. We cannot change our fate in any way! We must consistently accept the idea that providence is irreversible!

Even worse, other sciences studying human behavior point to further restrictions on free will. In choosing a way of life, we can be limited by legal deprivation of liberty, applicable laws, culture or moral norms, habits, lusts, dependence on drugs or alcohol and their neurological effects, or mental illness. These restrictions limit our freedom of speech and deeds, but are they also restrictions of our free will? Studies with laboratory animals have shown that all these factors can affect free will. A rat infected with the *Toxoplasma gondii* parasite changes its behavior. When the parasite passes into the rat's brain, the rat loses its aversion to cat urine and thus is less likely to identify areas where cats may be present. Infected rats are then more likely to be eaten by cats, and the parasite enters the cat's intestines, where it can reproduce. In this way, the rat's free will was affected by the parasite's blind-reproduction mechanism.

Similar effects of disease occur in humans. We know how much illness or disability can limit our lives. Stephen Hawking, from the age of twenty-one, suffered from lateral multiple sclerosis, which led his body to almost complete paralysis. What could be the scope of free will of a person who is unable to exist without the care of doctors, nurses, and loved ones? An alcoholic does not show free will if he wants to overcome his habit but cannot because of the alcohol-related changes in his brain. He must heal this addiction if he wants to regain influence on his fate.

The fate of people as social beings harnessed in civilization modes can be determined by the place and time of birth and their financial status. Our life options depend on where we were born, who our ancestors were, our education, our living conditions, and so on. A four-year-old Peruvian child who lives on the street, begging for food, doesn't have much choice. He is forced to beg to survive. Where is the free will in it?

The resulting lack of free will seems to free us from any moral responsibility. So how can we counteract evil or disaster? How can we prevent lies if we lie not because we want to but because the circumstances and physical and psychological processes in our brains have to go as they go? Therefore, we should not be responsible for crimes against the law and for our sins against God. This, in turn, seems to contradict our sense of responsibility, social norms, and the internal conviction that we are not totally helpless.

Psychologists point to the psychological reasons for limiting free will as another and the most important factor depriving us of freedom of choice. Our decisions depend on the momentary mental states in which the mind is in the process of deciding. They depend on the content of what we have in our memory at the moment, our mood, available knowledge, and our patience for analyzing different variants of behavior.

All assessments of our free will are subjective. If I want to prove that I have free will, I could, for example, decide that right now I will go to the kitchen and make a cup of tea. What does this prove? Well, I think I decided to do so, and that means I have free will. If I get up from behind the desk and actually go to make a cup of tea, will it be proof that I have free will? The question still remains: Why did I make this decision? Was it not to prove that I have free will? In this case, my decision would be based on specific premises: it had its cause in the decision-making process, and it was guided by the goal of showing that I can decide what I want. The choice was to make tea, but it does not matter what action I decided to take. I chose something that I am able to do and that shows I have free will. This is how our brain works (see part III)—searching for how to meet our needs, it chooses those options that are within its physical or intellectual capabilities.

One may suspect that the free will we have the illusion of is the result of a conscious realization of the available options and an understanding of how they were chosen. "Freedom is the recognition of necessity," wrote German philosopher Georg Wilhelm Friedrich Hegel (quoted after Engels 1962). The fact that all these options are determined by our genes, our past, and the motivations developed throughout our lives does not matter. Thus we reach the following conclusion: there is no free will in its canonical, nondeterministic sense.

Let's try, however, to challenge this counterintuitive conclusion. First, the expectation that cognitive science and neurology will remove all restrictions on the freedom of decision is unreasonable. Most people recognize that absolute freedom does not exist. Therefore, we have to find gaps in the barriers of thinking, presented point by point, to find a place for (albeit limited) free will and freedom of choice. Let's start by dealing with physical determinism, which creates the hardest, most fundamental restrictions on the freedom of thought.

Defending the idea of free will and at the same time wanting to remain materialistic, John Eccles, who received the Nobel Prize for his work on synaptic signal transmission, speculated that the immaterial mind affects the material brain through quantum phenomena in synaptic connections (1994). Penrose and Hameroff seem to agree with this view, suggesting the quantum principle of brain function (Penrose 1998). It turns out that physics itself suggests a solution and breaks the close dependence of the effect on the initial state of the phenomenon. Many cognitive scientists point to the probabilistic nature of mechanics and quantum electrodynamics, which at the most elementary level describes all physical phenomena. So the final state of the process cannot be clearly defined. It may be interference by various quantum states. The effect of a set of causes cannot be determined with finite probability. We don't know anything for sure. Maybe this is a place for free will. Unfortunately, in the vicinity of large conglomerates of proteins and chemical compound chains, quantum states collapse immediately (decoherence), and the calculated probabilities of specific processes take values very close to zero or close to one. The range of discretion resulting from the quantum nature of physical processes would therefore be extremely narrow. We need much more flexibility and freedom of decision making. The pure randomness offered by quantum physics would not give us the kind of free will that is worth having.

It seems impossible to maintain the position that the objective reality in which we live completely determines our fate. We have many examples of individuals managing to break the limitations created by objective reality. Thanks to stubborn efforts to get education or valued skills, many manage to move beyond the level of the seemingly unattainable. Often, chance or a happy marriage is the deciding factor. Changing one's place or lifestyle can help. Neglected children often show more determination to climb the career ladder than their privileged peers. An addict may give up his addiction. Diseases can be cured or plans realized regardless of the disease's destructive impact. During the fifty years of his terrible illness, Stephen Hawking wrote and published over twenty books, many of which are best sellers worldwide. He got married twice and had three children and four grandchildren. Almost everyone has had times in their lives when they could make decisions that completely changed their lives. Indeed, there are moments in everyone's life when the field of possibilities opens (though not always as wide as they would like). But it's hard to argue that our fate is irreversible and nothing can change it. However, doubt arises here. If these moments decided our fate, then we return to the third reason, a psychological one, determining decision making.

In the decision-making process, we are guided by current knowledge, current needs, drives, preferences, current mental and emotional state, awareness of our own capabilities, and assessment of the current external situation. The decision also depends on our intellectual disposition and the content of working memory—what we are thinking about and what we are paying attention to. These premises are interdependent, because each of these factors can affect all others. In addition, they change from time to time depending on the time dynamics of individual processes. However, the decision depends on the momentary balance and dominance of the scenario and takes into account all these factors. Free will in a nondeterministic sense must seem an illusion. Our mind will be influenced by these factors, not others, many of which are independent of our mental efforts, and will decide to act independently of what we call free will.

Many physical processes take place in a chaotic environment. Particles of matter—for example, the water particles of a descending waterfall—are often found at the bifurcation point, where the choice of the way ahead is ruled by chance. This is the theory of indeterministic chaos processes. However, in our minds used to cause-and-effect thinking, the temptation arises to state that apparently the molecule did not have freedom of choice. This position is fatalism, the belief that if something happened, it must have happened. This belief is also mystical. If we look for the reasons for this phenomenon, there will be too many of them to be able to draw any conclusion justifying this and no other behavior. Post factum, we can prove that the combination of all factors forced a specific course of events. But this is only our expectation and our hope. If the process is highly chaotic and of an indeterministic nature, then we will not be able to indicate any way of linking the dramatic effect of water molecules with the forces that caused it. There are no theories, no equations that can be solved. The result of the process will be unpredictable.

The rules for creating mental patterns by neural networks are associated with multiple nonlinear processes that generate chaotic results. For philosophers looking for an orthodox doctrine that frees biology from determinism, this may be an insufficient argument, because there may be a suspicion that there will be a so-called deterministic chaos, in which we will not be able in practice to anticipate the results of the process in the long term, but it will still be possible in theory. As supporters of this doctrine prove, deterministic chaos is unpredictable, because we do not have sufficient knowledge about all processes affecting the analyzed phenomenon. However, if we had such knowledge, any uncertainty could be removed, and the results would be predictable. The real world does not give such satisfaction to philosophers. Mathematicians studying chaotic phenomena have found that in the case of strongly nonlinear processes disturbed by noise, chaos takes a nondeterministic form. One of the important modern mathematicians, Charlotte Werndl, proved that chaotic phenomena of this type show complete independence of end states from the initial state (2009).

This applies in particular to processes of a high degree of complexity in which there is a blur of parameters and nonlinear interaction of many elements. Biophysical phenomena in neurons and synapses show all these features, and in addition they are constantly disturbed by the influence of external factors. These factors include all environmental influences, both influences of civilization in the form of chemical, acoustic, and electromagnetic pollution of the environment, geophysical processes resulting in shocks and vibrations of the Earth's crust, and atmospheric influences through changes in humidity, temperature, and pressure, as well as the impact of gravity of distant planetary objects, cosmic radiation, and all-penetrating neutrino streams. Many of these processes exhibit indelible shot noise. What's worse, all these processes and many others not listed here occur simultaneously, which in the presence of interdependencies causes unpredictability of the results of information processing and thus thinking. This is an immanent feature of such a complex structure as the brain. Therefore, one can assume that we will generally not be able to predict the decisions an individual makes.

Even if you found a system that wasn't deterministic, there is no particular reason to say you'd found free will. You can say that your mind is either determined or not. If it is not, then you are just a random thought generator, nothing of any particular value. On the other hand, if your thoughts are determined, then all the forces of the Universe have conspired since the Big Bang to make you what you are. Quite a fabulous thing to say about you (David Vogel – private communication).

The impact of momentary mental states on decision-making processes remains similarly unexplored. The decision-making process involves comparing mental correlates of the scenarios of planned activities. Decisions are made on the basis of comparing unpredictable neural correlates of real and imaginary objects. By analyzing a posteriori how correlates are created, it is possible to explain how they formed and determine that the system had no choice but to make one or another decision. However, *a priori* such a prediction is not possible. The system can analyze the problem for any length of time and recall an infinite amount of data, impressions, premonitions, and suggestions, changing the weight of the correlate. Its decision is not predictable and determined. So the feeling of "will of the system" is most justified.

This is confirmed by psychological experiments conducted using brain-imaging techniques. These experiments showed that brain activity typical of a conscious decision-making process is observed around two hundred milliseconds before initiating the activity with which this activity was associated (Ullman 1996; Crick 1994). Ordinary reflexes do not require consciousness. Therefore, all these processes take place in the subconscious. Admittedly, later studies have shown that the main peak of behavioral activity of brain fields can last for a few seconds and falls shortly after reporting the realization of the task being carried out.

Hypothetically, the reward system could also be activated to assign the value of the planned response, comparing long- and short-term effects. Simultaneous activation of the limbic system can give an adequate degree of emotional involvement. Premotor fields will be involved to plan the necessary moves to place them under motor control. All these activities can take place subconsciously, which raises suspicions about the lack of free will.

These suspicions contribute to the confidence of many researchers that there is no free will when we consider the numerous results that contradict Ulman and Crick's research mentioned above, showing that motor centers are stimulated in advance of the decision to move. Hans H. Kornhuber measured the activity of movement centers occurring half a second before the signaled decision of bending a finger (1984), and Katharine Baker and colleagues showed stimulation of movement centers even two seconds before an informed decision to take action (2011). If the decision to perform an action takes place without our consciousness, then how can we even think of free will? It must be an illusion. Repetitive experiments seem to indicate that we have no influence on what we do! At best, some scholars allow a kind of "veto," that is, stopping the action. In clever experiments they have shown that the mind, after initiating the reaction, has about two hundred milliseconds to modify it or, as Libet called it, report veto. Libet and Mele also postulated that this time-delay period could allow for learning and improving reactions (Libet et al. 1983; Libet 1985; Mele 2005; Kornhuber 1984; Dennett 2005).

In a worse case, it was speculated that consciousness would only be "notified" of the movements made, and its function would be "reporting" of completed tasks. But who would receive this report? There is nobody there. Another fundamental question arises: Who makes decisions for us? All misunderstanding is related to a misunderstanding of what is "we" and who we are. Yes, there is really no mythical free will. We are our mind.

What role can the subconscious have in controlling decisions, and what is the role of conscious processes? This question is not easy to answer. Despite many psychological and neurological experiments, there is no clear answer, and many subsequent experiments will have to be planned and interpreted to unravel the tangled decision-making paths. Let's try to trace the way in which consciousness can influence the subconscious, to what extent it can shape this subconscious. Well, motor responses to stimulation can be initially shaped in the subconscious mind for a time depending on the complexity of the task. Studies report times from 0.2 seconds, in the case of simple finger movements, to periods of many seconds, in the case of the need to make a morally relevant choice. At the same time, in other fields of the subconscious mind, mental states may be prepared by defining goals or assessing the effects of actions. These states can be realized in working memory before a behavioral proposition is formed under the influence of said stimulation. In consciousness, there can be a fairly quick comparison of the patterns of stimulation representing these two mental states.

Therefore, the role of the subconscious would be to prepare proposals for behaviors fighting for access to consciousness. While the subconscious is preparing a response proposal, the time to "consider" consciousness preferences would be specified. Comparison of stimuli and a possible change in synaptic coupling patterns would result in acceptance of the reaction pattern or preparation of a modified proposal. In this way, the conscious mind would influence which of the conclusions of experience would be appreciated and consolidated and which would be ignored.

It is also one of the ways in which emotions influence the process of remembering important events, consistent with the conscious values established in the mind in the course of experiments. It can be presumed that consciousness interferes with the thought processes related to the choice resulting from conscious logical analyses; decisions; beliefs on the basis of worldview (models of large-scale integration) in the fields of politics, science, religion, morality, and so forth; conclusions resulting from studying works of science and culture; decisions when planning for the farther and nearer future; and many similar issues. The subconscious–working memory–consciousness relationship allows for conscious control of the system operation according to the system's will. However, this is not a mythical "free will," because it is limited by this relation.

American neuroscientist William Klemm (2010) writes, "In learning a new skill, such as playing the piano, there is no way that the subconscious mind can control the initial movements (fingers), because there is no way of knowing what to do. Only the conscious mind can choose which keys to press, because only it knows what should be done. If this is not free will, what is ?" But we must deny this intuitive statement. Consciousness's involvement in the choice of piano keys does not prove free will. It's just the use of knowledge after a moment of deliberation. The conscious mind can show a willingness to learn to play the piano if it has such a need generated by relationships with society or the influence of authority. However, the mind's will in this respect is not free. It is conditioned by understanding and by taking into account all the conditions of the learning process and the conceptual, associative, and motor apparatuses it possesses.

This does not limit the richness of our mental life and shaping plans for the future. Our thoughts are not a foregone conclusion, although the paths of thinking are gradually being fixed. Some basic foundations of the worldview are shaped throughout life. The effect of gradual stabilization of beliefs, views, and values in the minds of maturing people is known. It involves the process of so-called myelination of nerve connections. Myelins are axol glycolipid coatings that greatly facilitate and accelerate the transmission of electrochemical signals in axons. This process takes place gradually starting in infancy and includes the most intensively used neural pathways. The myelination process ends at age twenty-two to twenty-five, but changes in myelin thickness up to age thirty have been observed in some frontal areas of the brain. Once the brain reaches maturity, axons in the frontal lobes are protected, which are responsible for cause-and-effect reasoning, judgment, and planning. Presumably, just myelin deficiency in these areas is the reason for teenagers' inability to make mature decisions. But even if there is no pathological degenerative process, we see older people stiffening their views. It is for this reason that it is so hard to convince an elderly person to change his rhythm of life, clothes, nutrition, or political preferences. All this causes permanent restrictions on the scope of choice of behavior and the horizon of decisions taken. This is the framework that limits the free will area of each individual person. No one can demonstrate completely free will; everyone has their own range of decisions determined by the memory of past experiences, professed system of values, and worldview.

We can now try to define what we mean by the will of man, animal, or robot. The will of the system (machine, human, or animal) is to act according to its preferences. These preferences result from the interpretation of past experiences shaping the system of values and beliefs, the results of logical analyses of historical, current, and future situations, anticipation of the effects of actions, and the emotional state of mind. We can't force preferences. For a prisoner, sitting behind bars will not be his preferred behavior, unless he realizes this necessity and accepts the seemingly unbearable situation. Bondage takes away free will when it is not accepted. We treat disability, financial and family situations, or illness as a God's will, as long as we rebel against them.

Mind flexibility allows you to accept the most severe judgments of providence. It keeps balance between the rationalization of opportunities and limitations and between positive and negative emotions. The manifested preferences contain a strong component of the subconscious influence resulting from emotions. However, free will cannot be imagined without interference by consciousness. Both of these spheres are connected. The influence of emotions can be realized and mitigated by a conscious imposition of restraint in their expression. Logical conclusions can also arouse in us emotions that will affect the choice of response.

If we decide to bet a large sum in a poker game, then fixed patterns of behavior will allow us to make routine movements to move chips or money to raise the stakes. It happens almost without awareness. We can now smoke a cigarette, drink, or exchange comments with other players at the same time. However, attention remains tense. Any disruption to the typical course of the game can change our decision. While our subconscious has prepared the motor system to perform the action of adding the amount to the pot on the table, consciousness works out the decision whether to raise the stakes, by how much, and how to do it so as to confuse the players and win the pot.

Various psychological factors control these choices. Decisive factors can include a cool analysis of card strength and calculating the probability of the occurrence of winning systems, estimation of results of the game-play system used, habits of applying specific betting rules, calculation of one's own financial resources and the consequences of losing, superstitions that order or prohibit certain betting choices, determination to play, a cavalier mood or willingness to bet against all rules, anger at a bad card or the luck of other players, the discomfort of sitting at a cluttered, dirty, or crooked table, and so on. Emotions mix with considerations, habits with rational arguments. The final decision is the result of the deliberation of a motivated emotional mind. The result is not a foregone conclusion. Until the last moment, no one can predict the final decision, including the deliberating player. The player will do what he wants, what will remain in his consciousness after considering various factors. This deliberation and choice are manifestations of the player's will.

Our thoughts and decisions have not been written in the stars, and we have the opportunity to make the best decisions at the moment. We have our own will limited by the circumstances but according to our momentary mental state. What's more interesting, machines that in the future will be equipped with emotions and the ability to learn independently in a complex environment will also have a will. Therefore, they will be conscious and capable beings, and therefore moral ones. We will have to give them the same rights of conscious beings that we give ourselves.

The will allows us to focus attention on the issues that interest us, so in fact, if we want to do what we want and think about what we want, we must have our own will. Selective attention must be guided by consciousness, and before we make decisions, the very choice of the object on which we focus our attention is a volitional process, that is, one guided by our will. Psychological studies have shown that man shows the capacity for intentional self-control, which is equivalent to following his own will (Holland et al. 2013). As described in chapter 3, conscious, embodied intelligence has motivation to act. This, in turn, create emotions that, from a psychological point of view, help focus attention and concentrate on the perception and learning process.

As we know, focusing attention is necessary for effective memorization. As German Austrian neurologist Lüder Deecke writes, "Without attention we can remember very little and recall even less" (2012). We know from our own experience that events accompanied by emotions are remembered extremely effectively. Neurological research and visualization of the brain using the latest brain-imaging techniques indicate that during remembering, the limbic system sends pleasant impulses to the hippocampus, strengthening the memory. This would be an experimental confirmation of the thesis about the controlling role of emotions in remembering. We are dealing here with a weaving of interdependent functions managed by consciousness and emotions. In addition, attention would be driven by needs, emotions, and consciousness. To make matters worse, all this seems to contradict our considerations regarding the functioning of the attention mechanism described above, where we pointed to automatic attention to the elements of novelty and dynamics of the observed scene. We must assume that the process of focusing attention has an involuntary, instinctive, and intentional component.

Let us summarize at the philosophical level what we call the will of the system. We already know that there is no system that has free will. However, the will of the system is expressed in the possibility of making a choice. The moment the decision about the choice of a given scenario is made, the mind recognizes it as a desire for something and hesitates before making a decision.

As we described in previous sections, in the course of life experiences in the subconscious mind, our worldview, our system of values, is created in our memory. Life experiences accumulate, allowing accurate prediction of the effects of our actions. Decisions are made as a result of competition and their association with the system of values and with anticipated effects. A system capable of defining its own good is an intentional being. Such beings make decisions that are best for them from the point of view of the good of the system. So only intentional beings can demonstrate their will. Intentional decisions are not accidental. The choice depends on the experiences of the whole life of the intentional creature, and their life experience is strictly individual and unique. A volitional act is determined psychologically and is felt by the desire to achieve the intended result. In this way the will of the system is manifested.

The presented reductive model of the conscious mind has shown its use in explaining whether conscious systems are capable of making decisions and to what extent their decisions are "free." The great advantage of the model is removing the suspicion that mental phenomena do not remain in causal relations with physical phenomena. It seems obvious that our thoughts, desires, intentions, and decisions can be carried out according to our will thanks to the movements of the body, hands, and other members, as well as voice commands, nonverbal signals, gestures, and other means of communication.

The problems of mental causation (Fodor 1989; Kim 1998; Davidson 1980) resulted from the difficulty of indicating how the mental sphere can causally affect the physical realm. Our model explains these doubts in two ways. First, it challenges the hypothesis of mental anomalism, the supposition that mental phenomena are not subject to any causal laws, that our thought soars freely without being bound or limited. Mental anomalism is very conducive to the idea of free will. Of course, the phenomenon of mental saccades introduces the element of randomness of directing the thought sequence and as a result of making decisions. The chaotic nature of neural processes also interferes with the ability to predict the effects of thought and decision processes. However, it is a factor of uncertainty at the level of deterministic chaos, which does not destroy the main attractors and does not disturb causal sequences defined by the laws and rules of the material world.

Consistent reasoning leads to the conclusion that feelings and phenomenal experiences accompany physical processes in brain tissues. In this way, we used the hypothesis of the mental sphere's supervenience on the physical realm to explain the causal relevance of the mental sphere, challenging the assumption of anomalism, just as Davidson (1993) argued. As a result, mental phenomena appear to be epiphenomenons of physical processes. Such an epiphenomenon is the feeling of "free will" arising in the minds of beings who are accustomed to thinking that their inner self, and therefore they themselves, decide their actions. In fact, it is not a momentary sense of self but the entire sphere of subconsciousness and conditioning in which the body's internal and environmental conditions are involved in elaborating decisions by the conscious system. It interferes with the system of values, beliefs, and knowledge that is the content of consciousness. The relationship of the subconscious mind to the determinants of the current states of physical reality is developed during lifelong experiences. In owning our decisions, we are responsible not only for the decisions resulting from the logical analysis of the situation, our knowledge, and conscious beliefs but also for the influence of the subconscious, our drive or patience, our rage, our reflexes and associations, and shaped patterns of behavior. You have to live so that your psyche is not able to make bad, unethical decisions. But not everyone succeeds, and it doesn't always work out.

For the mental stability of a thinking creature, a sense of confidence and a sense of understanding are important. Searching for these states becomes the same need as other ones related to the lack of resources. Unmet needs are motivations for action reinforced by accompanying emotions. It is this set of features that directs the actions of an autonomous agent.

%%

Consciousness and Intelligence

In the previous chapters, we made a distinction between intelligence and consciousness. These terms seem to refer to similar mental properties. If a being has a high degree of consciousness, we suppose it must be very intelligent. Conversely, intelligent people should probably show a high degree of consciousness. This common belief indicates that these concepts are interrelated in some way. However, they relate to slightly different spheres. Intelligence is related to action. An organism or artificial autonomous system can behave intelligently or quite stupidly, unintelligently. Intelligent behavior leads to the achievement of goals formulated before the system or that it sets itself. These goals self-formulated by organisms mean adaptation to the environment, optimal use of its resources, and maximum use of the body's capabilities. Effectiveness in achieving a goal means surviving to give oneself a chance to continue and acting. The implementation of such tasks in the case of beings equipped with intrinsic intentionality requires at least a minimum of consciousness. But the high intelligence demonstrated in achieving the goal is not the same as the degree of consciousness. You can be fully conscious of the situation and still perform careless and unintelligent actions. The opposite situation may also occur: problems can be barely noticed, but actions can be very effective. Intelligence is the ability to learn how to perform effective action.

So what is consciousness? Consciousness is a temporary state of mind. We may be conscious of our feelings, perceptions resulting from information provided by the senses, and understanding of concepts that reach the mind or of which we are reminded. Depending on the content of consciousness, we are dealing with the types of consciousness discussed in chapter 1. If we are conscious of qualia, then we are dealing with perceptual consciousness; if these are abstract concepts, then consciousness of space, place, or even wider self-consciousness is manifested. Perceptual consciousness does not require too high intelligence, but self-consciousness or reflective consciousness requires the degree of intelligence necessary to understand the relationship between concepts and the ability to even elementally process information occupying the mind at a given moment. If we are defining consciousness as a state of mind at a given moment, what will happen with our awareness in the next moment? Nothing special. We may lose consciousness and become completely unconscious, but no one will think that we have lost our intelligence. Thinking potential is usually maintained despite unconsciousness. Even if you can't stay awake for a long time, be it in your sleep or as a result of severe injuries, poisoning, drug use, or intoxication, your intellectual abilities, although temporarily weakened, can come back when your consciousness returns. Intellectual potential usually does not disappear in such situations.

There is another factor in the continuity of our consciousness. This factor is memory. Thanks to memory, after we've been awakened, we can restore conscious processes referring to previous experiences. Thanks to memory, we can remember, after waking up, who we are. Our experience is not lost, and knowledge acquired through long-term learning does not disappear. The effort and risk of learning and the struggles to date are not in vain. Our memory models of reality allow us to maintain not only a high degree of consciousness but also our identity.

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

Attention is needed to gain consciousness. Human perception can be subconscious when sensory information is processed quickly and without attention and conscious when attention is paid to what is perceived. Subconscious processing in sensory pathways takes place mainly from the bottom up the neuron hierarchy, where sensory information passes through several levels of the hierarchy, extracting features useful for object recognition. But an object can be consciously recognized only when it focuses on perceiving the object in the image. In experiments in which participants were to remember a sequence of digits in the presence of distracting images of human faces (Fockert 2001), the key role of working memory in controlling visual attention and consciousness was demonstrated.

Many sensory information streams stimulate our perception, but only a small portion of this information is significant and requires attention. Because the working memory capacity is very limited, one should pay attention to what should be represented in such memory at a given moment. The visual cortex alone receives over one million bits of information per second. Attention helps protect memory against information overload. An important aspect of goal-oriented behavior is understanding and modeling the processing of information relevant to its achievement. Sensory information is subconsciously searched for, significant stimuli associated with emotional responses. This inhibits the impact of the thalamus, which, with the participation of the frontal lobe of the brain, leads to the capture of important elements of sensual perception and causes focus on the selected input source. Only the selected sensory stimulus activates consciousness and alerts the cognitive system that leads to attention (Sieb 1990).

Although people have identified neural networks in the brain involved in the attention-focusing process, the dynamics of this process are not well understood. There is a belief that for conscious perception, neural circuits of the prefrontal cortex are needed (Kane 2002). Although this is only one of the critical structures in attention-control circuits, the dorsolateral prefrontal cortex plays a significant role in maintaining conscious attention, providing access to stimuli, motivations, and goals in system memory. The frontal lobes are most developed in humans, where they occupy 30–40 percent of the new cortex. This prompted many researchers to attribute the highest cognitive functions, including conscious sensations, to the frontal cortex lobes. And indeed, damage to this area causes a number of cognitive deficits, such as problems with attention and motor control, short-term memory, temporal and episodic memory, learning through associations, and social consciousness; an inability to switch attention; lower creativity; and limited reasoning ability. In particular, the prefrontal cortex is crucial for the working memory.

Conscious focus is needed to maintain attention. Irrelevant or less useful information can be obtained without proper attention due to the disturbing effects of long-term memory stimuli. This can disrupt the cognitive process that is used to achieve the goals and perform planned activities. When attention moves away from the representation of target information, it can be difficult to recover from long-term memory without the help of attention mechanisms. So actively keeping relevant information in the center of attention and blocking signals that distract this attention are two interdependent features of attention centers. They form the core of cognitive mechanisms that enable logical thinking and problem solving. If we do not pay attention to part of the observed scene, our memory of what we saw quickly disappears and is masked by other visual stimuli. This has been confirmed by many experiments. Our impression that we can see the whole scene very clearly is the result of our memory providing what has just been observed and our consciousness, using frequent mental saccades, renewing the memory of what we have seen.

Cognitive activities require focusing attention on a selected element of the scene. As a result, attention and consciousness are closely related. However, one of the leading researchers, Christof Koch, distinguishes between consciousness and top-down attention, indicating that we can consciously observe objects without paying attention to them, and we can also use objects without their conscious observation. He suggested that top-down attention and consciousness have separate brain mechanisms that regulate their functioning (2007). This should not come as a big surprise, because consciousness is a state of being conscious of something, but it is not necessary to focus attention on it. Dehaene (2014) agrees that conscious impressions, which are a true sign of consciousness and which can be observed in a laboratory environment, require more than just vigilance and attention.

Reaction (response) and attention (concentrated mental effort) can be easily confused; let's discuss why. In the strict sense of the word "attention," we cannot "pay attention" to an object without its conscious perception. However, reaction to an object can be automatic, so it doesn't have to be a conscious act. We do many things automatically without thinking about it (without being conscious of it). In this case, we can respond to the objects we use without paying attention to what we do or to the objects themselves. On the other hand, attention focused on the object triggers a conscious perception of the object that engages our working memory. After the object is spotted, the cognitive system can be conscious of it without paying much attention to it. Attention is the concentrated effort of the cognitive system, and consciousness is the general state of such a system. We may ask someone to pay attention to something, but we do not ask them to be conscious of it. Consciousness appears unconsciously as a result of concentration of attention on something or someone, but it rarely arises spontaneously.

Thinking and planning require concentration of attention on one of the concepts activated in working memory, which are suitable for achieving a specific goal (they are associated in memory with such a goal). This requires the selection of a subset of objects from the contents of working memory for cognitive processing (e.g., the addition of two two-digit numbers). Psychological experiments confirmed that the focus of attention in working memory is used to select a single object for such processing, switching attention from a previously selected object. However, as Oberauer observed, the choice of the object for such processing does not seem to cause an active inhibition of the previously selected object of attention (2003, 2013). The broader attention, which contains several concepts recently activated and stored in short-term memory, serves as a kind of scene from which a single object was selected to focus attention on and to assess its usefulness in a given situation.

Working memory, although it occupies large area of the cortex, is limited by the mechanism of attention. Therefore, it has a small capacity. Psychological studies show that the capacity of working memory varies individually, and one can, through training, increase its capacity. Usually, however, human brains remember only a few elements on an ad hoc basis: six to eight digits in a telephone number or bank account, or several names on an attendance list. Similar potential is associated with allowing a number of objects/processes into our consciousness. Other studies indicate that short-term visual memory capacity is limited to three or four elements (Todd and Marois 2004). This seems to be in contradiction with the need for this memory to process a significant stream of information. It should be noted, however, that the electrical states of synapses of sensory field neurons can simultaneously reflect a huge number of sensory excitation states. The attention mechanism, on the other hand, is a discriminatory filter of "weaker" signals. Thus, working and sensory memory have sufficient bandwidth before the filter and a very small capacity of elements passing through the attention and consciousness filter, remembered first in short-term and then long-term memory. It is assumed that the mind can simultaneously control 5 ± 4 activities. It seems that in fact the mind can focus on one activity at a time, and many activities are performed simultaneously by quickly switching attention.

The Subconscious Mind

When writing about consciousness, it is impossible to ignore the subconscious. The subconscious raises enormous and widespread interest, sometimes more than consciousness. Among amateurs, books describing the importance of the subconscious, its impact on our behavior, and guides on how to shape, improve, and use its power in everyday life enjoy enormous readership. Why is this? Presumably, consciousness is something obvious to most people. It is something we are used to, that we know we have, and that we cannot imagine living without, because loss of consciousness is associated with loss of life. This belief is opposed to the idea of our book, which is to explain consciousness, which seems obvious to most people, and it leaves aside the subconscious that seems so attractive. There is a secret in the subconscious mind. We hear that it is a powerful force that directs our activities but that does not manifest itself and that is hard to master. Let's try to explain what the subconscious is and how it relates to consciousness.

The term "subconscious" was introduced to psychology by the French doctor Pierre Janet, who at the turn of the nineteenth and twentieth centuries studied various psychopathologies and treated hysteria (what we now call neurosis) using hypnosis. He drew attention to the importance of what is unconscious and what remains in the human mind and called it the subconscious. However, this concept was made widely famous only by Freud's work. Freud's main theory, psychoanalysis, assumes that suppressed thoughts and feelings are found in our subconscious mind and manifest in the form of irrational behavior that ranges from seemingly insignificant lapses or forgetfulness and transitions to clear symptoms of mental illness.

According to Freud, unconscious mental processes have a significant, if not major, impact on human behavior. The source of his theses is the belief that mental processes are unconscious, and the most important role in the subconscious mind is played by the primary sex drive (libido). The relationship between consciousness and the subconscious consists in the conflict and censorship of the subconscious through consciousness, and their sources should be sought in the early stages of the development of child sexuality, despite the fact that social customs also affect the development and formation of the individual as a social being. The subconscious is a part of the mind that lies beyond the boundaries of consciousness because its "content" is often too painful to remain in the conscious sphere. The subconscious is a vast deposit of subconscious instincts and memories pushed out of consciousness.

Freud thought that the "content" of the subconscious mind was permanent. According to Freud, the subconscious is stronger than consciousness, and therefore stimuli, desires, and thoughts from the subconscious can cause consciousness to be suppressed by fear. He also believed that there were no psychological coincidences in human behavior. The choice of friends, apartment, or favorite food is associated with unconscious memories and provides tips for rational explanation of our conscious life. Therefore, the task of each person is to delve into the secrets of the unconscious mind and thus themselves. According to Freud, everything we do, think, and feel is determined by what happened to us during adolescence and what influences the course of adult life.

Does contemporary knowledge about the construction of the mind, thinking paths, and decision making confirm Freud's hypotheses? Can we show where the supposed subconscious force comes from? What brain structures and functions that they perform are responsible for the supposed subconsciousness, and does it exist if we don't feel it directly? To answer these questions, we need to consider whether the description of the functions of the mind so far was complete and whether it certainly includes subconscious processes, and if so, which processes run consciously and which create the subconscious.

We wrote about the fact that perception of qualia takes place consciously. But what happens when the act of perception ends? We stop looking, admiring, patting, smelling, and tasting. Then the sensual impression ends, but it does not disappear without a trace. As we described in chapters 1 and 4, the feeling that accompanies perceived objects is remembered. It is often classified and associated with other insights related to the event perceived by the senses in the process of interpretation and memory as a pattern of a specific quale.

The subconscious is also something more. It is a huge memory storage in which we keep traces of everything that has been reaching us since the third month of fetal life. The subconscious contains all the information resulting from our experiences but also the words spoken in our presence and those we said ourselves. Latent content associated with harm suffered is removed from the area of associations reaching consciousness. They can be inhibited by feedback that causes unpleasant feelings whose transmission to consciousness is blocked. In this way, they become forgotten episodes that subconsciously affect our actions. They were recorded with all consciousness; however, the trauma accompanying them could suppress them, and the role of psychologists is to bring them out of hiding.

The information stored creates a variety of connections that can trigger quite different, often unexpected stimuli. The view of an unknown landscape sometimes brings back memories of childhood, such as detecting a strange smell in some situation in which such a smell could not appear, experiencing positive feelings the first time we hear a song, and feeling irritation at someone’s voice although we do not know why. Sometimes activated connections may cause behaviors in us that we do not understand. Someone makes us angry, we are afraid of something, we are drawn to something else. It seems that the string of thoughts sweeping through consciousness is evoked within us in a completely independent way. It is the subconscious mind that dictates the sequence of thoughts.

Of course, sudden events, some categorical command, or a new, important task can switch the mode of thought flow and concentrate thoughts around new matters. But when the engagement imposed from the outside ends, thoughts follow their own path again, and this path is marked out largely by the subconscious. The observed flow of thoughts has its biophysical explanation in previously discussed brain processes. The key aspects are operations of adjusting the stimulation configuration to the patterns. Stored patterns in a decisive way limit what we can finally think of. They take away some freedom of thinking and decision. The impact of the subconscious mind on our choices can be decisive. We described this process in more detail in the section on free will.

However, we can already see that Freud and many of his successors were right. There are many more unconscious thoughts than conscious ones. In fact, perhaps these dominant conscious thoughts occur only one at a time. We wrote about it in describing the functioning of working memory and the sequential mechanism of “attention.” Modern psychologists like to compare consciousness and subconsciousness to an iceberg, only a small part of which protrudes above the water level. This small upper part corresponds to consciousness, whereas the huge number of associations that can become the basis of thought processes are hidden underwater. Some of these associations may never be realized as a separate conscious memory or thinking pattern. However, this does not mean that they do not affect the path of our thinking. They create a ground that participates in the formation of thoughts, plans, and informed decisions. Mental events, which take place in the subconscious, create premises for conscious mental processes, such as dislike of or sympathy for newly met people. The human psyche tends to push unpleasant sensations or preferences condemned by society to the subconscious. Injuries, passions, and aspirations are directed to the subconscious, where they develop, sometimes creating complex mental structures. Although the mainstream of the mental life of an individual, his feelings, ambitions, complexes, and aspirations, takes place in the subconscious, it plays an important role in human life, because it determines his views, motivations, and attitude toward other people.

A number of patterns or paradigms anchored in the subconscious define what we believe, what we expect, and how we think we should act. This is the result of learning and upbringing. Parents, educators, guardians, and teachers have embedded this into our subconscious. We treat them as our own, and we are convinced that they describe objective reality. We recognize them consciously when we notice that the results of our thoughts and our decisions are always similar and never violate the instilled worldview, instilled principles, and adopted paradigm of social behavior.

But was Freud right when he thought that everything we do, think, and feel is determined by what happened to us in adolescence, that our choice of friends, home, or favorite food is associated with unconscious memories? Such theses reject the possibility of the influence of genetically determined character traits. If that were the case, it would be difficult to explain the cases, known in the literature, when identical twins separated after birth and raised by different families, in different economic and social conditions, in different cultures and different countries, show an amazing similarity in lifestyle preferences, type of friends, or choice of place of residence. This indicates that the features of the mind are influenced by both our genetic makeup and behavior patterns collected in our subconscious mind throughout our lives.

If we can reach the subconscious mind, allowing thoughts that it suggests to us, can we use this huge treasury of knowledge to make meaningful decisions? Of course we can reach into memory, trying to remember what we currently need. However, sometimes there are situations in which all the knowledge we accumulate is insufficient to make an informed decision.

If primitive man traversed forests and valleys, looking for the best place to build a house, then he would have to get a huge amount of information to do it right. He did not have access to most of this information. He could not study the history of floods and landslides, could not assess volcanic activity, soil composition and acidity, the degree of aggression of neighboring tribes, the presence of dangerous animals, and the risk of disease. However, having personal experience, he could notice the color of the leaves, marks on the grass, noises, and common animals and insects. Based on these faint premises, he could usually make the right decision. He was helped by an unusual form of knowledge: intuition. Intuition refers to the subconscious. Our wanderer would probably not be able to justify his choice rationally. He would probably claim that he feels that this is a good place or that good deities tell him so. The subconscious has allowed him to associate a large amount of data and poorly defined parameters that consciousness would not be able to process.

Conscious, logical thinking is helpless if the problem is poorly formulated and there is no theory combining numerous parameters. The subconscious, however, works differently. The pattern similarities sought are set much less stringently. Instead, they can cover more data. The conclusion is less certain, but it is possible to draw it. This is the great power of the subconscious mind and intuitive cognition.

Intuitive thinking has one more advantage over analytical thinking which based on propositional attitude to the problem under consideration. It is much faster, which allows one to make quick decisions in real time with much less effort. Kahneman Daniel describes this advantage in detail in his book "Thinking, Fast and Slow"(1994).

We can reach our subconscious by yet another way: by interpreting our behavior, especially dreams. This is in line with the thesis of Freud, who treated dreams as a source of direct insight into our unconscious mind. In his opinion, the driving force behind dreams is the drives and desires we had in our childhood, which were suppressed by the regime of home, school, and social education. What the dream tells us is called the explicit content of the dream, and the latent thoughts we reach through the chain of associations are the hidden thoughts of the dream.

Symbolism of speech containing permanent and characteristic elements of the mentality of a given society penetrates dreams. The ability to symbolize is therefore a species trait that is not acquired; it is part of the inherited culture of human nature. However, keep in mind that new symbols are constantly being created that attach to the old ones. Since desires are suppressed during wakefulness, they are revealed only in conditions of reduced censorship during sleep. The essence of dreams is to be hallucinatory fulfillment of irrational drives and desires. Dreams can also reveal the independence of simultaneous associations generated by the subconscious.

One of us, Wiesław, received illumination due to an unusual dream he had, which confirms this peculiar character of the subconscious. The dream was expressive and realistic.

In this dream, a mother reproved her daughter that she was learning badly and apparently was insufficiently clever or distracted because she was busy with other matters. To confirm her position, the mother gave her daughter a difficult puzzle. The daughter obviously had no idea what the correct answer was, and sleeping and dreaming Wiesław did not know the answer either. The mother's aggression increased, and the conversation was like tyrannizing an incapable child. Wiesław decided to defend the daughter, arguing that the puzzle was too difficult for a small child and suggesting that the mother certainly did not know the answer either. However, the mother gave the correct answer. Wiesław was surprised, because the answer seemed obvious. The consternation caused him to wake up…and then he began to wonder about the meaning of the dream. How was it possible that he did not know the correct answer to the puzzle? It was presented by a "mother" whom he dreamed of, a mother created by his subconscious. The puzzle was formulated in such a way that the person who asked it had to know the correct answer in advance. So the "mother" in his subconscious knew the answer, but he didn't know it himself? And yet his thoughts, while dreaming, were also generated by the same consciousness. This means that different associations are swirling in our mind quite separately.

In this strange dream, the collision of two threads was amazing. They were not independent but interacted with each other. Perhaps such a phenomenon is only possible in a dream, because a waking conflict of thoughts would be censored by the logic of consciousness. The fact that consciousness comes to them separately (and sequentially) does not mean that we have two personalities, although psychologists also take this into account. Perhaps there are people with expanded consciousness who can simultaneously draw knowledge from different areas of the mind. Perhaps one day artificial intelligence will be encouraged to engage in such a dialogue on the border between consciousness and subconsciousness. These are fascinating questions facing cognitive science and robotics.

The question arises whether we can influence the subconscious mind or break free from established paradigms and persistent preferences. The father of American psychology, William James have believed that man can consciously control the subconscious (1890). James's statement was not entirely confirmed. The experiences of the twentieth-century psychology have shown that the impact on the subconscious mind is clearly limited. The subconscious is created for too long and too arduously to be easily modified with a simple trick afterward.

Pharmacological agents used for this purpose have very limited effects both in strength and period of effectiveness. The psychological therapies used gave psychologists more satisfaction (mainly in the form of fortunes obtained thanks to them) than patients. The most effective therapies are your own efforts to rebuild your psyche. It is often the case that people in the immediate environment say a lot of bad words to a child, give them bad grades, and provoke all their childhood behaviors. All this can cause the child to have lower confidence, a lack of self-esteem, pathological reactions, psychoses, and complicated mental deficits. This negative information will always remain in the subconscious mind, permanently burdening the child's psyche.

However, an adult already can consciously enrich the subconscious mind with such concepts, ideas, and words that build it and serve it. He can get new experiences. Change the environment and change the way of life. This will gradually build a new subconscious. This is worth consciously dealing with. A friend, wife, teacher, or psychotherapist can help. In practice, this is not easy. It requires freeing oneself from inhibitions resulting from traumatic experiences of the past. This is known to those who survived the phenomenon of paradigm shift and who managed to see reality in a different light.

Modern popular psychology calls this phenomenon an "aha moment." The "aha" provides a new insight into the situation or a new way to solve the problem, or makes one conscious of the chance to change the way one acts. Under this trendy name, there is sometimes a deep remodeling of the subconscious. It contributes to the new human personality.

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

The Circle of Emotions

Consciousness can be found in the minds of intelligent beings whose purpose is to remain in attainable well-being. As we already know, intelligence enables one to learn how to survive in a hostile environment. It also helps one to avoid adverse factors and find any beneficial effects on the physical and mental state. The struggle for beneficial factors and avoidance of difficulties raises a variety of needs that must be met in order for an intelligent animal or human to succeed in the struggle for survival. Failure to meet these needs raises stress, which we can treat as a punishment that fate or its inefficient action gave to a living being. Even partial success in meeting needs can be considered a well-deserved reward of fate for effective, wise, or merely clever actions.

Needs that exert pressure on the functioning of people and highly intelligent animals can have a diverse nature. They may result from aspirations:

1. To meet the need to acquire information and knowledge about the environment.
2. To maintain the homeostasis of the body.
3. To satisfy drives.
4. To satisfy instincts.
5. To meet psychological needs.

Stress resulting from unsatisfied needs evokes emotions. Generally speaking, these are emotions of willingness or unwillingness to do something to meet the needs. They can be explicit (conscious) or hidden (unconscious). They enable people to understand and be made conscious of their situation, because they give value to their experience, transforming it into feelings. To a large extent, they decide the efficiency, energy, and pace of actions taken. We can experience many of these emotions as a general emotional state and as an attitude toward the surrounding reality. Here, we qualify feelings of anticipation, longing, interest, and apathy. Usually they are unconscious and difficult to define.

Emotions are accompanied by feelings conducive to meeting the needs. Generalizing the list of sources of emotions, let’s enumerate some of them, assigning them to the specific needs indicated above in points 1 to 5.

1. Curiosity and willingness to understand
2. Hunger, thirst, various types of pain, and desire to escape
3. Sexual arousal, disgust, disgust, fear, appetite, and satisfaction
4. Discomfort of touching, disgust, desire to blink, and itching
5. Love, trust, excitement, and apathy

The line between sensations and feelings is blurred. We can have a lasting sense of overwhelming love for the whole world, or a clear and intense feeling of passionate love for a specific person or object at a given moment. We have already discussed the feeling of pressing concerns and persistent fear of often undefined threats and sudden fear causing a panic reaction in a dangerous situation.

Curiosity can be a general state of interest in the surrounding world, or it can be an urgent need to explain something that just interested us. Curiosity is an everlasting flame that burns in everyone’s mind. The same applies to apathy, a lack of interest as a general mental state or a sense of indifference, of detachment from a given attitude of “what do I care,” often supported by a shrug. The criterion for distinguishing these states is the degree of consciousness of the state in which we find ourselves.

Affections are emotional states that we often don't realize. They are often pushed deep into the subconscious. We do not realize that we are depressed or overly optimistic. These states usually last longer, and we recognize them after changing our behaviors and responses to events that happen to us. However, we are able to react immediately to feelings of pain, fear, hunger, itching, or passion. These reactions are fully conscious, and without consciousness, we are not able to feel them.

Feeling drives sensitivity to sensory stimuli and prompts us to act, search, acquire, and perform all other activities triggered by these emotions. As a result of these activities, the needs are satisfied or not met, the goal of action is achieved or not, and the signal associated with this reward or penalty is obtained. The signal of reward and punishment is used to learn which actions appear to be the most effective and how to behave in the next similar circumstances. By associating the effects with emotional states, a person or animal learns to recognize the importance of individual affections and feelings and thus develops his qualia and enriches his psyche. The result of the actions taken is a change in the state of needs. Here the circle of emotions closes, because the body is ready for the next cycle of adjusting the mental state necessary for planning and making decisions about the activity most favorable from its point of view, taking into account its current needs.

The matter gets complicated because another sequence of psychological reactions may be operating in parallel:

Perception of new stimuli that cause pleasant or bad impressions. A stream of information from the senses flows in, constantly affecting our psyche.

Arousing emotions associated with these stimuli, including hidden ones. Sensory stimuli evoke new emotions, mainly impressions and sensations at all levels of consciousness.

Intensification of emotional needs. If we have a need for company, then a scene in a movie can strengthen this feeling so strongly that we realize our loneliness is unbearable.

Taking actions to meet these needs. Realized needs and emotions connected with them constitute motivations to remove discomfort resulting from their dissatisfaction.

Both of the abovementioned paths of throbbing feelings, making decisions, actions, and changes in well-being can function in parallel and interact with each other. This creates an endless wealth of emotions and a variety of behaviors.

Another complication is that the feeling itself or the action taken as a result of the feeling can spontaneously become so pleasant or terrible that it determines further actions regardless of existing primary needs. Then this feeling or action becomes a self-driving need. This happens when, for example, the feeling of anger causes someone to exact corporal punishment on subjects, and after prolonged practice, beating and the accompanying emotions become pleasure, turning the severe punisher into a sadist. Generally, repetition of activities involving emotions can turn into a new need. In particular, if the sensations and the subsequent actions start to give us pleasure, they may be accompanied by the release of endogenous opiates, which will have the same effect on these activities as addiction. What's worse, there are also mental associations of these emotional and functional sequences that strongly change the personality of a person.

As we can expect, the strongest and most widespread stimulation will become the object of consciousness. It doesn't have to be a uniform feeling. It can be a complex of feelings and emotions that creates a complex mental state. We can imagine that a wounded soldier, though feeling severe pain, will also feel the desire to survive, which will determine his actions in order to hide, escape, and so on. The lover can feel hot passion for the beloved while at the same time delighting in the scenery accompanying the date.

Other feelings and image schemas of emotional patterns will remain dormant in the subconscious. Stimulation of the object's pattern is not a feeling in itself, but its reminder may arouse the same emotional responses as when we first encountered this object. The decisive role in evoking feelings is played by associations that were established during previous experiences. In a similar way, through co-occurrence, associations of emotional states with effects of actions and achieving satisfaction or disappointment while satisfying needs are established. This description makes us aware that it is emotions that determine the behavior of living things to a large extent. It confirms that humans and animals are what they are because they are emotional beings and because they have motivated emotional mind (MEM).

Probably not all manifestations of the human and animal psyche, not all phenomena and problems, have been solved or even discussed here. Mind, perception of the world, and therefore the level of consciousness changes from generation to generation. We understand our psyche better and better.

People who are mentally ill or only slightly disturbed are not discriminated against and removed from social life. Contemporary language, liberal culture, and scientific knowledge limit the manifestations of mysticism, or at least its demonstration in interpersonal relations. Semantics and image schemas found in social and media communication adapt to the expanding scientific worldview. Perhaps this is reflected in the phylogenetic development of humans fixed in the genetic code. However, the truth is that there will be no time for further natural evolution of the brain! If we want to keep up with the development of technology and scientific knowledge, it is not enough to raise the level of education.

Already today, only a small percentage of people understand the physical, astronomical, economic, and biological theories that modern science presents. This also applies to the theory of mind, which was the impetus to write this book. Soon we will begin to repair, improve, and supplement our brains with artificial elements. This may be an effective way to continue cultural progress. Another way is to interact with self-conscious artificial intelligence, which will be the subject of the next part.

Particular caution is needed here, because some philosophers believe that epiphenomenalism is a form of dualism, where brain matter and physical processes are accompanied by mysterious intangible epiphenomena. Does this not mean that we are abandoning the reductive model? Fortunately, Jaegwon Kim (1998) proves that such ontological epiphenomenalism is not contrary to reductionism. Thus the presented model can be safely called the reductive model. And on the other hand, the epiphenomenality or causal inertness of mental properties seems equally mysterious. How and why would something with causal powers (i.e., the relation of neural representations) ever produce something *without* any causal powers?

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

Part II Machine consciousness

The Embodied Mind

The notion of embodied mind was introduced in part I, proving that the embodiment of the mind is essential for the intelligent system to interpret simple sensory impressions and turn them into qualia. Discussing the natural mind, we drew attention to the importance of the notion of an embodied mind in the shaping of speech, and we even quoted the concept of embodied mathematics to explain how mathematical abilities have evolved in human minds. Now we will examine this concept as fundamental to the design of artificial minds and to understanding what qualities can be achieved by an autonomous agent—a robot operating in a natural environment. Moreover, we claim that no intelligence can arise without this embodiment.

The embodied mind is the mind of an autonomous agent who possesses the body. The principles of designing robots using the idea of embodied intelligence were first described by Brooks (1991) and were characterized by several postulates for the development of embodied agents. The first assumption is that agents develop by acquiring experiences in a changing environment in which they act and observe the effects of their actions through their senses. Another important assumption adopted by Brooks is the postulation that there is no need to build an environmental model. Instead, the robot simply can observe the environment in which it works and learns. This approach has revolutionized robotics and has changed the research direction of cognitive systems.

Animals, even those with very simple nervous systems, learn. Eric Kandel of Columbia University studied the learning process in the example of marine gastropods (2007). The sea snail only has about twenty thousand neurons, so it is relatively easy to study its learning process. It was observed that when the snail was gently touched, it withdrew—sensory neurons directly stimulated motor neurons for immediate withdrawal. However, repeated gentle touches caused a gradual habituation to touching, and the withdrawal action was weakened. As a result of learning, the strength of connections between sensory and motor neurons weakened. This is already known to us as the effect of getting used to a stimulus—habituation. However, this was not the case with painful touches. In this case, the sensitivity and strength of the response grew. The snails learned a more radical reaction to an unpleasant stimulus.

What really is the embodiment of the mind? How can you understand it in the most general sense? The simplest is to understand it as an agent's body. If the robot is to function autonomously in the environment, its body must contain sensors and actuators that activate motor functions. The only forms of true incurred intelligence we know are in humans and animals. So maybe the embodiment should be defined as the body of the intelligent system. However, such a definition would be erroneous. Goldman and de Vignemont (2009) explain how we should understand our embodiment. Although they focus on social cognition, their observations can be applied to the embodiment of intelligence. They extract the brain from body and the body from the environment. Their key thesis is the claim that the body cannot be excluded from the environment. Also, let's add that the brain cannot be detached from the body. Scientists explain that anatomy and body movements are important for determining perceived sensations and processing sensory signals in the brain. Gallagher claims that our sensory experience depends on how the head and body move, and also on where the body has different senses (such as eyes and ears). In addition, our motor functions are conditioned by the construction of muscles and tendons, their elasticity and interdependence on other muscles, and the history of their use (Gallagher 2011). He indicates that movements are not always fully planned by the brain, and some of the necessary signal processing takes place in the environment. The nervous system learns "interactions with internal body and environmental restrictions to obtain desired results" (Clark 2002). Behavior of the embodied system is not merely the result of an internal control structure (such as the central nervous system). Its morphology (body and limb shape, and type and location of sensors and effectors), as well as the material properties of the morphological elements (Pfeifer 2007), are also influenced by the behavior of the system. The body determines how we perceive and experience various phenomena and objects in the environment and ultimately how we understand the world. Our feelings, understanding of the observed environment, and motor skills are the result of bodily sensations (Dehaene 2003). From such impressions arise the patterns of orientation in action (like the front, back, side, and top), which are analogous to the imaging schemes representing these activities in the mind. The body plays a special role in sensory consciousness, which determines the impressions or the qualia.

Liao conducted experiments with trout in water with turbulence and without (2003). These experiments show that fish use the physical properties of their environment through body morphology to reduce energy consumption while swimming. The trout flows between experimentally generated vortexes, using the vortexes to reduce the effort of their muscles compared with the effort of fish floating in the water without turbulence. The essence of embodied cognition is that cognition is not only done in the brain but is dispersed in the brain, body, and environment so that corporal processes are shaped by and directly contribute to the creation of consciousness and cognition. Noë presented a detailed picture of conscious perception in which the sensory-motor relationships and properties of the environment are responsible for the construction of the brain-recognized representations (2004). Thompson and Varela (2001) extend this to human interactions that involve facial expression and create gestures, movements, and sensory-motor associations.

If the intelligent system is conscious of its own body, it must have a way to determine what the body is. Many psychological experiments have shown that it is not as easy a task as you might think. The illusion of a rubber hand is a good example of such difficulties. In this experiment, the experimenter touches the hand of the subject. The subject's hand was hidden from her gaze and resting beside the rubber hand, which was clearly visible to her. The subject simultaneously touched the rubber hand with her own hand in exactly the same way (by scratching, stroking, squeezing, tickling, etc.). After some time the subject had a deluded impression of touching felt in the rubber hand. She identified and felt a tactile impression in a place that she could see on the rubber hand. She felt as if she had been touched at exactly the same place where the experimenter touched her during the initial phase of the experiment, when her hand and the rubber dummy were irritated at the same time. This can be explained by the degree of functional synaptic coupling between the synapses responsible for the visual perception of the touching activity and the neurons in the sensory system. As we recall, when such a coupling arises, the mere irritability of the visual neurons generates an illusion of receptor irritations in the learned and previously sensitized site. This phenomenon is known in psychology as a visual capture, characterized by the domination of vision over other senses. This phenomenon is used by ventriloquists in such a way that the viewer has the impression that the voice he hears comes from a dummy whose lips move in sync with the ventriloquist's voice. Another example shows that understanding speech may depend on visual observation. For example, if you hear the recording "ba" while observing a video in which a person says "ga," you will hear "da" as a result of visual impact on the hearing.

Other unusual phenomena are phantom arms or legs. A person whose hand has been amputated may feel a hand even when he does not have it anymore. If he feels itching in such a phantom arm, he cannot scratch it, because the arm is gone. However, a psychologist can deceive him by using a system of mirrors and showing him the image of another hand in such a way that he perceives it as his missing hand. Because this hand visible in the mirror can be scratched, the discomfort of the itching sensation can be removed. V. Ramachandran, who worked with people suffering from various neurological disorders, described many cases of false and strange sensations, including phantom limbs, presence outside one's body, visions of a personal God, and even the denial of his own existence (2011).

The impression of presence outside of one's body is when a person seems detached from his body. These include the sensations of swimming, flying, levitation and rotation, or the experience of seeing only part of one's own body. These experiences relate to the pathological feelings of position, movement, and perception of the completeness of one's body and are the result of the fact that the brain is unable to combine sensory impressions into coherent information about its own position (Blanke 2002). According to Dehaene, the impression of being outside of one's body is a stronger form of dizziness that we can all have when our perception is not compatible with other senses (2014). By studying and understanding these strange cases, he could better explain human behavior and help some patients overcome their fears.

%%

Perceiving and identifying with your body (the belief that "it is me") depends largely on observing the body in the environment. It is also a mental image of yourself. What qualities are most often associated with self-understanding? According to Gallagher, they usually contain basic biological and ecological aspects that allow the system to distinguish between what is itself and what is not. Other embodied aspects of the "I" are those that define an egocentric reference point that reflects the perspective of the first person and allows action in the personal space (Gallagher 2013).

Self-consciousness requires distinguishing yourself from the rest of the environment. Conscious states are caused by externally supplied signals from the environment but also by internally generated mental states or signals coming from inside the body. Self-consciousness requires the ability to distinguish between externally generated stimuli of conscious experiences from those internally generated. It requires that we can distinguish our own thoughts from external images and can perceive the needs of our own body. This task is facilitated by the sensors of touch and temperature that are placed in many parts of our body. When we touch different parts of our body, we receive signals from both the hand we use to touch and the place it touches. When we touch an object other than our body, the impression of touch comes only from our hand. In addition, the feedback derived from the observation of the touched parts allows the association of touch signals with body parts. This self-identification method is not flawless, though, as evidenced by experiments with phantom limbs, which we have already written about.

Distinguishing oneself from others can be confirmed using different sensory and motor modalities (e.g., sense of balance, touch, and vision). They confirm the embodied agent's effective exercise of body control, movement, and readiness to act. Other aspects are related to social existence, which manifests itself when one realizes that another person is looking at him or has a sense of recognizing himself as being separate from others (which is associated with the ability to recognize himself in the mirror). Important cognitive aspects of self include continuity in time, memory about existence, and recognition of one's personality traits. Finally, the concept of self can be extended to the family, the environment in which it grows, the culture, and the things it has. We distinguish things like our family, our community, and our country. The concept of self is neither perfect nor permanent. You may not be able to recall your own life in the past, as in the case of amnesia or Alzheimer's, and your character and personality may not change. But even in such extreme cases, there may be a limited form of self-consciousness, such as identification with your own body and relations with the outside world.

Self-consciousness helps in observing your own body and perceiving your own needs, thoughts, emotional states, character traits, skills, and knowledge. This consciousness affects our self-confidence, relationships with other people, life plans, ambitions, and goals that we set for ourselves. At the same time, we realize that others are also conscious of their own selves, that they act in accordance with this consciousness, and that they may take into account our reactions to their behavior and understanding of their mental states. Understanding other people allows us to explain their behavior, needs, and motivations and allows us to take their needs into account in our activities and to influence their behavior, preferences, and views. What's more, self-consciousness prompts you to form a belief that other people can be conscious and have a conscious mind. It makes it possible to create the theory of not only of one's own mind but also of foreign minds, which is the basis of reflective consciousness. One of the functions of mirror neurons is imitation of the behavior of others. Young children cry when they hear the cries of other children, although they may not know what the reason is. We can feel pain or sympathy when seeing the suffering of others.

In psychology and neuroscience, the body schema is known as the ability to organize sensory information that provides effective control over the body's position in the environment and control of its activities. Control over the arms and legs is unconscious, but it can work precisely when reaching for objects or performing activities. In this way, the body is integrated into the mind, and control over its movement becomes automatic. We do not need to be conscious of having hands to perform actions using them, but as experiments with phantom limbs show, the body schema may be disturbed.

We can also change the body schema using tools. When a blind person uses a cane, it helps him perceive the environment, and the cane becomes part of his sensory system. In a sense, the cane becomes part of the body that the mind directs to view the immediate surroundings. Another example of body extension using tools is the use of eyeglasses to improve vision. Glasses become a part of the embodiment that the mind uses to perceive objects in the field of view. Tools do not have to be attached to the body to create extension as long as they can be manipulated (perhaps even remotely) by the body (Noë 2009). A surgeon using remote manipulators can stretch the schema of his body at a distance of up to several thousand kilometers. Another example of using tools to extend the body schema is to use cell phones to communicate. Such communication is an extension of the body schema, providing a virtual presence and facilitating contacts between people. The mobile phone becomes part of the remote sensory and motor system controlled by the body. Many car drivers or aircraft pilots experience the unification of their mind with the vehicle they're driving. The driver of the sports car senses every vibration of the machine, hears the sound of the engine's intense work. The suspension of the car must be sufficiently hard to allow the driver to feel the smallest object under the tires. The pilot senses overload when rising, falling, or rotating. Driving a luxurious limousine with a soft suspension and muted interior does not provide such impressions. Often the tired driver observes the road as if behind the glass, like on a movie or TV screen. This reduces the driver's ability to react effectively and poses the danger of a delayed reaction or even falling asleep behind the wheel. Therefore, the automatic support with which modern vehicles are equipped is desirable. Steering in these conditions resembles attempting to control a vehicle in a poorly designed computer game. However, in good games with sophisticated graphics, the player can feel united with a game character (especially in the first-person mode) and have a feeling of full participation in the action. The body of the character becomes the body of the player, the senses of the character his senses, and the movement his own feet and hands. Only the mind remains singular—the one in the player's skull.

Just as tools can extend the body schema, so can language expand the thinking schema. In this case, we share our thoughts and plans so that we can collaborate more easily and achieve better results in our conquest of the environment than without such communication. Language tools, concepts, and imagery patterns resulting from the use of the language provide an extension of our minds and help us to think, solve problems, and exchange information. Good examples of such language-based mental tools are numbers and the numeric system. Great advances in mathematics, engineering, and science became possible as soon as people adopted the Arabic numeral system with positional notation and zero. However, mental tools such as language are based on the perception and understanding of the world around us. Therefore, language (including thoughts) is not limited to our skulls but exists within complex interactions between humans and their bodies, minds, and sensory-motor skills.

A blind person does not consider his guide dog as a tool. He recognizes it as part of his environment, which he can trust, so the dog is integrated into his conscious perception of the world. They largely live in a symbiotic relationship of mutual trust and dependence, and to a considerable extent, mutual trust develops gradually based on the daily routine and friendship between the blind person and the dog. They are present in mutual consciousness, as the mother is present in a child's understanding of his world. A guide dog or rescue dog cannot be considered a tool, because the cooperating person does not exercise full control over it. The dog has its own preferences and must have its own willingness to perform the task, else it cannot be relied on. It is therefore a partner, not a dumb tool.

The embodiment of the mind is defined as the part of the environment that is under mind control and includes sensors and actuators through which the mind communicates with the environment. Sensors send sensory data from the environment to the mind, including information that comes from the body. This way the brain can feel the pain signals generated inside the body as well as those coming from outside the body. Actuators use brain-generated signals to control movements to perform operations. Brain-controlled activities can be performed in the external environment as well as inside the body (such as secreting hormones).

The embodiment of the mind does not have to be predetermined, unchangeable, or even have a physical body form if the brain can only use it to perceive and affect its environment. The boundaries of the embodiment of the mind can change, forcing it to relearn the inner model of its own body and change the understanding of itself. Tools can be thought of as an extension of the body if the brain exercises control over them, but also vice versa—lack of control over the former parts of the body excludes them from the embodiment of the mind. Embodiment need not include the brain of intelligence that controls it but must communicate with it through channels that ensure the perception of the environment. It must also contain actuators and their communication channels, through which they receive motor signals from the brain. The precondition for remote communication is the assurance that each sensor stimulates a clearly defined region of the brain and that each actuator receives signals from its assigned memory regions, which control the body's movements.

%%%

The idea of the agent's body and its relationship with the environment is increasingly popular in philosophy, psychology, neuroscience, robotics, education, cognitive anthropology, and linguistics, as well as in systems of dynamic approach to behavior and thoughts. The role of the body is explored and appreciated in many applications, from robotics to perception and action control. The use of body dynamics and environmental properties facilitates control, planning, and decision making. Animals have learned to use the body's interactions with the environment to move, maneuver, and exploit a specific location in the environment. Fish use naturally occurring currents to increase speed, and robots can exercise physical principles to maintain balance or walk efficiently. Animals use easy-to-detect environmental hints and active vision (supported by activities such as head and eye movements) to improve perception and repeat observations to obtain more detailed models of observed objects. Vision is explained as an active cognitive process that includes the study of the perceived object based not only on rich, detailed internal models used for object recognition.

Conditions that exist in the real world of things create a framework for proper action and reflect the compatibility of agent's competencies with environmental properties. Used correctly, they facilitate the effective implementation of the actions of the embodied system. For example, diving birds use the relative speed of the optical expansion of the observed image to accurately predict the time of collision with the water's surface so that the bird can collapse its wings at the right moment. Another example is a player tracking a baseball and changing his position so that the observed ball approaches him in a straight line in his field of view. Thanks to this, instead of observing the curve of the ball movement and predicting where the ball might fall, which would require a very complex calculation process, he can be sure that he is in the right place to catch the ball. In this way, using body movements, a complex problem can be replaced with a simpler one. Such sensory-motor coordination replaces internal representations and comprehensive planning with less costly strategies. Its task is not to first model the world and then plan and act but to maintain the adaptive balance between the agent and its environment. This leads to potentially different strategies to represent the world in memory, effectively integrating observation, operation, and the environment. This model based on higher-order invariants explains some of the adaptive reactions in rhythmic motion, the ability to capture and use objects and visually directed activities. This type of action, based on the observation of changes in the environment, has led to many useful adaptive mechanisms, such as the evolving mechanism of a sunflower tracking the movement of the sun in the sky. The sunflower does it without understanding, consciousness, or intention. But does it feel the sunlight?

In a conscious embodied mind, representations are created and used for actions with objects regardless of whether they are present, and reasoning can be treated as simulated perception and action. In many cases memory represents sequences of sensory stimuli, feelings, and low-level motor control. Consciousness at a higher level can be treated as an abstract (i.e., detached from real observation and action) version of the coordination of perception and action, though it is less consciously felt than lower type qualia (based on real experiences and sensory feelings).

Needs and Motivations

Natural brains have a complex structure in which individual parts play specialized roles. We revealed and presented this in a shortened description of important brain functions. Neurobiologists and psychologists study the functional features of different areas of the brain, trying to understand their structure and the connection paths between them and to find out how they interact with each other and learn about the organism's environment. Although natural brains exhibit intelligent behavior and consciousness, we are not able to construct such complex artificial brains. The question is whether we can propose artificial brain structures in machines that will be able to create the basis of intelligence and consciousness, even if this intelligence and consciousness are different from those of people. The approach we propose is a constructive one. Wanting to build an artificial brain, we propose what properties it should have, how it should be organized from the functional side, and how it could perform these functions, using known technologies, algorithms, network structures, and computing methods.

The first step is to define the embodied intelligence so that it not only can be scaled from simple to more complex but can also clearly distinguish intelligent entities from nonintelligent ones. In nature, we have many specialized organisms that perform complex operations in order to survive and pass on their genes. The definition of embodied intelligence that we have adopted is an arbitrary division between intelligent and nonintelligent entities, aimed at building an intelligent and conscious machine. We have recognized the ability to learn as the most important feature of intelligence, which is why we consider beings that do not learn anything as not intelligent. We can distinguish different forms of intelligence; among them are individual intelligence, group intelligence, or intelligent processes resulting from natural selection. These three types of intelligence are reflected in research on artificial intelligence, leading to such sciences as computational intelligence, swarm intelligence, or genetic algorithms.

We limit ourselves to understanding individual intelligence as that of an embodied agent who learns how to survive in an unfavorable environment. We also ignore the discussion of whether intelligence can develop without embodiment, although we share the widespread belief that it is impossible. We wanted to avoid discussing the usefulness of a potentially intelligent device that decided not to do anything. That is why our embodied agent's mission is to survive in an unfavorable environment. What do we mean by that? We know what survival is for people: living and working in an environment where one needs to eat, seek a partner, or avoid cold. When people are in an unfavorable environment, their intelligence and knowledge about the environment ensure their survival.

Our agent will not have the same needs as people but must have needs whose fulfillment is a measure of success (survival). Meeting these needs will require physical and mental effort, and the development of useful skills will be associated with the development of intelligence. The agent's environment must therefore be unfavorable. We must provide conditions in which individuals with intelligence will not only be able to survive but will also be subjected to pressure from an environment in which better solutions, greater skills, and broader knowledge count. Such an environment stimulates the agent to develop.

The opposite can be argued: that the environment that provides the agent with everything it needs is not conducive to the development of intelligence. In a mythical paradise, man would have everything he needs—that is, the environment would not be conducive to the development of intelligence. But deteriorating conditions in the environment can stimulate people to make a greater effort to survive. Many people claim that the fastest development of science and technology takes place in periods of wars, when the lives of many individuals and the existence of nations are threatened, and mobilization for defense requires extraordinary mental and physical effort and cooperation.

An artificial brain directing a virtual or real agent acts to ensure that its needs are met. An agent may have such needs as preservation of a certain level of energy, protection of its body against destruction (which may include avoiding extreme temperatures, strong magnetic fields, excessive humidity, or collisions with objects), or reaching the goals set by the designer. These needs that are predetermined by the system designer are called "primary needs," in contrast to the higher-order needs that the agent will develop in the process of learning and implementing the tasks set before it.

The agent treats unmet needs as a signal to act. The strength of this signal depends on the degree of unmet needs so that the agent can differentiate between them. This facilitates an automatic needs-management process and indirectly motivates the agent to take specific actions. Using the analogy of pain forcing an immediate human reaction, let us call these signals representing unmet needs the pain signals. The sizes of these signals can be measured and compared with each other. Various pain signals not only provide motivation for action but also control the learning process. If the agent took some action, and as a result of this action, its pain decreased, it means that the right response to a given need was found, and it is a signal to learn. If the effect was the opposite, and the pain increased, it means that such activities should be avoided and different action tried out.

The decreasing pain signal can be perceived as the reward signal (although it is not the same as reward), and the conscious agent can feel the resulting enjoyment from it. Based on the reward mechanism, a machine-learning system known as "reinforcement learning" was developed. Reinforcement learning is considered the most effective method of machine learning. This method allows the development and control of very complex autonomous systems that achieve the set goals in the natural environment, where the environment model is difficult or impossible to develop, and that require changes in the way the agent works.

But should we restrict the machine learning to meet the needs given to the agent by the designer (like homeostasis or simple drives)? Can an embodied agent be truly intelligent if it only pursues the set goals? Can it achieve consciousness and self-consciousness without possessing curiosity about the world, without seeking new solutions and setting new goals for itself—goals that its designer would not know would be useful in the environment in which the agent acts? Intuition tells us otherwise: that an intelligent agent cannot be a slave that blindly obeys its master. And what if the agent, conscious of its actions, finds out that it violates social norms? Can it then (and should it) refuse to carry out the orders?

Proponents of reinforcement learning say that a system of goals and rewards specified by the designer is enough for the agent to achieve any degree of mental complexity. In particular, they argue that the agent can realize sub-goals on the way to achieve the main goal for which it receives a prize. But is it enough?

In the 1990s, researchers had already found that machine learning can be enriched by adding (in addition to the external reward used in reinforcement learning) an intrinsic reward that the agent receives when it satisfies its curiosity. While the reward for fulfilling the designer's goal comes from outside and requires an external evaluation, the internal reward is an internal to the agent's mind, helping it to learn new things and expanding the agent's knowledge—hence the postulated drive of curiosity discussed in our work.

Let us consider for a moment whether there would be useful way to extend the internal reward, which the agent receives after the discovery of novelty, as a reward given for its own deliberate actions. The internal reward has the advantage of not requiring external interference. In addition, the fact that the agent received an internal reward does not have to be disclosed to an outside observer. Because the internal reward is a signal to learn, the agent changes the state of its knowledge without the knowledge of the observer. If the reward is also connected with achieving the goal that the agent set for itself, then the fact of achieving this goal is also unknown to the outside observer. The agent becomes unobservable, and consequently also unpredictable, if its internal states are unknown. This is similar to what we know about our own intelligence, that no one other than ourselves knows our thoughts or goals. Nobody knows how our motivations change. But these are only the alleged qualities of our intelligence. Are they useful for something other than our private world, which we have for our own use? In particular, do they result in higher efficiency of our activities? We will explore this in the next section.

We may consider whether the agent should have pleasure-generating centers in addition to pain signaling. We have already mentioned that pain relief is can be perceived by the agent as a reward, but wouldn't it be better if the agent simply received the reward signal? This is how it is done in reinforcement learning. The agent learns to maximize the reward, and if it can get several different rewards, it chooses the action for which the value of the reward is the highest. What can be wrong with that? An additional argument for using the reward signal is that in the human brain, there are centers that secrete dopamine, serotonin, and other neurotransmitters associated with the reward. If these centers are directly stimulated (e.g., chemically or electrically), the pleasure feeling arises. Wouldn't it be a wonderful world if pleasure were at our beck and call? Perhaps it would be a good solution in a world where there are no resource constraints and in which everyone is good and kind to each other. However, as we said, an unfriendly environment is needed for the development of intelligence; without threats to health and life, there would be no progress, and without signs of pain, there would be no compulsion to seek solutions—there would be no intelligence. For now, such a world exists only in fairy tales.

In the brain, dopamine functions are associated with reward signals. In the context of the reward of learning, dopamine also acts as a prediction error signal in a situation where the reward is unexpected. Predicted rewards do not produce dopamine, but rewards that are unexpected or greater than expected produce short-term increases in dopamine, while skipping the expected reward causes a decrease in dopamine release, below its normal background level. This reward-prediction mechanism is implemented in the reinforced learning method. Since one can stimulate dopamine-secreting centers in the brain, one can artificially induce the impression of reward and the happiness associated with it. Is this not how narcotic drugs work? So what is wrong with that? The fact is that in this difficult-to-survive world, this effect would not be a measure of success but would rather indicate the avoidance of problems that should have been solved. Drug addicts not only lose their will to act, but, addicted to the drug, they also lose the will to live and create, and they lose family and friends. Their worlds die around them with their consent. In one animal experiment, a rat with electrodes implanted in a dopamine-secreting part of the brain would activate it by pressing a lever in its cage. When the rat discovered that it could stimulate its reward center, it did nothing but stubbornly repeat the stimulation, giving up eating and drinking. As a result, it died (probably happily); the uncontrolled reward system led its body to destruction. In reinforcement learning, there is no limit to the reward—the bigger the reward, the better.

If we want to build machines with habits and weaknesses similar to people who act against their own interests in certain situations, then perhaps we should equip such machines with a reward system. Don't we feel better among people who have weaknesses, which we understand well? Would it not be easier to coexist with such machines, knowing that they are also craving the reward, that they can be intoxicated or bribed with a reward—machines in which we can find kindred spirits that understand that we can behave irrationally? Maybe so. However, until we have mastered the technique of building rational, intelligent machines, and until we understand what dangers may arise from their intelligent, autonomous behavior, we suggest not engaging in such experiments. It is easier to cooperate with someone rational than with someone whose method of reasoning, system of values, or motivations deviate from the known social norms. At the beginning of the era of conscious machines, there will be enough differences in the way of understanding the world and our role in it. Why hinder coexistence with artificial entities?

A mechanism that reduces the pain signal representing unmet needs suffices to control the conscious system, to develop its motivations and manage goals. The human organism also does not seem ready for unlimited reception of the reward, and the reward signal is to teach him favorable behaviors. Psychologists know that limited rewards provide the optimum stimulation for learning. The monkey learning new tricks only gets a drop of sugar water as a reward when it performs its task well. Further desire for the reward maintains its cooperation with the trainer. If she could drink her fill, she probably would not be interested in further cooperation. The human body has built-in defense mechanisms against overdose of the reward. After euphoric sex, the body needs to rest to feel the excitement and satisfaction again. After drunkenness, the head hurts, the reactions slow down, the speech is slurred, and thinking is disturbed. Our consciousness tells us that we are not ourselves. These defense mechanisms are not always enough to stop us from serious harm—hence, warnings and prohibitions. You should not drive a car after drinking alcohol, because it creates a road hazard. Drunken drivers react more slowly to changing road situations. Their consciousness is disturbed. But what does this mean? Does that not mean that they are differently conscious? We have different levels of consciousness, and sometimes a level is so different that it is clearly noticeable to outsiders who know our normal behavior.

Are there any socially beneficial effects of irrational behavior, including actions to the detriment of one's own body, which result from the reward overdose? Do joint feasts and celebrations strengthen social bonds, build trust and identity with a given group of people—with a sports team, with a nation? All these result from the desire to obtain a reward in the form of social acceptance. In turn, the social ties binding a group or a nation can lead to acts of heroic defense against an enemy attack. Japanese kamikaze pilots sacrificed their own lives in order to destroy enemy ships. Acts of self-immolation are powerful signals of opposition to war or social injustice.

On the other hand, a system that maximizes its rewards leads to an unlimited collection of resources, unlimited ambitions, and religious fanaticism, so it is a serious social threat. An economic system that allows uncontrolled collection of wealth by a small group of people is potentially socially unstable when the differences in possession of goods lead to the impoverishment of a large part of society and cause protests and riots. Religious fanaticism compels people to sacrifice their own lives in acts of terror in the name of eternal happiness in heaven. Perhaps from a social point of view, the desire to obtain rewards and not only to satisfy needs may be beneficial. However, it creates clear threats. Are we ready to install it in intelligent machines? Let us protect the world from the bane of fanatical machines. Fanatical people suffice.

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

Curiosity is not aimed at a specific goal. When we start exploring, we do not know what we will discover. A child or animal, curious about the world around it may explore without a specific need. There is nothing wrong with that if one knows little about the world and has time to learn. However, in everyday life we have to solve many specific problems. There is a small chance that by learning aimlessly about the world around us, we will also learn how to solve the problems that lie ahead. But we have a much greater chance of finding out how to solve such problems if they become the goals of system actions, and if, instead of being rewarding for discovery, the system receives larger internal rewards for achieving the goals it sets for itself. In reinforcement learning, the prize is awarded only for the goals set by the designer (primary goals). We argue that a system that sets itself internal goals (higher-order goals)—and gets an internal reward for achieving these goals—is able to learn better in an unknown environment (i.e., is more intelligent).

Let's think how an artificial brain can generate higher-order goals. Do we know the mechanisms by which we set ourselves such goals—goals not directly related to the needs of homeostasis or procreation, goals built into our bodies' homeostatic system of penalties and rewards? Although we do not know with certainty how our own brain does it, we can postulate how an artificial brain can do it. An embodied agent has a built-in system of needs, for which it receives an external reward. If these needs are not satisfied, pain signals increase, and the agent's goal is to reduce these pains. The reduction of pain (or getting the reward) will occur after the agent performs the correct action. This (getting the reward) is the basis of reinforcement learning. In order to go beyond the primary needs (set by the system designer), the agent may introduce an abstract need binding the received reward with the possibility of its re-receipt. After the agent learns the rewarding action, the impossibility of doing this in the future will cause abstract pain. The new abstract need of the agent is to ensure that in the future, it will be able to perform the activities for which it was awarded the internal reward.

There may be many future conditions under which the agent will not be able to perform the activities for which it received the reward. For example, if an agent has satisfied the need to keep its energy resource at the appropriate level by finding a source of energy (e.g., a battery) and then recharging from that source, then predicting or estimating the impossibility of doing this in future is his new abstract pain. Under which conditions that may happen? An agent may not be able to find a source of energy, and even if it does, it may not be able to apply it if, for instance, a plug used to connect to the power source is broken. No matter the reason why the activity for which it received the reward cannot be repeated, it is a source of abstract pain for the agent, and its abstract need is to remove obstacles that it senses or predicts. However, the agent does not have to wait until it has to fulfill its lower-order need to become interested in a higher-order need. On the contrary, the agent assesses the state of the environment (and its own state) in terms of its ability to meet lower-order needs. If, in its opinion, it will not be able to repeat this activity, then a higher-order need arises, which the agent itself supervises and for which it receives an internal reward.

This scheme is fundamentally different from hierarchical reinforcement learning, where in the course of one goal, the agent can set up subgoals whose implementation leads to fulfillment of the basic goal. Such a scheme leads to the breakdown of a more complex task into subtasks and does not lead to the creation of abstract goals. In particular, the agent performs subgoals only during the implementation of the original goal, for which it receives the reward. There is no need for any subgoal when there is no need to achieve the initial goal. To illustrate, imagine that the agent must eat in order to avoid being hungry (original need). After eating food, the agent's blood sugar level rises, and the pain associated with hunger goes away (reward). If there is no food, the agent cannot satisfy hunger. If the agent is looking for food when it is hungry, it implements the concept of hierarchical reinforcement learning. However, an agent who has created an abstract need (getting food) looks for food when there is no food in stock, not only when it is hungry.

Having abstract goals, the agent learns how to achieve these goals, and each time it achieves the abstract goal, it learns. After learning how to meet a higher-order need, the agent may generate the next-level abstract need, if it estimates high likelihood of that performing desired activities in the future may be difficult or impossible in given environment conditions. This impossibility of performing an activity needed to satisfy an abstract need creates an abstract (higher-order) pain, which raises another (abstract) need for reduction of this pain. For example, if an agent learned that money is needed to get food, then the lack of money in the future generates the abstract need for earning the money, which the agent can achieve, for instance, through paid work.

In this way of learning, the agent's goal is to meet the higher-order needs. The signals that trigger these goals are abstract pain signals that result from the assessment of the agent's ability to meet its needs. The greater the pain, the faster the agent will pay attention to the need to lower this pain. The agent's motivation to act in the environment is to satisfy all its needs (and minimize all kinds of associated pain signals); processes associated with this type of learning constitute motivated learning.

Mathematically, these two learning systems—motivated learning and reinforcement learning—treat the reward signal differently. While in reinforcement learning, the goal is to maximize the (usually one) reward signal, the goal of motivated learning is to minimize (usually the greatest) pain signal. In motivated learning, the reward signal is equivalent to a reduction in pain. It is easy to see that the reward is limited. When the pain signal is reduced to zero (or below a set threshold), there can no longer be a reward associated with the reduction of this signal. In motivated learning, we have a natural way of choosing variable goals of action. Once the dominant pain is reduced, the system can switch to the minimization of another pain signal (which dominates after reducing the previous dominant pain), changing its goal. The goals are thus determined autonomously by the system. This is important because the motivated learning often leads to the emergence of new abstract goals, and the choice of what goal to pursue (which need is the most salient) at a given moment is very important for effective action.

In the described scheme of motivated learning (Starzyk 2012), the pain signals compete with each other, setting the actual goals of the agent's work. In the simplest scheme of motivated learning, the agent tries to minimize the strongest pain. This leads to a simple system of managing the agent's motivations and actions. However, other motivation control schemes can and should be considered. For example, at work (Graham 2015a), we showed that the agent can more effectively meet all of its needs if it takes into account the state of the environment, its own location, the cost of activities, and other aspects of achieving goals. Then the agent can decide that it is more profitable to pursue a different goal than the one with the dominant pain. Motivated learning is more akin not only to how people solve their problems but also to how they set their own goals and how they organize their work while having many different goals in life. As we have shown, in many tests in a simulated environment (Graham 2015b; Starzyk 2017a, Starzyk 2017b), motivated learning leads to a more effective resolution of the agent's needs.

The more complex the environment in which the agent works, the greater the advantage of motivated learning over reinforced learning. In motivated learning, each need has an associated variable pain signal. These signals result from the assessment of the conditions of the environment and the condition of the agent and are measures of how difficult it will be to perform the action necessary to meet a certain goal. The harder it is to do this, the stronger the pain signal. In addition, the strength of the pain signal is also affected by the importance of the need that this pain represents. The more important goals for the agent will cause stronger pain signals. Because the agent can have many different needs at any time, the agent's evaluation with pain signals facilitates the management of these needs and is one of the causative forces motivating the agent to act. The actions that will be taken by the agent in a given situation depend on its needs, the state of the environment, and the assessment of how such needs can be realized. The agent uses pain signals to assess the situation, plan, and make decisions. Potential changes of these signals, related to the activities planned by the agent, are critical from the point of view of controlling the agent's thought process, its attention and choice of action. They are used both for planning actions as well as for responding to what is happening in the environment and what may result from the actions of other agents.

Feelings and Emotions

Since pain (determining unmet needs) is a signal motivating the agent's actions, what is the role of feelings and emotions? Where do they come from, and what is their importance in the agent's learning process and activities? In particular, what is their role in the artificial system, and how can it lead to conscious sensations? In part I, we recognized qualia as a necessary requirement for conscious sensations. The qualities of living organisms are related to their needs, but after all, sensations such as the color of light, sweetness of an orange, or fragrance of a rose are difficult to associate with needs without complicated philosophical constructions.

%%

In the artificial system, we have introduced a motivation mechanism based on pain signals. These signals were enough to build a coherent system of motivations, emergence of new needs, and determination of the goals of the activity. If pain reduction is enough, what could be the supposed role of reward and punishment signals?

Comparing our learning motivation system with reinforcement learning, we pointed out that pain reduction is better suited to constructing stable learning systems and developing agent motivations and needs than maximizing reward. We pointed out how a pain-based agent can create new needs unknown to the designer, depending on the relationship between objects and other agents operating in the environment. We have already written extensively about why such a choice of agent learning and development methodology is more justified in the case of machine learning than reinforced learning. But can we introduce a reward signal in such a system, a signal that is different from pain reduction? How do we incorporate a positive reward signal into a learning mechanism based on pain minimization, and what should the source of such a reward signal be if it is not the agent's need? The answer is an award that is represented by a positive signal as opposed to the negative pain signal used to represent the agent's needs and motivation. We don't want a mechanism based on maximizing such positive signals, because we know that it leads to unstable systems and that maximizing rewards does not lead to the development of motivations and goals and their efficient management. But we also do not want to deprive ourselves of the satisfaction of eating a favorite cookie, savoring the taste of ice cream on a July afternoon, admiring the landscape, or feeling the elation of love and romantic dreams or plans associated with it.

How can we determine the source and the real role of reward and punishment signals observed in humans and animals stimulating pleasure centers in the brain? Let's take the example of a hungry jackal who looks for food and finds a dead hen. His motivation is to find and eat food. But the jackal sniffs it carefully before it begins to devour its prey. If the hen stinks, it means that the meat is spoiled and that eating it, instead of satisfying hunger, might cause pain or even fatal intoxication. A good or bad smell does not satisfy any of the jackal's needs, but it is a signal predicting reward (satisfying hunger) or punishment (causing new pain). So stimulating the sense of smell can be considered pleasant or not depending on how the sniffed object affects the body's needs. But can we consider the unpleasant smell as a signal of punishment and the pleasant one as a reward? Probably not, if the reward is to represent meeting the needs. Pleasant smells could be a preview of an actual reward (fulfillment of needs), and an unpleasant one might warn of real punishment (a signal of pain). This is one of the possible ways to reconcile the needs and feelings, which after all do not satisfy needs in themselves. In a similar way, the smell of a ripe apple encourages one to eat it. The smell of a loved one's hair could be a promise of sexual intercourse. Similarly, we can relate to feelings provided by other senses. If the sugar-sensitive taste buds are irritated, it means that the picked fruit is ripe and suitable for consumption. If the taste is bitter, it is better not to eat it.

The signal provided by the senses to anticipate reward or punishment (pleasant or unpleasant feeling) is therefore only a measure of whether the need will be met or not. This is a signal related to the physical properties of the observed objects, which is directly related to the needs of the agent, although it is not a signal of satisfying these needs. The reward signal understood in this way is a kind of consent to the action that leads to satisfying the need. The presence of a pretty woman (associated with the reward) encourages the man to have sex with her (meeting the needs). The smell of browned roast is an incentive to eat it, but the smell of burned meat can be a signal to deter eating. A warm quilt (associated with a reward) is an incentive to cuddle up and spend the night under it (meeting the need for sleep). The softness of the surface being touched means safe gripping, in contrast to the sharpness of the blade, which requires maximum caution when handling it. The snake's hiss is a warning against deadly danger, and the sound of the waves of the sea is associated with the safety of the mother's womb and the beating of her heart. Observed colors are a signal of whether the fruit is ripe, whether the girl is healthy, whether the mushroom is poisonous, or whether the neighbor is angry. The sense of balance warns us about a painful fall and forces us to take action to save ourselves from this fall.

So pleasant or unpleasant signals provided by the senses are different than the reward for meeting the needs, and thus we have to name them differently. The sensory-generated pleasant or unpleasant feelings result from the genetic development of sense organs, which signal that what we perceive is useful for meeting physiological needs, as we described in chapter 4. The senses provide an illusory reward, which is a prediction of a reward but which in itself is not a reward. Let's call this kind of reward pleasure (or pleasurable feeling).

You can use pleasant or unpleasant sensory feelings in learning to encourage or discourage certain activities. Electrical current stimulation, used in experiments, is a known form of punishment for improper behavior. This unpleasant feeling is, however, a direct stimulation of the pain signal, and the resulting learning can be classified as learning by minimizing pain. However, associations that are learned can be transferred to any situation in which we want to evoke the desired behavior. A milder form of learning by associating with pleasant or unpleasant impressions is to use sound signaling, as in Pavlov's experiments with dogs. It is these associations with pleasant or unpleasant impressions that, like Pavlov's experiments, became the basis for the development of reinforcement learning, where the right behaviors associated with pleasant sensations are rewarded and the wrong ones punished.

Sometimes the difference between pleasure and reward seems to blur. Let's take an example of licking ice cream. Pleasure is perceived by the senses in the taste of ice cream, while the little ice cream that we consume is a part of the reward. Devouring ice cream can be satiating (fulfilling the need), but at the same time, further food ceases to be pleasure. Wine tasting is a pleasure (stimulation of the taste buds) but does not have to be associated with drinking all the samples that could lead to getting drunk (meeting the need of drinking wine). Therefore, wine gourmets, after tasting, spit out the contents of their mouths and are ready to taste other wines. On the other hand, the pleasant aroma of cooked food does not lead to satiation, and here it is easier to separate the pleasure of sniffing the dish from satisfying the need in eating it.

Since pleasure is the promise of meeting a real need, administering it is the best way to systematically encourage activities that result in receiving it. In primate learning experiments, monkeys only receive a drop of fresh water after performing each desired action, not to satisfy their thirst but to keep them interested in participating in such experiments. However, in light of what we wrote in the previous paragraph, it is not really a dose of pleasure but a dose of reward. Pleasure does not accumulate as it does in the case of maximizing the reward in reinforcement learning. We are not dealing here with maximizing rewards but reducing the need. So there is no fear that the system will become unstable, maximizing the reward indefinitely.

Similarly, the best punishment that can positively affect a child's behavior is not a painful loss but taking away small pleasures so that much remains to be lost. Punishment that is too painful can have the opposite effect in triggering rebellion, stubborn silence, and misunderstanding of the punishment. Very painful punishments leave a mark for life, but they do not necessarily lead to the desired change in behavior. Many criminals were unloved in their youth, harassed, beaten, and abused by their loved ones. Considering the enormous damage done to the minds of children, societies rightly impose high penalties for child abuse.

Finally, pleasant or unpleasant feelings are directly related to the perception of the world, seeing it as it is, with qualia, with perceptual and phenomenal consciousness. This is done by giving meaning to all sensations and perceptions, as described in part I. These feelings are gradual and thus allow the evaluation of sensations. It is to distinguish these pleasant and unpleasant feelings that we discern colors, the sound of the wind, the smell of a rose, a warm touch, and the hardness of a rock. It is for such an understood award that we overcome shyness in approaching a beautiful woman or before a public speech. It was for such an abstract reward that people have been willing to be a kamikaze, commit hara-kiri, or engage in war or suicide attacks. It is for such a reward that we work, invest, start families, and organize countries. Let's treat pleasant feelings as promises of meeting the needs and unpleasant ones as a threat that we won't get what we want. The system of motivations and goals can still be based on unmet needs, represented by pain signals, but the system is supplemented with pleasant and unpleasant feelings that are used to recognize the possibilities of achieving goals.

Pleasure is not the "reward" that is used in reinforced learning. There, the more rewards the better. Mathematically, the real reward we use to develop motivations and goal setting results from minimizing the negative pain signal. Its level depends on the level of pain (i.e., the agent's condition) and changes over time. A pleasant feeling can be compared with a derivative of the pain function. It mainly depends on the properties of the perceived object (to what extent it is suitable to meet the needs) as well as on the state of the agent (whether it has a need that this object is able to satisfy), and it also changes over time. But here it is not true that the more the better.

A pleasant sensory feeling is how quickly the pain signal can be reduced. The pleasure of consuming ice cream means that the blood sugar your cells need is quickly delivered. The sensory feeling is directly related to perception or phenomenal consciousness. Its gradation and variability over time does not depend on the realization of these sensations but is ontogenically regulated by the body and its needs. Hence the change in the sense of pleasure (the taste of food) as the degree of saturation increases; hence the change in the sense of pleasure in a romantic relationship (the euphoria of lovers turns into a much less intense experience of happiness as the relationship continues).

Sensory pleasure has different values, shades, and changes even with the same stimulation (e.g., with the same source of smell). What is the gradation of this pleasure? Let's take an example of food. We can eat different things, but we like some more (greater pleasure) and others less (less pleasure, with the same degree of unmet need—pain). A juicy steak tastes better than plain potatoes. Why? Both eating meat and potatoes will reduce hunger pain, but eating meat will provide the body with the necessary calories faster. However, the pleasure of eating does not accumulate; we do not fall into euphoria as we eat. On the contrary, we feel more satiated, and the need decreases. Pleasure, which is the feeling of the taste of the food eaten, may also decrease. The first bites of steak taste much better than the next ones.

Due to the suggested functional separation of pleasure and reduction of the pain signal (meeting the need), pleasure signal manipulation may occur. If we drink tea sweetened with sugar, then the signal from the taste buds is a scam and does not represent a promise to satisfy hunger. People who like sweet tea but watch the amount of calories consumed cheat their own body by experiencing a pleasant feeling. The use of drugs (a way of delivering pleasure), which often mushrooms to the limits of the endurance of the body, is a scam and does not really meet any of the body's needs (except to escape from problems and drift away into the unreal world). Even painkillers deceive the body that nothing is really happening, that we don't have a terminal illness, that everything will be fine. Everything will not be fine—we will die of cancer—but at least we will not suffer. Thus, not all scams are bad. Sometimes it is better not to know the truth about certain things, especially when we have no influence on them. Nature created pain to warn of danger to the body, but humans overcame nature and learned to block this signal. This allows one to perform surgery and save lives, put away suffering, and forget. That is why societies legalize the use of cannabis, use opioids to control pain, and tolerate drunkenness.

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

Emotions are a reflection of the state of mind. We experience many different emotions, with higher or lower intensity or duration, and they are caused by various reasons. Emotions can be externalized or hidden and have different effects on our behavior, social relations, and decision-making. They include such states as love, jealousy, fear, panic, anger, anger, joy, contentment, happiness, despair, sadness, longing, nostalgia, irritation, anxiety, greed, occupation, disgust, compassion, gratitude, trust, grief, guilt, hatred, hope, pride, shame, and so on. We can define "curiosity" as an emotion associated with the drive to explore the external environment or the internal knowledge stored in the mind. However, due to curiosity's specific role in the learning process and motivation to take actions, we classify it as a need in our artificial mind model. Additionally, because in motivated learning, motivations for action are specific needs (including the need to satisfy curiosity), we do not consider emotions as motivations but only as modifiers of decision-making or activity-planning processes.

It is true that we react emotionally to certain threats and that these emotions will modify the actions taken to avoid these threats. An example is the feeling of fear we have when a bear appears in our path. This fear can paralyze our movements, accelerate the rhythm of our hearts, or cause an adrenaline rush, but the plan of what we have to do to get out of trouble must be based on a rational assessment of the situation and a logical analysis of how to remove the threat. Therefore, the reaction to a threat has two components: a decision about a specific action (e.g., escaping from a dangerous place) related to satisfying a specific need to ensure safety, and its implementation under the influence of emotions of fear. The motivation to act will be to remove a specific threat (e.g., escape from a burning house), not an attempt to change the emotional state that is connected with this threat. Emotion may influence the choice of action (fight or flight response) and also may change as a result of such action.

These types of motivations include other psychological behaviors. For example, if you run away without thinking, this is an unconditional reflex that can be called instinctive. Unconditional reflexes can be hard-coded in motivated emotional mind as an automatic backward movement in response to pain. However, if the action occurred after reflection, it is the result of a conscious decision to avoid the threat. Emotions arise as a result of a conscious assessment of the situation or as a result of the subconscious process of responding to threats. In fact, most emotions are subconscious. They result from internal brain processes that are not consciously observed or controlled. They reflect the state of the mind that escapes conscious control. Emotions are all-encompassing and affect the method and time of responding to stimuli, perceptions, decision-making, and assessment of the situation. So they modify the thought process in a very general, nonspecific way. Emotions affect the general mobilization of the body's resources to either attempt to fight or help escape from the place of danger. However, the dominant goal of a conscious mind is to seek optimal solutions in the context of current situation and under the influence of emotions.

Often, emotions accompany subconscious reactions to what has happened. For example, we may reflexively evade a stone falling on us during mountain climbing, and the emotion that arises results from realizing what would happen if this stone were to hit our head. Different people will react differently to this situation. Some with stoic calm will say that nothing has happened, and others will be shaky and unable to continue their climbing. Some emotions, such as the fear of a snake, seem genetically conditioned. Perhaps avoiding snakes was so important to the survival of the species that emotional reactions evolved and were genetically coded as automatic panic at the sight of a snake. It is known that the mere sight of a snake can paralyze many of its victims. Are they paralyzed by the emotion of fear? Is this paralyzing fear to protect them from attack? In the biological world, playing dead can sometimes save a life. Isn't this emotion a defense strategy elaborated upon by nature?

These considerations raise the question, Should machines also have emotions? Not fake smiles or frowns, artificially generated by human-looking robots, but genuine emotional states that relate to their consciousness? For years, attempts have been made in robotics to design machines that pretend to understand human emotional states or demonstrate sympathetic reactions to humans. However, these are not real emotions (i.e., the robots don't actually feel anything). The people who build the machines design programmed responses, understanding that people will trust robots more when they display apparent emotions such as contentment, surprise, or compassion.

People experience and communicate a rich variety of emotions. Laughter indicates joy and communicates friendly intentions, while crying indicates sadness, dissatisfaction, or frustration and often evokes a reaction of compassion or assistance. Accelerated breathing and a sinister facial expression indicate anger and can be interpreted by an opponent as a sign of an impending attack.

It is difficult to assess the reasons for the emergence of emotions and expression of emotional states. Although emotions can be shown—or suppressed—consciously, many are not subject to conscious control. It is not always easy to determine the role of emotions, yet they clearly serve an evolutionary purpose: they help increase the odds of survival. Animals express emotional states to scare away predators, win partners, and gain acceptance by social groups.

Although many complex emotions have evolved, this does not mean that they are all necessary or useful for robots. Sometimes expressing emotions is useless or even harmful. Feelings of embarrassment can paralyze a shy person preparing to speak to a group of people or establish intimate contact. Do we even understand all the social consequences of emotion, including those that, like shyness, appear to have little utility? Evolution and natural selection seem to confirm the usefulness of most features associated with the development of species, even if we do not fully understand their roles.

In addition, these are human emotions related to the human body, the way a human perceives the world, his or her goals, and methods of survival. Because machines have a different body structure and other possibilities of impacting the environment, they do not have to benefit from having human emotions. Before specific emotions and their mechanisms are introduced into the organization of artificial minds, their consequences should be carefully examined. After all, we cannot count on the natural selection of machines to establish suitable emotions. Machine engineers will have to equip them with the mechanisms of producing desired emotions, but only after testing how they affect the machines' behavior and ascertaining the conditions under which they are beneficial. So should machines have mechanisms to create emotions? What emotions are to be used? Emotions' roles and the methods of their creation are important to consider when creating artificial intelligent entities.

Emotional states are experienced and felt internally, while their physiological reactions can be demonstrated externally—and rather subconsciously. Moreover, emotional states can be perceived by realizing the body's reactions, such as having sweaty palms or trembling from fear. The state of strong nervousness can be confirmed by the measurement of one's blood pressure or increased heart rate. One can describe the feelings of happiness and/or love with full consciousness, especially when there are cultural patterns that describe the state of elation, total acceptance, and the feeling of being loved. These subjective feelings pertaining to the recognized emotional states are qualia, yet they concern internal states of the system. The familiar feeling of a tightened throat associated with panic and/or anxiety, a feeling of nausea associated with disgust, or a feeling of pressure in the stomach associated with mourning are examples of such feelings related to the emotional qualia.

In terms of artificial mind architecture, we are interested in the mechanisms of emotions and their impact on one's behavior and decision-making. It is believed that emotions are responses to sensory experiences. However, they differ from the qualia, which are the subjective impressions of these experiences. Qualia resulting directly from observation do not have to be associated with emotions. Observing a table's roughness or the coolness of a metal handle does not necessarily evoke emotions. In such situations, the role of qualia is similar to their role in causing pain, pleasure or discomfort. Feeling the heat radiating from a stove prompts warmth (and may invoke a pleasant emotional response), whereas touching a hot iron causes a feeling of pain. In case of the former, it is just a pleasant feeling that triggers no action, while for the latter, the consequent pain results in certain reactions of the body and is a motivation to act. Similarly, observation of a person crossing the road may leave us indifferent, but seeing the same person being run over by a car triggers a strong emotional response. Pain is associated with unmet needs and motivates one's body to act in order to minimize it. If we do not have money and we need it, we seek ways to earn it.

Emotions can be results of complex interactions and effects of many interacting processes. Most often they do not constitute specific motivations for action but only modify the process of conscious searches for ways out of situations that caused these emotions. We do not always know how to control the destructive feeling of jealousy; it is not always the result of what we do. Our beloved can, whether intentionally or unintentionally, please many men and/or women. Seeking a solution for growing conflict can lead to the loss of love for a loved one. A feeling of anger may result from inept attempts to find a solution to the problem; whom are we to blame if these attempts fail? Feeling despair after losing a loved one cannot be removed by chopping wood or abusing alcohol. This emotional state can change our view of the world for a long time—the willingness to act, the organization of work, our interests and life plans. Emotions, although often the results of specific actions or events, usually do not lead to constructive solutions. Longing expresses the state of melancholy for what we do not have and often cannot have. Greed is a state of mind through which we filter our decisions. Should we treat this feeling? Perhaps, if we understand that greed prevents our being accepted in a group, that there are people who wish us badly. But if we somehow get rid of greed, we simply remove the modeling effect that it had on our way of thinking and acting. Therefore, in our architecture of consciousness, we treat emotions as behavior modifiers.

Similarly, we differentiate between emotions and needs. The needs are associated with specific goals. They result from lower-order needs, and their fulfillment indicates a successful action. Emotions arise from actions and observed events; they are modifiers of behavior, thought processes, and decision-making. Strong emotions, such as fear or anger, are often associated with the acceleration of the decision-making process, omitting detailed considerations of the social effects of actions, with aggressive acts, extortion, or forward pressure. Calm emotions—such as sadness, longing, nostalgia, or compassion—increase our sensitivity to how our actions affect others, cause greater caution, and lengthen the decision-making cycle. Passions, such as love or jealousy, increase susceptibility to retrospection, to experiencing and observing our own moods, but also are the cause of extreme efforts and unusual solutions to evoke admiration.

At this point we can consider the need to introduce mechanisms for mapping and using artificial drives that determine the behavior of living beings to such a large extent. According to Sigmund Freud, the forces determining human behavior are the drives characterized by the source, goal, and object. The source of drive is internal and bodily—the goal is to relieve tensions and get pleasure—and the object is the person or thing to which the action is directed. Freud distinguished two main drives: libidinal (sexual need; the search for love and pleasure) and aggressive (all destructive tendencies, including autoaggression). The basis of drives is biology, of which our artificial systems are completely devoid. Do we have to introduce them in the machines? What would their purpose be? In motivated learning, *drive* is a growing signal of unmet need. The sex drive is one such signal. All this can be represented and dealt with by the same motivational mechanism: pain increase and pain reduction. Finding the reward, as we wrote, is not preferred in motivated learning. In man, after satisfying a need (e.g., sexual drive), there is a decrease in this need (corresponding to pain reduction). There is no continuation for this need (continuing to maintain the drive or maximizing the reward); otherwise, the body will be exhausted and eventually destroyed.

Simple emotions, such as fear or disgust, can be subconscious reactions to stimulation. These simple emotions are associated with the subconscious processes of the body's reactions to harmful substances or sensory arousals. The foul substance causes an immediate reaction and a sense of disgust, as does a feeling of fear evoked by the sight of a snake. However, the majority of complex emotions require an assessment of the situation in relation to one's own needs and cause modifications of behavior according to the situation. This assessment is also associated with the ability to observe changes in the environment and, in particular, the reactions of other people to our activities. In this context, expressing emotions through changes in voice articulation, posture, or facial expression are manifestations of the form of reaching an agreement (Griffiths 2005). Smiling at a nice girl encourages her to start a conversation. A threatening face is used to scare away an intruder. A gentle voice conveys a peaceful attitude and promotes building trust.

How can a machine integrate observations to evoke emotions on one hand and to use them to modify its behavior on the other? Because the types of emotions are extremely diverse and have different backgrounds, mechanisms of formation, and impacts on the body, we are not able to reproduce them all. We can only suggest how some of them can be made in the machine. Remember, the machine will not have to create human emotions. In addition, a machine's emotions can be expressed in a different way than people's emotions and can have a different impact on its behavior. However, when describing how a machine can create emotions, we will stick to the designated pattern of creating emotions and their importance for system functioning.

Let's start with a simple example of creating complex emotions that require assessments of situations—for example, irritation. Artificial irritation can be caused by repeated unsuccessful attempts to do something the machine wants to satisfy its need. After observing another individual performing a useful activity, the machine would like to perform a similar operation (perhaps using the concept of mirror neurons), but its skills are too limited. It repeatedly tries to recreate a series of observed movements, each time watching its defeat. Because the machine does not understand why its action failed, it generates an internal signal of emotion related to failure. This signal is generated as a result of a failed action and accumulates in the center of emotion called artificial irritation. Such a center needs to be previously designed and programmed, and the designer must determine how to accumulate a parameter (e.g., the number of unsuccessful attempts) that will indicate this artificial irritation. In contrast to the reward or punishment signal, the artificial irritation signal is independent of the type of action being performed. The machine knows what it should do, but it fails; the signal of irritation is an expression of its failure. It results from disappointed expectations. With each unsuccessful attempt, the artificial irritation signal will grow. This artificial irritation can cause a change in the rational strategy of achieving success or discontinuing further attempts. The irritation signal will disappear when success is achieved. The artificial irritation signal will gradually decrease if, for example, the subject resigns from activities that caused it.

Artificial anger can be generated in the machine by behavior by other agents that is unfavorable to the agent, as well as by its inability to perform tasks or the lack of expected results. As with artificial irritation resulting from failed actions, the cumulative anger signal causes aggressive behavior as well as modifications in planning and decision-making processes (attained in the machine by changes in threshold values). In addition, it may cause higher energy expenditure in preparation for the acts of aggression or fearful behavior to scare an opponent.

Another example is artificial satisfaction. This may be triggered by the machine's success in the activities it undertakes. In contrast to the irritation resulting from a situation in which nothing worked, here is the opposite: every attempt ends with success. Such an uninterrupted range of successes is easy to assess and to use to stimulate artificial satisfaction. The signal of satisfaction built in this way will influence the machine's behavior during periods of ongoing success. The function of artificial satisfaction can be fulfilled by another well-designed system stimulating the satisfaction signal. Proper design of the artificial-emotion mechanism should be as general as possible, correlated with motivations and observation of the effects of the action. After a good streak, the signal of satisfaction will gradually die out and may even have negative values, leading to reverse changes in the machine's behavior. What can the role of such artificial emotion be, and how will it affect the planning of actions or the way of making decisions? Such a satisfied machine may not be fully aware of what favors it, but the feeling that everything is going well under the circumstances may encourage it to take riskier steps.

Artificial shyness can be introduced as an emotion of a machine that has reached awareness of its own existence. A machine's self-awareness can lead to reflection on how it is perceived by others. Is she liked? Do others trust her? How is she comparing to peers? Can she cooperate with them? Do they approve of her actions? Is she a leader, or does she prefer to follow? If the machine is very sensitive to the opinions of others, it can develop artificial shyness, and its shy but considerate actions can be approved of and appreciated. There must have been an evolutionary advantage of being shy. In a society of collaborating individuals, not everyone can demonstrate self-confidence, be a leader, and dominate. It can also happen the other way around—its self-confidence will ensure its greater success. Thus, in a group, there is a need for both types of individuals: those who are shy and those who are self-confident.

Artificial fear may be associated with pain prediction. The machine, aware of the situation in which it is threatened with pain, remembers what it previously experienced in similar conditions and generates a signal of fear. It is an emotion that changes the way information is processed and the manner in which it responds to environmental stimuli. The machine, fearing the impending painful experience, quickly reacts to the approaching danger. Its attention is focused on what is happening in the environment, and it is ready to escape quickly from the threatened place or to take up the fight with the imminent threat. It is ready for increased energy expenditure to either quickly escape or attack the enemy, depending on its assessment of the situation. In humans this is associated with the secretion of hormones. Machines can change threshold levels when processing information, switch to more battery power, reduce reaction time, or reduce the time they spend for deliberation in considering various options. Such modifications of the machine's functioning are not difficult for its constructors and are relatively easy to program. The fear signal in the machine, similar to that of humans, increases the stress on the whole body. In this case artificial stress comprises excessive energy expenditure, higher voltage levels that control information processing, or current overload that may damage connections between circuit elements.

Among people, shyness or confidence may be genetically conditioned, and the usefulness of many emotions has been verified by evolutionary natural selection. In the machine, we must try out the usefulness of emotional states and their types. We must propose mechanisms for their buildup and their impacts on decision-making. Regardless of how we accomplish this, we can say that the machine is able to have emotions and that these emotional states can influence its way of associating, learning, planning, and making decisions. These impacts will differ from the impacts of the needs and goals that the machine implements. Unmet needs motivate the machine to act. With learned ways of achieving goals, the machine takes actions aimed at achieving a specific goal (e.g., how to get money or how to turn on the generator). Its method of planning, related to the implementation of needs, seems to be the equivalent of rationally made decisions. The machine tries to perform the necessary activities, expecting that after their completion, it will achieve a reward in the form of a realized need (reducing the abstract pain that accompanied the unrealized need). It expects success based on the fact that in the past, the implementation of these learned activities also brought success.

If these emotions are artificial, how much can we say about the similarity of the machine psyche to the natural psyche? We can assume that it will be different, that—just as in the case of a psychopath—the feelings that guide us will be foreign to a machine. However, for fruitful cooperation with machines, it will be necessary to recognize and respond to their emotions. Just as we have learned and are still learning how to respect the emotional states of animals and protect them from suffering, we will have to learn to recognize, consider, and influence the emotional states of conscious machines. Do we need to develop all the emotions we know in the machine (or emotional states similar to those we know among people)? Not everything that nature has produced for us is worth imitating, but before we reject any of the bad emotions, we need to know why such emotions have evolved in people and think about whether we understand all the consequences if we deprive our system of such emotions. Thus, designers must study how various emotions they put into machines' responses influence their psyche and their behavior. This will take time.

The influence of emotions on the machine's behavior will be different than it is for humans and animals. Satisfying specific needs may take place as a result of actions performed as schemata recorded in procedural memory (if they were routine procedures) or declarative-episodic memory (in the case of learned reactions). Emotions, on the other hand, affect all the neurons grouped in the neighborhood of emotional arousal. Emotional impact can be transmitted in the form of spatial signals reaching many neurons simultaneously. Functional synaptic coupling or other information carriers corresponding to, for example, potassium waves generated by stimulated astrocytes in natural brains may be used for this purpose. Even if emotions do not have a global impact, they largely influence many decisions. In artificial systems, such a quasiglobal interaction can be implemented, for example, by changing the threshold values of stimulated neurons in a region or changing the supply voltage of the electronic system implementing artificial neurons. Raising these threshold values (or reducing voltages) will reduce the number of options considered, as fewer of them will be considered sufficiently correlated with the selected objective. In this case, decisions will be made with caution, and only those that guarantee certain success can be chosen for implementation. Another example of global interactions that may result from emotions is the change in the rate of signal transmission between neurons, affecting the speed of decision-making or shortening the time to make a decision. This will influence the number of options considered before the action begins, after which the machine expects to achieve an intended goal.

Emotions should be propagated using paths other than sensory-motor signals that stimulate individual neurons. The proposed solutions depend on the hardware implementation of neurons. Neurons located in a specific neighborhood can have their transfer functions, threshold values, or speed of operation affected by signals representing emotions. Such solutions are often used in the design of integrated circuits, where local changes in the level of power supply affect the speed of the system or its energy expenditure. This does not mean, however, that the implementation of emotions requires a hardware implementation, because these changes can be simulated equally well in implementations of neural network software, in which the parallel processing of information in the network is emulated on sequential computers.

In practical implementations of intelligent systems architecture, experiments are needed that show the benefits of the machine's emotional states. In one of his works, Starzyk showed how an intelligent robot can build artificial trust and showed the benefits of using such a behavior modifier (Starzyk 2016). Trust can be a basis for the development of social relations, friendship, and cooperation. These and other emotions can improve efficiency of the embodied agent's actions in various environments, especially where other (perhaps intelligent) agents work. However, full consideration of the various types of artificial emotions requires careful research on their impact on the behavior of the machine. In addition, the generation of different emotional signals should be appropriately selected in order to organize the decision-making process and perceive the results of actions. Finally, the legitimacy of introducing a particular type of emotion should be documented by the benefits that these emotions can bring in the functioning of the machine. When designing artificial consciousness, we do not have as much time as nature needs to verify the suitability of emotional states in the process of species development, which is why intensive tests will be needed before introducing the emotional mechanisms in the machine.

%%

Motivated Learning Robots

Writing about minds we naturally must address the necessity of teaching them. Human minds require a long period of teaching a child almost everything—moving, grasping, walking, speaking—not to mention the many years of learning knowledge about the world, about society, and about oneself. Without this, children will not become full-blown adults, and perhaps they will not be able to survive on their own. Many animals come into the world with a resource of skills and innate knowledge, but for the hatchling to become an eagle or a gull, the kitten a tiger and the puppy a wolf, the animal must learn a lot from its parents, other members of the pack, or its community in general. We know that the higher the intelligence of the animal, the more complex the social behavior, the longer the period of education, and the greater the dependence on the parents and the learning period.

We have tried to show that if an intelligent system is capable of processing information in such a way that it can benefit from the acquired knowledge, then such a system must have the ability to record previous experiences, which means the ability to learn. What is more, we have demonstrated that to achieve deep knowledge about the world, a knowledge that allows for the development of logical relations between abstract concepts and underlying sensory feelings, it is necessary to link sensors and effectors to the mind with its embodiment. Such embodiment allows the self-learning process to be accomplished by manipulating the environment.

The question arises as to whether we also have to teach robots. We have already asked this question, suspecting that it is enough to build appropriate algorithms to control the robots, and they will do what is programmed for them. This claim is true. We already have such robots. They behave according to predetermined rules, goals and procedures. We have already given examples of autonomous cars, drones, and other vehicles. However, having such robots does not satisfy people who build them and want to use them to perform increasingly difficult tasks in an increasingly complex and unpredictable environment.

The implementation of tasks in city traffic or on the battlefield is already sufficiently complicated. And what will happen if the battlefield is crowded by other autonomous vehicles, which will also try to fool and mislead enemy combat systems? Robots designed for such sophisticated missions should be able to adapt to the changing environment and constantly develop their abilities to survive and achieve goals. That is why we will continue to deal with robots capable of learning and developing.

If the adaptation to the operating conditions is not evolutionary involving the principles of the genetic transmission of traits but instead involves open-ended learning in embodied machines, then such systems are called epigenetic ones. Epigenetic systems, both natural and artificial, are characterized by a developmental process in which the change in the behavior of the subject is the result of interactions between this subject and its physical and social environment. Epigenetic robotics has two related goals: 1) understanding biological systems through social science and engineering, and 2) enabling robots and other artificial systems to adapt to different types of environments, instead of programming them to function in a specific environment.

Epigenetic robotics explores mechanisms and architectures for lifelong learning of embodied machines in natural environments. Learning must lead to the gathering of new knowledge of increased complexity and is based on the exploration of the world and social interactions. The morphology of the biological system arises with the development of sensory, motor, and social skills. Describing the architecture of consciousness, we will be largely based on the achievements of epigenetic robotics. Development work is trying to imitate the development of young children, observing the acquisition of sensory and motor skills in embodied systems. Basic sensory-motor skills, such as hand-eye coordination, walking, approaching and manipulating objects, using tools, and so forth are acquired through interaction with the environment.

Another group of skills is communication skills, such as gestures, speech, or facial expressions. Symbol grounding is used to combine sensory observations with memory organization and skill development. Developmental robots learn from people how to participate in social situations through cooperation with human teachers and imitating people after they demonstrate their desirable behaviors. Teaching sensory and motor skills is facilitated by communicating these skills between people. Other areas of interest are social behavior, speech recognition and vocalization, concentration of attention, and development of empathy. Developmental robotics is also interested in higher-level cognitive skills. This includes attention, representations of higher-level social concepts and skills, empathy, emotions, and theory of mind.

We are trying to understand the cognitive mechanisms of a higher order in humans (Asada 2009), so before artificial systems can perform these functions, we must learn and understand how these functions are implemented in natural minds and how they can be imitated and designed in artificial systems. It is important to note that instead of developing each higher-order function separately in the form of a separate module, we should examine how the higher cognitive functions result from the dynamics of the complete system operating in the environment and how they interact with each other. We claim that only by creating an integrated robot control system designed to operate in a specific environment—capable of perceiving, creating representations, organizing memory, planning, and controlling the performance of activities, as well as predicting and observing the effects of actions—can we build intelligent robots. The artificial mind architecture discussed in this book is attributed to such an integrated concept of the development of intelligent conscious systems. At no time have the minds of robots been assigned the attribute of consciousness. In this book, we will try to test how the properties we postulate may appear in real constructions of artificial brains.

Intelligent machines cannot have skills, motivations, or higher cognitive functions programmed by designers but require autonomous (independent) development of these functions in their minds (Weng 2001). Developmental robotics is based on a specific machine-learning method, mentioned in previous chapters, known as reinforcement learning. Reinforcement learning is inspired by behavioral psychology, which deals with the agent's activities in the environment to maximize a certain type of accumulated reward. This allows the machine to automatically define the ideal behavior pattern in a specific context in order to increase the efficiency of its operations. For an agent to learn proper behavior, simple feedback in the form of a reward signal is required. This coupling is known as a reinforcement signal. Reinforcement learning applied in developmental robotics has been enriched recently by adding internal motivation based on curiosity (Oudeyer 2007). Motivated emotional systems, equipped with MEM, go further: they generate internal reward signals to discover new things, learn data compression, or enrich accumulated knowledge.

The development of social behavior requires interaction between people and allows robots to learn from people how to behave in social situations. It is worth recalling that the imitation of human behavior is one of the most effective ways of learning. There is a convergence with the cognitive behavior of people and animals, demonstrated in the learning processes, especially of young individuals. To achieve this effect in the robot's mind, the relationship between perception and action should be established so that the observations are associated with proper behavior. An unexpected benefit from developmental robotics is the fact that the obtained results can help psychologists understand animals' learning and perception (Asada 2009).

Developmental robotics also investigates the reuse of previously learned skills and behaviors in new situations. This includes the dynamics of the body, its shape and size, and the expected environmental impact, called morphological computations. Examples of morphological computations include displacement, in which the dynamics of the body are used to facilitate walking and gripping the observed objects. Walking is the result of mutual interactions between material properties, motor control, and the environment (flexibility, strength, friction, surface shape, gravity, etc.). In order to grasp an object, the agent does not need to know the shape of the object beforehand; the shape can be sensed during grasping, and is a direct result of the morphology of the object and hand (Pfeifer 2005). The use of morphological computations facilitates control operations and enables work in real time. Research on biomechanics confirms that morphology plays the same role in the rapid movement of animals, where the dynamics of the body are used to ensure precise control of the muscles, which leads to rapid action without a precisely calculated trajectory (Blickhan 2003).

Developmental robotics requires a mechanism that can set action goals and organize learning new skills, as well as identify subobjectives in complex sensory and motor operations (Baranes 2009). Psychologists have identified the internal motivation responsible for curiosity in people, and some forms of such internal motivations have been proposed in developmental robotics (Oudeyer 2010). Internal motivations lead to exploration in various subspaces in search of expected information gain (e.g., heuristics to maximize prediction errors, variances, entropy, or uncertainty) (Thrun 1995).

In 1981, Sloman and Croucher wrote an interesting article (Sloman 1981) wherein they argued why robots will have emotions. The basis of their reasoning was that for a robot to be intelligent, it must have different needs, including physical, spiritual, and social. These needs will create motivations for action. Old motivations will generate new ones, requiring a mechanism that enables this. New motivations differ from those originally set by the designer. We developed these ideas, and they took a form of motivated learning during which robots create new needs and goals based on previously developed needs. In the next section, we will describe how such motivations can be generated and how the robot can manage them.

%%

Motivated learning was developed by Starzyk as an alternative to the reinforced learning method (Starzyk 2011a). In many aspects, it is similar to reinforcement learning. For example, an agent learns how to achieve goals using a signal similar to the reward signal. But there are also important differences between these methods. In this section we will describe motivated learning and compare its properties to reinforced learning. We will show that this type of learning is close to learning natural systems and yields better learning results in artificial systems than reinforcement learning.

The creators of autonomous robots face a difficult and important problem: how to create a mechanism that will implement many functions related to the perception, processing of the reward signal that stimulates learning, management of competitive motivations, and creation of new motivations, as well as the control of machine planning and operation. The problem designers face is that the machine's casing is not a body equipped with nociceptors and other pain receptors. Pain in the machine must be made artificially in the form of an adjustable parameter measured in the artificially created pain detection center. An intelligent machine must be able to learn how to reduce external pain signals received from the environment and obtain internally generated abstract pain. External pain signals are predefined and connected to pain detection centers that trigger learning mechanisms whenever such pains increase (bad) or decrease (good). So the motivation of the machine comes from the response to changes in external pain signals. The machine will learn how to minimize these pains and thus learn about the principles of operation and the laws in force in the environment. The constant hostility of the environment can become the basis for motivation to learn and act, create goals, plan, think, and solve problems. Knowledge is a by-product of learning. Thus, it is not necessary to include preexisting knowledge in the machine's memory. However, some basic knowledge (such as breathing or sucking, or the reflexes of children) can speed up further learning in a smart machine.

Intelligence cannot develop without embodiment or interactions with the environment (Pfeifer 1999). Using their embodiment, intelligent agents act and cause changes in the environment through their actions. The environmental responses (including changes in pain signals) are recorded by sensors located in the body, the housing of the agent. At the same time, the embodiment of an agent is a part of the environment that can be perceived, imagined, and recognized by an intelligent machine, leading to self-determination.

But why are we doing anything at all? Why do we undertake activities sometimes subject to considerable effort? We are compelled to do so by unmet needs. In motivated learning, we simplified the description of all motivating signals to pain signals. The goal of the machine is to reduce these pains. We have done so to develop a unified mechanism for introducing abstract goals and managing these goals through signals that we compare with each other in order to decide on action. As for the simplification of functional mechanisms of the artificial brain, we do not consider the reward signal, only the pain, so we do not try to implement separate mechanisms of instincts, drives, and other such psychological terms. Instead, instincts in the machine may be treated as subconscious implementations of a strategy that reduces pain signals. If a machine tries to reduce a pain signal without thinking of how to do it or even understanding why it is doing it, is it following its instincts. This does not mean the machine *knows* that it is following something we call instincts. Its response is processed in subconscious, in which all sensory and pain signals are available, but a machine doesn't use its conscious mind to decide what to do. Yet it follows the need signals within its mind without focusing its thoughts on them. It uses gradients of changes in the pain signal as artificial feelings that guide its act of searching for the proper action.

Thus, we can explain the functionality and usefulness of instincts, feelings, and emotions for the machine in a uniform way through the mechanism of meeting the machine's needs. This does not mean that there cannot be other structures organizing the fulfillment of the machine's needs and that perhaps some of them may lead to better brain organization of the machine and more efficient solutions to its problems. Our task in this book is only to show that it is possible to motivate the machine to learn by using pain that results from unmet needs.

What is the role of pain in motivated learning? Although we can use both reward and penalty signals to train a machine, avoiding penalties may be enough to develop an agent (at least in simpler systems), and unlike maximizing rewards, this leads to stable systems with limited amount of satisfaction delivered. Philosophically, this may be at odds with the universally recognized pursuit of happiness, but one must distinguish between striving for happiness and remaining happy. We can quote the lyrics of the song by the Polish rock group Skaldowie: it's not about catching a bunny but about chasing it. There is a lot of life wisdom in this simple text, but not only that. In the process of searching for the desired outcome of action, the machine experiences feelings and uses them as guides for its learning. Pleasurable feelings are associated with increasing chances to achieving the goals while unpleasant feelings indicate that environment conditions are not conducive to achieving the goals.

Please recall that feelings are impressions that triggers no action, while pain related to the specific need, results in certain reactions of the body and is a motivation to act. So when we have a need we are guided by feelings, and if there is no need in motivated learning there will be no actions and no feelings. Artificial feelings may play similar role for development of machine consciousness as hostility of environment for development of artificial intelligence. While hostility of environment threatens such primary needs as homeostasis, and is conducive to development of intelligence in motivated learning, feelings built into the system primary needs are foundation of machine consciousness.

Since primary needs are set by the designer, only these needs may have built-in mechanisms to detect if the environment conditions are conducive to satisfy them. These mechanisms correspond to artificial feelings and do not require learning. Thus they are not cognitive. On the other hand, feelings associated with higher order abstract needs are cognitive and have to be learned.

In our understanding of the needs-control mechanisms, it is simply a better solution. A system remaining in a state of eternal happiness would not learn anything (did it already possess all knowledge and can automatically satisfy all needs without pain and associated feelings?), and therefore it would not be an intelligent system (although it could be infinitely wise). We know such infinitely wise systems only from mythology, such as the all-knowing Greek god Helios.

There are many instances of instability in award-based systems. We have already presented an example of a rat that electrically stimulates its reward centers while ignoring food until it dies of hunger. It is also known that drug abuse, which stimulates people's pleasure centers, can lead to death. Although we can always interpret the reduction of pain as a reward, maximizing the total reward leads to different solutions than minimizing the dominant pain (reduction of the negative signal). The first approach can lead to unstable systems (with an infinite reward), and the second stimulates the process that will be completed when the negative pain is reduced below a certain threshold. In addition, in multitasking systems, the pain-reduction mechanism provides a natural way of managing motivations and selecting the goal of action. Mathematical pain reduction corresponds to a mini-max problem, where the optimization effort focuses on the strongest pain signal and automatically moves to a different goal when the dominant pain represents another need. This approach allows one to create a simple system that is able to simultaneously manage multiple objectives.

In the proposed motivated-learning approach, the machine uses neural structures to self-organize the system of motivations and goal creation. The system stimulates motivations and creates goals not only at the level of external motivations (in order to avoid external pain) but also at various levels of abstraction of the goals the machine developed in the learning process. The goal-creation system is responsible for assessing activities in relation to the goals set, stimulating the learning of useful associations, and building the representations of ideas for sensory stimuli and representations of motor actions at the output of the system.

The motivated emotional machine uses internal reinforcement signals to more efficiently learn the goals of action. Because internal rewards are dependent on achieving goals set internally by the machine, learning may proceed without the participation of a teacher. Once the machine has learned how to achieve a lower-level goal, it develops the need to find in the environment objects or situations necessary to carry out beneficial activities (those that facilitated achieving this goal in the past), and this need is used to determine the motivations and goals of a higher level. In this way, the agent uses an integrated system of motivations and goals, derived from the original motives (pains) and external rewards, to create internal mechanisms for the selection and evaluation of its activities.

Motivated, emotional, embodied agent working in a hostile environment learns actions motivated by the pursuit of pain minimization, leading to the creation of abstract goals and learning behaviors that reduce this pain. Let us list the distinguishing features of a machine in motivated learning mode:

- The machine uses clearly defined primary pain signals.
- The machine is rewarded for minimizing the primary pain signals. (This defines primary goals.)
- The machine creates abstract motivations and sets abstract goals based on primary pain signals.
- The machine receives internal rewards for satisfying its goals (both primary and abstract).
- The motivation of the machine is to minimize all pain signals.

Motivated learning requires a mechanism for creating abstract motivations and related goals, as well as a mechanism for selecting objectives and supervising their implementation. The machine is constantly creating new motivations and reacting to motivations that were previously developed. Competing signals representing various types of pains (both primary and abstract) guide the machine in choosing the objective of action and acting in accordance with the designated goal. These signals change as the effects of machine actions and changes caused by these actions appear in the environment. In the search for new solutions, the machine can use reinforcement learning and artificial curiosity, using their corresponding properties and methods of implementation.

The mechanism of building motivations and choosing goals supports perception, initiating the learning of new concepts and distinguishing objects and activities that were useful in achieving the goal; this helps in building internal representations of used objects. In addition, this mechanism establishes a connection between sensory perception and proper activities on observed objects needed to achieve the goal. This learning defines new categories of objects and concepts useful for effective machine operation. Objects the machine encountered in the environment are no longer inert. They evoke emotional associations and lead to the emergence of higher-order needs and feelings. The machine learns better and faster the category of useful objects versus other objects discovered by curiosity or frequent observations of these objects in the environment. The machine will remember the objects and situations that have had a significant impact on achieving its goals or changing its level of pain. The effective use of this type of learning concept requires an episodic memory mechanism. A detailed discussion of the episodic memory model will be presented in another section.

The motivated-learning mechanism produces different centers of abstract pain responsible for the assessment of pain and learning on the basis of changing pain signals. Pain signals compete for the machine’s attention, and the winning signal motivates the machine to operate. The machine's aim at any time is to reduce the dominant pain signal. At the same time, the machine has no specific motivations for action when there is no such pain (or when all the pains are below the set threshold). It can then use the drive of curiosity to explore the environment. We write about how this can be practically done next.

%%

Now we have to take on a challenging task: designing a system that will allow a machine to feel pain and to develop the right response to reduce that pain as much as possible. We also set ourselves an even more ambitious task: using pain signals of different levels to set goals. At first glance this seems absurd. What are the aches and pains that are to be set by the system? Yet, as we shall see, pain signals are perfectly suitable for this. The primary pain signals reach us from the outside, because there, outside the casing of the system, they arise. In this delivery of pain signals, the external environment manifests its unfavorable nature to the agent, as we discussed in previous sections. This does not mean that the environment must actively inflict pain on the agent. This happens only in some situations when the agent is attacked by other agents or hurts itself. Other pain signals may be an expression of the agent's perception of the threat to his needs—for example, energy supplies are shrinking, or someone has broken into his home and stolen a valuable painting, or there is a resource shortage (and the world is such that there is almost always a shortage of resources, whether these resources are space, time, or sexual partners).

In complex environments, there are rules that regulate the relationships between various objects affecting the perception of the machine and especially its pain signals. By discovering these principles and learning how to use them to its advantage, the machine gains extensive and complex knowledge about the environment. The motivated emotional systems formulate their goals in accordance with their desire to create conditions in an environment where implementation of such goals is possible. In this way, the intelligent machines learn to actively change their environment to their advantage and not just to respond to the existing state of the environment. The organization of creating machine goals presented in this section is a sketch of one of the possible approaches. Because the description is a bit more technical, the reader who is not interested in detailed aspects of the organization of architecture or the implementation of algorithmic solutions of the machine can ignore this discussion without greatly detracting from understanding the ideas of machine consciousness discussed in our book.

The development of the needs and skills of a motivated machine in a specific environment is based on a simple built-in mechanism of creating and perceiving pain. Primary pain (a negative signal that can be considered the opposite of the reward) comes from a hostile environment and forces the machine to respond (i.e., the machine's actions are forced by a pain signal). Primary pain leads to the original purpose and the need to fulfill it through appropriate actions. This in turn stimulates the development of higher centers of pain signals and creates a higher level of motivation. The motivational learning mechanism uses basic pain and learning circuits, as shown in figure 1. In this figure, stimuli of sensory and motor neurons are represented for simplicity by individual artificial neurons (S and M), although distributed representations of perceived objects or motor activities are more effective and can be used in this method.

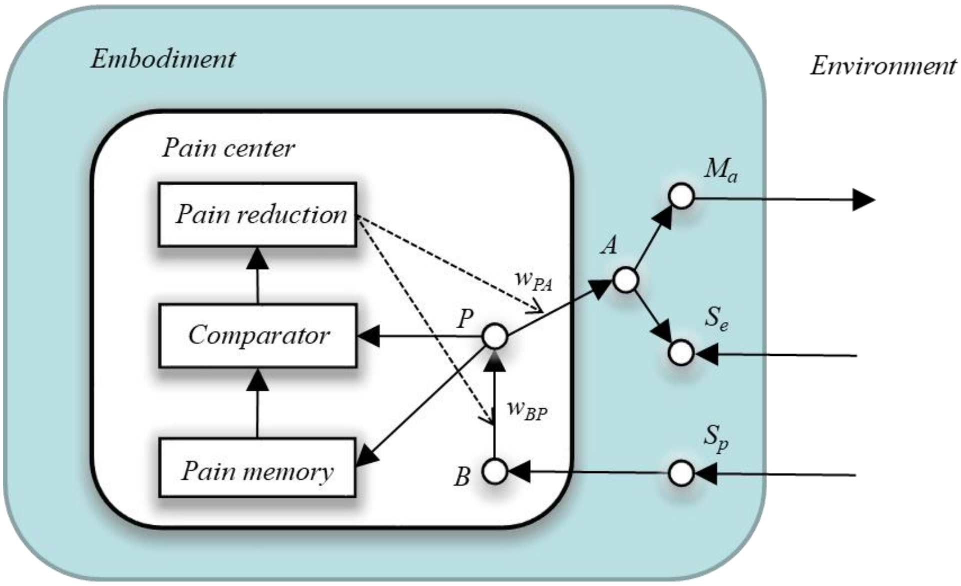

Figure 1. Basic diagram of pain detection and learning unit. The pain signal P (pain) is generated in the pain center block as a result of observing the current state of the environment, represented here by the sensory signal S_p. The pain signal is compared to what is stored in the memory value of this pain before taking the action. After the detection of pain reduction, the weights of synaptic connections in w_{BP} and w_{PA} increase, leading to the memorization of the proper action needed to reduce this pain. The action, represented symbolically by the M_a motor functions, is possible if the necessary objects represented by the sensory signal S_e can be observed in the environment.

The pain detection center responds to the pain input signal representing the negative stimuli that the machine must minimize. If the pain is caused, for example, by the lack of resources that the machine needs, then the action that will lead to finding such resources in the environment will reduce this pain signal. In figure 1 the level of real pain P (pain) is controlled by the preference signal B (associated with the level of the resource bias) multiplied by the weight of the connection w_{BP}. The pain memory center also stores the previous level of pain that occurred before taking the action. The new pain signal is compared to its previous value in the second group of artificial neurons responsible for learning control.

The growing pain signals force the machine to search for motor activities by stimulating the action of the actuators, recorded in the process of learning the system through changes (initially random) of the weight of the connection (represented in the figure by w_{PA}). The machine is looking for the right action, starting with the strongest activation of action neurons (the strongest weight combined with the pain signal). All neurons of action activation and pain neurons compete with each other using the mechanism Winner Takes All.

The victorious signal of pain forces the machine to explore the environment to reduce this pain. The solution can be found through its own search or observation as another agent performs the desired task (Rizzolatti 1996). In this way, the machine detects observed sensory relationships between objects and the response of the environment to the activities it performs. Observed objects, concepts, and activities are not predefined or learned but are observed in the environment and recorded in memory as results of a successful operation. Therefore, the concept of an object is related to the useful and predictable features that the object has in relation to the objectives of the machine and its ability to achieve these objectives through proper action.

In order for the machine to associate its observations with the features of an object, it must be able to recognize such an object in the natural environment. The machine can observe the environment using cameras and properly process the image from the camera to distinguish individual objects and their mutual relations. This is one of the tasks belonging to the field of machine vision. In machine learning, deep artificial neural networks called convolutional networks have been successfully used to analyze visual images. The convolutional neural network consists of the input and output layers, as well as many hidden layers. Hidden layers are convolutional, sparsely or fully connected. Convolutional layers apply the convolution operation to the input data, passing the result to the next layer, while the connecting layers connect the outputs of the neural groups in one layer to the stimulation of one neuron in the next layer. In multilayered convolutional networks, objects are recognized mainly by correlations and self-organization of similar features, while the function of building invariants of representation in the machine's brain (recognition of objects and ideas) is accomplished through continuous observations and correlations over time. Reliable perception and building unchangeable representations are the subjects of active research and are related to the concept of symbol grounding.

An important element of motivated learning is the mutual support loop, which consists of the fact that building the representation of objects in the environment (which results from the association of the observed activity with the reward for satisfying the needs) is the result of motivating the machine to action, while the motivation to action comes from the properties of the constructed representations, performed activities, and perceived effects of actions. New representations of concepts, objects, and problems to be solved may bring new motivations to protect or acquire desired resources, while new motivations force the machine to discover new ways of solving problems, learning new concepts, and recognizing objects that it should use. Internal motivations are a form of mental scaffolding on which the agent builds his understanding of the world and develops motor skills, autonomy, adaptability, and social skills.

As soon as the machine discovers a proper action, any difficulties that prevent this action from occurring in the future (e.g., a lack of resources or the inability to act) will cause an abstract pain. For example, if a machine needs certain resources to lower the original pain, and the resource is not available, it generates a signal of abstract pain. This abstract pain motivates the machine to look for a way to obtain the missing resource. The center of abstract pain uses a similar schema to create a motivation, like the one shown in figure 1. However, the center of abstract pain is not stimulated from the physical pain sensor; abstract pain only symbolizes the fact that there are not enough resources for the machine to satisfy its original or other abstract pain.

At any given time, the machine may have a number of different pains, each creating a different goal. Changing pains change the machine's motivations into actions as it concentrates its efforts on reducing the dominant pain. The same mechanism that stimulated action and learning at a lower level of pain would regulate learning how to respond to abstract higher-level pain. This will result in the emergence of a complex system of needs, values, and concepts regarding the observed environment. In addition, this motivating mechanism stimulates the machine to interact with the environment and develop its skills. For example, an agent may feel the initial pain when it is hungry. When food is available and the agent eats, the initial pain subsides. After observing that eating food has reduced hunger, a center of abstract pain is automatically created that becomes active in the absence of food. An inhibitory synaptic effect is also generated between the signal representing the presence of food and the center of abstract pain, and finding food in the environment suppresses this pain. When food is not available, the agent takes action to reduce the abstract pain of food shortage. Therefore, the agent may feel abstract pain (lack of food) without feeling the lower-level pain (hunger).

Motivated by the dominant abstract pain, the agent is forced to look for a method to reduce it. Finally, reducing the abstract pain of food shortage may result from the action open in conjunction with the observation of a refrigerator, in which it can find food. This means that the abstract pain caused by the lack of food will be associated with the sensory-motor pair fridge-open. In case the machine opens the refrigerator and sees food, the abstract lack-of-food pain is suppressed. Strengthening the importance of the connection between the abstract pain (lack of food) and the goal that alleviates this pain (fridge-open) after confirming that food is in the refrigerator will strengthen the usefulness of the action to achieve the goal of reducing this abstract pain. In addition, the expectation of food after the action of opening the refrigerator is created. It is like creating an imaginary pattern in the natural brain. This expectation of food will be used for planning future activities in which the expected result is known (in this case, the appearance of food after opening the refrigerator). This process is illustrated by figure 3.

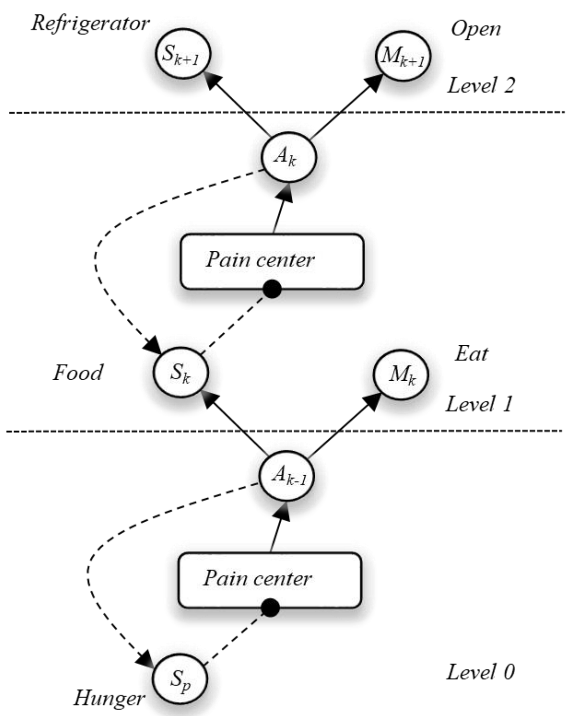

Figure 3. Creating an abstract pain signal. Primary pain Sp felt by the senses stimulates the pain center, which controls Ak-1 to ease this pain. Successful action causes abstract pain at level 1. Abstract pain can be alleviated after carrying out the correct action (opening the fridge). After this action, the higher-level abstract pain is created (when, for example, there is no food in the fridge).

This abstract pain and associated hierarchy of goals can be extended. If the agent opens the refrigerator but does not find food there, the machine must try other options to reduce this pain. It can further explore the environment or take advantage of the teacher's help. When the machine spends money (in the store) to buy food, food becomes available and reduces the abstract first-level pain (lack of food). Such an action is rewarded by the internal reward signal, which depends on the effectiveness of the action leading to the reduction of pain. In the future, this action will be strongly stimulated by the center of abstract pain (no food in the fridge). However, when there is no money, the center of abstract pain at level III is activated, creating a connection that inhibits pain after detecting the presence of money. Then the device has to learn how to quench the abstract pain at level III associated with lack of money and so forth.

In motivated learning, at every step, the machine learns an action that fulfills its purpose, and this action and associated representations of objects can cause the creation of further motivations and abstract goals. Using this simple mechanism, the machine not only learns to match the objectives to the intended activities but also learns the expected results of actions, ways to recognize and acquire the objects it needs, and the relationships between objects. It also learns the importance of objects that are related to its needs. The device regulates activities to achieve its goals and determines their priorities at any time.

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

Attention Switching

By building a cognitive system, we hope to create a conscious agent that can choose which scenes to remember and how important (from the point of view its needs and goals) are the events it observed. An agent that could focus its attention on selected elements of the scene could switch this attention depending on the state of the environment and its own needs; it could have impressions and emotions and would learn its behaviors in the natural environment. Because this would create an agent more complex than a motivated agent, we will call it a conscious agent. The critical function of the conscious agents' mind that distinguishes them from the motivated agents is their ability to switch attention; thus, the mechanism of switching attention is an important element of such functionality. By using this mechanism, agents can concentrate their attention on a single object or idea, consider a single operation, recall a single episode from episodic memory, or plan to satisfy their particular needs.

The mechanism of switching attention is an important feature of working memory that focuses on the selected object in semantic memory. The attention switching responds to various types of signals, from sensory stimuli through planning and monitoring of performed activities to associative activation of semantic and episodic memory. It results from constant rivalry between these signals for attention. However, sensory arousals that describe a significant change in input signals (movement, high intensity, bright color, pain) will take precedence over other signals so the system can pay attention to likely important events in the environment. Such a significant change in sensory arousal inhibits other signals from dominating attention switching. The next priority is assigned to signals monitoring actions performed, using cognitive support from the semantic memory. Activity-monitoring signals inhibit the effect of associative activations in semantic memory. This interdependence of the three types of competing signals is shown in figure 9.

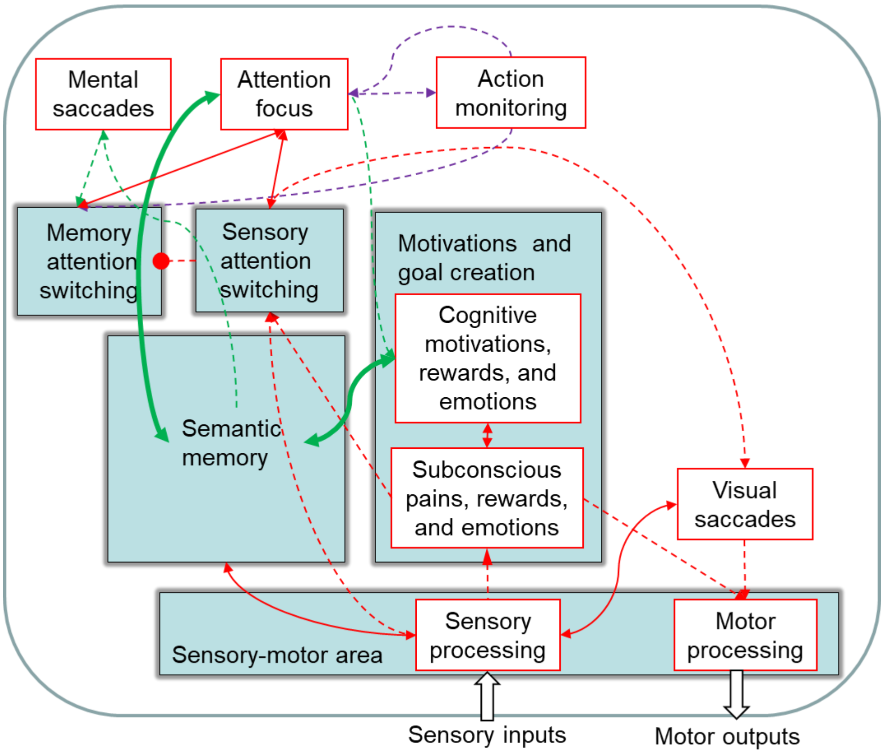

Figure 9. Competing signals in switching attention. In the conscious switching of attention, there are signals of planning and monitoring actions, associative activation of semantic and episodic memory areas, and motivation signals of actions. Subconscious switching of the action is controlled by sensory stimulation signals and subconscious pain signals representing the agent's needs.

Let us discuss how these attention switching signals impact the effectiveness of a conscious agent's actions. Subconscious sensory signals must take precedence over internal signals, at least until they are recognized, because they may signal a threat or pain that needs attention and an immediate response. On the other hand, frequent interruptions resulting from repetitive sensory signals can be irritating and can limit the effectiveness of the agent's thinking, so after the sensory signal is properly evaluated, it should be blocked by raising the threshold of sensitivity to this signal. As an example, one blocks noise coming from outside the window when trying to focus on his or her work.

While there is a need for an external signal to switch our attention because it can mean a serious threat (e.g., a fire alarm), after evaluating it, we should be able to ignore it so it does not constantly call for our attention. Here we can multiply examples confirming the usefulness of both switching attention to respond to a threat from the environment and blocking such signals when they are repeated and, in the agent's opinion, do not inform it about the threat. In the first situation (e.g., when an agent is attacked by another agent or detects that the level of radioactive radiation measured by its sensors may harm it), then the agent must interrupt its intended actions and respond to the threat. The agent can also benefit from observing what is happening in the environment. For example, an agent might have unexpectedly encountered a hidden treasure. Although the agent did not plan to look for treasure, it analyzes this finding and determines that it could use the treasure to meet its needs. As we have shown, such opportunistic behavior, when the agent uses the observed changes in the environment to meet its needs, translates into better satisfaction of all its needs than behavior that first tries to fulfill its most urgent needs (Graham 2015a).

Conscious signals representing internal motivations and action monitoring must have a higher priority in switching attention than those unrelated to a specific goal of action, thought, or association. After all, they are responsible for the proper functioning and implementation of the agent's needs. One can raise thresholds for these signals if they are repeated too often for a similar reason as blocking subconscious sensory signals. As an example, we can provide the need to block conscious information about hunger in order to focus on how to best satisfy such hunger.

Thoughts and associations compete with each other using the mechanism of mental saccades, so they do not require thresholds that block their influence on the attention. In this mechanism partially excited neurons compete with each other by their excitation level, resulting in a winner. This leads to focusing attention on the neurons representing the object, activity, or concept that is dominant at the moment, which will be consciously recognized and evaluated in working memory in terms of their suitability to satisfy the agent's needs. Then the inhibitory signal that is part of the mechanism of the mental saccades blocks the previously winning thought after it is evaluated, thanks to which other associations can take over the machine's attention and become consciously realized and processed. These conscious signals represent observations, fragments of the scene recreated in memory, plans, personal memories, and associations in semantic memory. When there are few new happenings in the environment and its primary needs are satisfied, the conscious system's attention is switched by internal thoughts, action plans, situation assessments, memories, and so forth. Mental saccades are therefore responsible for the continuous flow of the conscious agent's thought.

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

Episodic Memory and Learning

Episodic memory plays an important role in the cognitive system and must fulfill a number of useful functions. In particular, it is responsible for the learning process that leads to acquiring and storing knowledge, forming emerging concepts, tuning the representation of concepts in semantic memory, grouping time memories, building relationships between concepts and events, realizing events, finding previous episodes, quickly memorizing new ones in real time, removing less useful episodes, and finding a way. Some of these functions have been implemented in a previously developed model (Wang 2012), but some of them still require development and implementation. According to the theory of motivated learning, the agents learn by creating goals and satisfying their needs. Internal-needs signals determine the level of motivation, and meeting the needs determines the amount of internal reward that the agents receive. The agents learn how to meet their needs and are able to predict the results of their actions. When planning their activities, the cognitive agents estimate the effort and potential reward or pain that they may incur. Each planned activity or observed situation can be assigned a value from the number of rewards/penalties that are either expected or obtained by the agent.

This value can be used to determine the meaning of objects or the entire scene observed. The memory may contain sparse coding of terms using the hierarchical structure of features for each object or concept in synaptic memory. An agent can use a small number of features to represent an object, its properties, associated memories, or episodes. There may be a separate person-identification structure, another to associate this person with his home, spouse, work, and so forth. These characteristics may include episodic memories of important events, with less important background details. For example, if we meet a friend in a restaurant, we can remember several things about this meeting. We may recall dim lighting in a restaurant with many crystal chandeliers, or an obtrusive waiter who often interrupted our conversation with questions. But we may not remember how the person who sat at the next table was dressed. We might also remember the general topic of conversation but not detailed sentences or words. Thus, concepts and events that are significant or often observed will be more easily recognized than those that have little meaning.

If two concepts are associated with each other, some neurons that represent one concept can also be used to encode the characteristics of the second concept. These neurons provide associations between the two concepts, and their use may result in the mechanism of associating episodic and semantic memory (Horzyk 2017), laying foundations for mental saccades. In addition to the quick association of concepts, such sparse representations of objects, scenes, and episodes greatly increase the memory capacity.

The ability to plan and think distinguishes the cognitive agent from the reactive motivated agent. While the reactive agents react directly to the input signals from the environment, trying to minimize the pain signals, the cognitive agents use their associated memories (semantic and episodic) to plan the best activities that meet their needs. Suppose, for example, that a hungry agent sees coconuts on a coconut palm and does not know how to climb a tree. If, going further, the agent sees a ladder, the cognitive agent will be able to recall the coconut tree using its episodic memory, prepare a plan to get the ladder, and then will return to the tree with the ladder and climb it to pick coconuts. The reactive agent will not be able to do this because it has no episodic memory that interacts with the semantic memory and its needs and goals, so the agent will not remember previously seeing a coconut tree. According to psychology, the agent will not have an appropriate imaginative scheme to use for planning actions that it did not previously perform. This justifies the need to build episodic memory for the cognitive agent to effectively plan its goals.

When organizing episodic memory, it is necessary to record each episode and dynamically forget less important ones. The scene significance determines duration of the episode in memory, as well as the ease of associative recall when fragments of the scene are observed or recalled. As a result, more significant elements of the scene will have stronger connections with the concepts stored in semantic memory. Because episodic memory must operate all the time, an effective real-time mechanism must be designed to remove minor events from episodic memory. Episodes will be invoked through associative links from semantic memory. When a scene similar to that recorded in episodic memory is observed, or when elements of such a scene are activated by association with observed objects or actions, the whole episode will be recalled. The significance of the recalled episodes will be reevaluated, and their duration and ease of recall will be changed accordingly. If, in real time, the number of collected episodes is too large and the system's memory is filled with a lot of episodes that the system has not been able to assess and organize, the system should be temporarily disconnected from active decision-making so that it can organize its episodic memory. We suspect that one of the tasks of sleep is to organize the memory in terms of the significance of the memorized episodes, to remove unimportant episodes, and to simplify recorded knowledge by creating links associating knowledge with significant episodes.

Usually, frequent recall of an episode increases its importance and extends the duration of an episode in memory. Events that were not repeated and were not significant to begin with will be quickly forgotten, while those of high importance or frequency will be stored for a long time. Episodic memory has many other functions. For example, it supports the process of learning the sequence of spatiotemporal events needed to store knowledge in semantic memory or to control motor functions through procedural memory. In addition, episodic memory can be used to assess the values of observed scenes by analogy with previous experiences and rewards received and by forecasting the expected results after the planned action.

Knowledge is gained through conscious activation of semantic memory in the process of recall (perhaps as a response to the activation of episodic memory or a thought process controlled by working memory) or repeated observations. Knowledge is stored in weights of connections between neurons in semantic memory. These weights combine concepts with their defining features and combine these features with sensory activation. Thus, knowledge is represented by a complex heterohierarchy of interrelationships between cognitive concepts, their representations, and their features, as well as their use and potential impact on system values. The cognitive aspect of this knowledge is based on personal experiences. For example, the importance of money determines its usefulness for an agent. The concept of apples will be determined by taste and smell in addition to appearance and texture, as well as a way to satisfy hunger.

This complex representation of concepts and their relationships with sensory perception, motor activities, and the needs and internal pains of the agent constitute qualia—individual cases of subjective, conscious experience. Philosophers talk about qualia as if they were completely biological phenomena, creating a fundamental problem for the materialistic explanations of the mind-body problem. But individual conscious experiences are exactly that—individual—and depend on the whole system of perception and learning of the action, as well as the system of needs, values, and associations. In addition, the perception of the same objects will be different among agents that may have different sensory systems. Even among people there will be significant differences in the perception of various objects or their features, despite almost identical systems of perception and similar biological and social needs.

Episodic memory supports acquiring knowledge by recalling previously stored scenes. The scene is first created by activating semantic memory areas that represent concepts (perceived or imagined) in the center of attention. Usually observed objects are consciously recognized along with their relative location in the observed scene and are associated with the activities performed. The activation of concepts in semantic memory can be the result of, for example, visual saccades and switching attention to the observed elements of the scene. Figure 2 shows an example of an input image and its reconstructed symbolic scene represented by the activation of concepts in semantic memory.

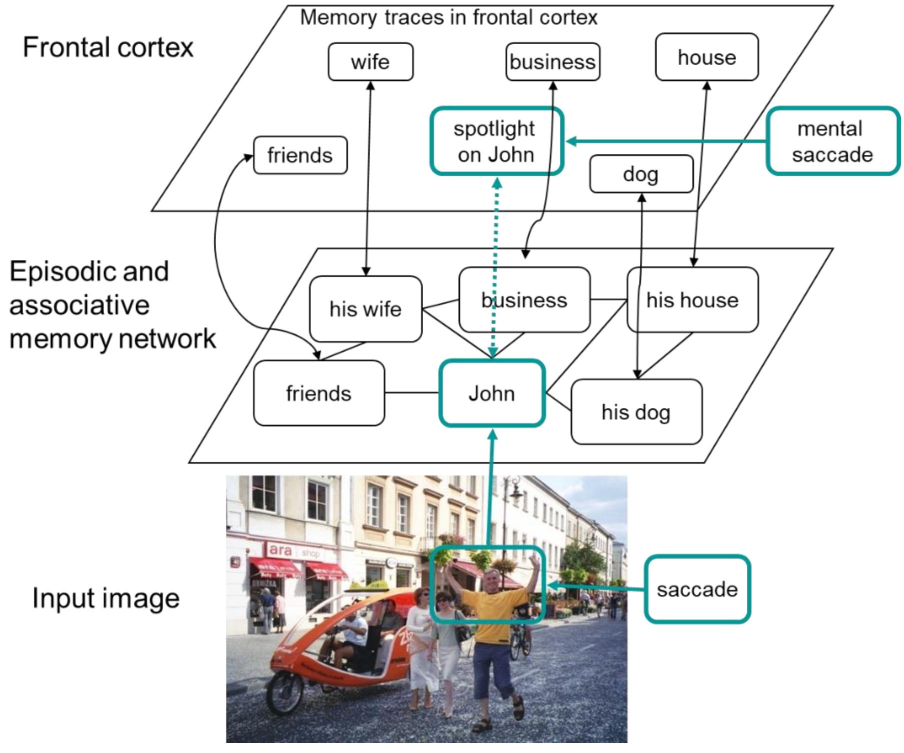

Figure 2. Reconstruction of an input image in working memory, associations in long-term memory, and focusing on working memory.

Figure 2 illustrates the process of creating associations in semantic memory. First, due to the saccade, the agent's attention is focused on the joyful man in the yellow T-shirt. After recognizing the person (John), the observer can stimulate his representation in semantic associative memory. In this memory we stimulate objects, things, people, and so forth associated with John (home, business, wife, friends, dog). Mental saccades allow the agent to focus on these objects related to John. Thinking about John's wife, the agent can remember her name and the fact that in two days it will be her birthday. This may lead the agent to start a conversation with John and send birthday wishes to his spouse. The thought of John's dog may evoke a funny episode related to how the agent got scared of a cat who jumped on him. These associations follow each other quickly, creating an uninterrupted stream of thoughts related to the observed scene.

Cognitive Architecture

The cognitive conscious system must have an imagination in order to understand the surroundings. It must also be able to imagine a hypothetical situation and plan its activities. Only the cognitive conscious being has the ability to remember episodes from its experiences. Because episodic memories require the structures of the hippocampus or its equivalent, if the body has a hippocampus, it is potentially cognitive and conscious. For simplicity we will refer to conscious, intelligent beings as conscious. This potential consciousness indicates that an organism capable of possessing states of consciousness does not have to be conscious all the time. The fact that episodes are stored in the hippocampus does not mean that the hippocampus contains all the details of the episodes, representations of objects that were observed in them, or detailed descriptions of the activities performed. This knowledge is found in large areas of associative semantic memory and in areas of knowledge about activities. The hippocampus is also considered a part of the declarative memory.

Although having a hippocampus seems to be a sufficient condition for the existence of cognitive consciousness in the body that has such a brain, this is not a prerequisite. The body can be intelligent and conscious even without the hippocampus, as is well known in neuroscience. The most famous case was the patient HM, whose hippocampus was damaged after the surgery, and as a result, he lost the ability to create new episodic memories, although he kept the memories of events that took place long before the procedure. Thousands of neurological experiments on synaptic connections in the hippocampus confirmed its importance in creating episodic memory. Part of the hippocampus is involved in the detection of new events, places, and activities.

What happens to sensory stimuli before they reach the level of cognitive perception? We do many activities on autopilot, without any cognitive supervision. How do these subconscious stimuli affect our understanding and decision-making processes? Subconscious processing was used as a tool for differentiation of conscious and subconscious perception, and it was helpful in answering these questions. According to the results of modern psychological experiments, many complex operations are performed by our brains without consciousness. Even the routine association of sensory data (including multimodal information) can occur without consciousness. Nonroutine association, which requires the creation of new combinations of sensory inputs, needs a conscious state of mind to explain and understand perceived inputs. As a result of subconscious processing, what we perceive as a conscious image is highly processed information and is an expert reconstruction of the external world, which differs from the data actually appearing in front of our senses. The working memory plays a significant role in the process of understanding the observed scenes and planning, thinking, and making decisions.

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

Working memory is a part of the cognitive system, responsible for the short-term storage of information available for processing (Miyake 1999). It records stimuli reaching the senses processed in the deeper layers of the brain. A scene image or other sensory impression can be registered and remain in working memory as long as the electrical states of the neurons involved in the coding process persist. The same image can be recalled or reconstructed from the state of electrical action potentials. This phenomenon of intentional remembering has been confirmed experimentally (Gelbard-Sagiv 2008). During the remembering, the same set of neurons that was involved during observation of the object is stimulated. It is believed that this mechanism is responsible for the functioning of short-term memory.

The short-term memory is important for reasoning and setting guidelines for planning, decision-making, and behavioral control (Diamond 2013). Anders Ericsson and Walter Kintsch (Ericsson 1995) introduced the concept of long-term working memory, which they define as a set of reading structures in a long-term memory, allowing easy access to information relevant to everyday tasks. In this way, part of the long-term memory functions as a working memory. Cowan also does not treat working memory as a structurally separate system in long-term memory (Cowan 1995). Representations in working memory are a subset of representations in long-term memory.

The working memory is divided into two levels. The first consists of activated representations in the long-term memory. The second level is the area of focus. Attention in working memory has limited performance and can hold up to four active representations simultaneously (Cowan 1995). There is some evidence that optimal performance of working memory is combined with the ability to focus attention on relevant information and ignore disturbances (Zanto 2009). Research also suggests a link between the efficiency of people's working memory and their ability to control the concentration of attention to environmental stimuli (Fukuda 2009). Such control allows people to pay attention to information that is important to their current goals and to ignore stimuli that are not relevant to these goals and that tend to shift attention due to their sensory significance (e.g., an ambulance signal). Working memory, with its attention mechanism, serves to dynamically bind all neurons involved in conscious recognition. Symbol grounding works on several levels of hierarchical memory structures, and the connections at the bottom of these structures can be used to build structural representations of objects to be stored in long-term memory and recognized in working memory. Bindings (couplings) of lower-level neurons combine features of elements or parts of them into objects, concepts, ideas, events, actions, and so forth. For example, objects in visual working memory can be represented by a set of different features such as shape, color, size, and structure, and each object is a combination of these features. Complex objects can be represented as spatial assemblies of their parts, their relative positions, orientations, and sizes.

The couplings at the lower level are crucial to remember which features must be combined and recognized as a single object (Luck 1997). At a higher level, the elements that make up complex objects or concepts are related to the representation of the context—for example, time or location. Temporal and episodic memories require sequential linkages of elements to locate them in the time or order in which they are observed (Nguyen 2012). Hawes and others developed the architectural integration mechanism to link cognitive architecture functions like action planning, language, and visual properties such as color, shape, or position (Hawes 2007) and used it in the development of schemas' cognitive architecture. Oberauer and Lange presented four hypotheses regarding the grounding of symbols and long-term memory (Oberauer 2009): 1) activation and binding are two basic functions of working memory, reflected in two different processes to identify and assess the level of familiarity and reminding; 2) the limit of the working memory capacity concerns connections and their maintenance, but it does not result from the activation of representation in the long-term memory, and therefore the ability of working memory to recall something is limited, but the assessment of familiarity is not; 3) knowledge does not depend on feedback, while reminding does; 4) temporary associations (connections) in working memory are saved as associations in long-term memory through gradual learning.

The cognitive process is based on representations of concepts (objects and relations between them) stored in a declarative long-term memory and procedures (actions and sequences of tasks binding various objects) stored in the procedural memory. Attention switching and control testing studies have attempted to explain how people prepare to perform a task or switch between tasks. After selecting the task and planning the operation, the procedural representation of this task is stored in working memory. Working memory is responsible for temporarily storing information that has been perceived in the environment or has just been retrieved from long-term memory. Thus, memory keeps activating the few elements that we want to manipulate or with which we want to plan the action, so the information stored in working memory is there for a specific purpose. In addition, working memory combines temporary storage and manipulation of selected information to support cognitive functions. It has three distinct states: activating a subset of long-term memory, limiting area of direct access, and focusing attention (Oberauer 2002). The role of focusing is not to keep the set of memorized objects active in the working memory but to provide conscious access to a single object selected for manipulation.

The elements in attention focus are selected for the planned operation from the consciously activated part of the long-term memory. A subset of active elements in the long-term memory creates a direct access region (selection set); new elements are selected from this set for cognitive processing one by one. The content of the active part of the long-term memory is the support of the elements of direct access. One can choose them indirectly by associating with perceptions, hints, thoughts, and motivations. Excited neurons in the area of direct access can be selected depending on the degree of their activation resulting from the context of stimulation of neurons on which attention is focused. The mechanism of the mental saccades described in the section on attention switching serves this purpose.

Short-term memory, closely related to the working memory, stores only the information manipulated by the working memory. The selection of a subset of this information is critical to the focus of attention and cognitive operations. Short-term memory refers to recently experienced conscious events, such as what just happened, where I was, what I did, where I put my glasses, and so forth. In short-term memory, one can store practically everything that was the subject of attention. Short-term memory is therefore an indispensable element of many functions fulfilled by working memory and is the basis of consciousness.

Short-term memory contains a small amount of information active in the brain for a short time. It lasts a few seconds, but this time can be extended by active sustaining and repeating. The capacity of short-term memory depends on the complexity of stored concepts and their universality and spans from four to seven unrelated elements or pieces of information. It should be distinguished from working memory, which not only stores information but also uses it to organize a process of conscious comparison and forecasting, reasoning, and planning activities. Working memory uses focus to select and process temporal information.

To store information for a long time, information in short-term memory must be periodically repeated or practiced. Grouping can help to overcome the small capacity of short-term memory by organizing objects in easily recognizable groups. For example, a long sequence of letters can be grouped into syllables or known acronyms. The name "short-term memory" means that activated memory traces disappear spontaneously and are forgotten over time. As we can observe in everyday life, new information, if not used, is quickly forgotten. Similarly, artificial short-term memory models assume automatic disappearance of neuron activation over time.

Working memory relies on short-term memory to perform its functions and is often confused with it. Cowan stressed that working memory correlates better with intellectual abilities than short-term memory because it also includes information processing supporting cognitive functions (Cowan 2008). In contrast, short-term memory is only a temporarily activated subset of long-term memory. It has been proven that long-term memory supports short-term memory, which can only handle several elements at once. Loss of continuous activity in short-term memory is caused by lack of attention and not by memory disturbance, as noted by Jarrod et al. (Lewis 2012). The focus of attention seems to differentiate these two forms of memory more than anything else. Irrelevant activations in short-term memory quickly disappear from the focus of attention, but they can be brought back to the center of attention as soon as they are considered important—for example, when their participation is needed for the purpose of an action chosen by the working memory. The assessment of the significance of short-term memory elements is therefore variable; it depends on the context and the course of the thought process referring to short-term memory, the needs of the agent, and the ways of their implementation.

Oberauer proposed a functional approach to modeling the working memory. He assumed that working memory is needed to provide access to information both perceived and stored in long-term memory in order to plan and take action and achieve set goals. In human brains, this corresponds to the support of higher cognitive functions, such as speech, planning, thinking, and problem-solving. For this purpose, working memory temporarily stores and manipulates several concepts in its declarative and procedural elements. Declarative working memory keeps objects consciously manipulated, while procedural working memory refers to procedures that can be used to manipulate these objects. Hence, switching objects and switching tasks result from the interaction of two different subsystems, so increasing the cognitive load to a declarative part of working memory does not affect its procedural part. Oberauer claimed that both procedural and declarative working memories have analogous structures and operate using similar principles. According to Oberauer, both types of working memory consist of: a) activated representations in long-term memory, b) a mechanism for creating temporary bonds to activated memory elements, and c) concentration of attention on a single object. Thinking and planning activities require the manipulation of concepts and their relationships, so the feedback must be quickly established and quickly dissolved. This requires different mechanisms than building associations in long-term memory.

Because the working memory supports cognitive functions, its design must meet several requirements specified by Oberauer (Oberauer 2009). First, it must be able to build and maintain structural representations through dynamic constraints; second, it must be able to manipulate such structural representations; third, it must be able to be flexibly reconfigured; fourth, it must allow partial disconnection from the long-term memory; fifth, it must enable controlled downloading of the required information from the long-term memory; and sixth, it must be able to store newly created structural representations in long-term memory. Below, we describe how to meet these requirements.

1. To build and maintain structural representations, working memory must be able to represent new concepts, new sequences of action, and new plans. This requires creating schemas and procedures for dynamically binding elements such as objects,

words, actions, and events. These bonds must be quickly established and quickly dissolved.

2. To manipulate structural representations, the system must be able to selectively access the required concepts and subject them to cognitive activities. This requires an attention-based mechanism of selection and evaluation of performance in accordance with learned principles (e.g., capturing and transferring a single object).

3. To provide a general, flexible reconfiguration, an executive process is needed that is able to control its operations and configure parameters according to the required task. Such a process must be flexible enough to change the strategy of planning activities depending on the situation.

4. To enable partial disconnection (inhibition of associated elements of acquired knowledge) from long-term memory (both declarative and episodic) and effectively manipulate structural representation, the system must be able to separate learned knowledge from the newly planned operating scheme. For example, a chess player should be able to detach layouts of the board from the previous games to be able to create new positions and follow them a few steps ahead. In addition, to avoid routine, the system must be able to free itself from typical activities.

5. In order to obtain a controlled retrieval of relevant information from long-term memory, the system must be able to decide at which point it accepts prior knowledge, and when such knowledge must be omitted, in order to possibly create a new solution.

6. Finally, to save new structures in long-term memory, the system must be able to transfer a new, beneficial solution from working memory to long-term memory to remember it. This transfer can be implemented in a neural network through the use of context existing in working memory to create representations of new knowledge in semantic memory.

According to Oberauer and his colleagues, working memory is first of all a system of information integration and building relational representations that bind different elements of memory, not a traditionally understood system of storing, processing, and focusing attention (Oberauer 2008). He showed a direct connection between such elements of intelligence as reasoning, memory, and creativity and various functions of working memory, especially its capacity. Both the working memory capacity and the intelligence reflect the effectiveness of focusing attention and the ability to maintain goals and relevant information in the face of disturbances. It was observed that building relational representations in working memory predicts the ability to reason, as well as the quality of storage and processing, while the switching of attention and supervision were not related to intelligence and were only moderately related to the reasoning skill.

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
We will now go on to describe the architecture of artificial cognitive systems. It consists of interrelated functional blocks that process information, leading to the creation of representations of objects, assumptions, beliefs, and knowledge that the system possesses about something specific (people, groups, categories, concepts, activities, objects, etc.) or abstract (thoughts, theories, information, skills, beliefs, ideas, etc.). The cognitive system has its own individual view of the world and decides how it perceives, summarizes, filters, and organizes information about the world around it. Additionally, in accordance with the assumptions of the embodied intelligence, the cognitive system has the ability both to observe the world around it using its senses and to act on the surroundings using the actuators and

embodiment. Examples of such systems are natural systems, animals having at least the seeds of consciousness and self-consciousness, and of course people.

Currently developed artificial cognitive architectures such as SOAR (Laird 2008), ACT-R (Anderson 2004, 2007), Icarus (Langley 2005, 2006), LIDA (Baars 2009), Polyscheme (Casimatis 2002), and CLARION (Sun 2007) must either rely on predefined goals (without self-motivation) or predefined rules (without autonomous reasoning). Due to their dependence on predefined heuristic rules, these cognitive systems do not have the autonomy, adaptability, or reasoning needed to perform complex missions or to realize constantly changing goals in uncontrolled, complex, dynamically changing environments. Because we want to develop machines with human learning abilities, we must go beyond the reactive response used in reinforcement learning, introducing internal motivations and creating new, previously unknown (to designer or robot), and perhaps abstract goals. Such organized systems must be able to develop knowledge about their environment and must have their own beliefs, motivations, and goals; they must communicate, observe, plan, act, and anticipate the effects of their actions (Voss 2006). Intelligent systems adapt to unpredictable and dynamic situations in the environment through learning, which gives them a high degree of autonomy and makes them the ideal solution for controlling autonomous robots and virtual agents (Floreano 2004; Pfeifer 2007). The mechanism of motivated learning described in this book provides such development opportunities in the system of an embodied intelligent agent (Starzyk 2008).

This section presents a functional organization of the cognitive memory of embodied agents. We will start by presenting a higher-level description of functional blocks, their interactions, and ways of information processing. This description is aimed at presenting ideas, not their detailed implementations. After presenting the organizational structure of the cognitive agent, we will emphasize the description of semantic and episodic memory blocks (Horzyk 2017) and their functions.

Next, we will discuss the structural organization of cognitive agent memory, presenting functional links between various functional blocks and explaining their mutual relations. Different functional blocks process information simultaneously, sending interrupt signals to direct attention, change plans, monitor activities, and respond to external threats and opportunities. They also provide a conscious agent with personal memories, accumulated knowledge, skills, and desires, making the agent act fully autonomously, satisfying his needs, building his motivations, and influencing his emotions.

Consider the structural organization of the cognitive agent's brain, in which functional areas that interact with each other can yield conscious actions in the environment, model the environment in memory, accumulate knowledge and skills, create scenes and record them in episodic memory, create motivation and determine the objectives of actions, focus and switch attention, plan and supervise performed activities, evaluate these activities, and generate feeling in the agent's mind. The presented organization seems to meet the basic requirements for information processing by a conscious intelligent robot system. It seems to contain the minimum organization of a conscious system. This does not mean that other mechanisms that support the functional architecture cannot be added to this organization. An example here can be the mechanism of creating emotions and its influence on the way the information is processed by the brain. Because the research in organization and the role of artificial emotions is only in the initial phase (Starzyk 2016, Starzyk 2017), we do not postulate here specific structural or functional solutions for emotions.

Figure 3 shows the main functional areas of the cognitive agent. As we can see in this figure, this artificial brain is not homogeneous, but it contains specialized areas of memory whose functions are clearly defined and whose pathways of synaptic interactions with other regions of the artificial brain determine their functional separateness. It is highly likely that some specialized brain functions will have a specific connection structure between neurons, providing the desired functionality. This is what happens in the brains of animals, where, even though the neocortex has a relatively uniform structure of connections between neurons in different places of the cortex (a familiar structure of minicolumns), the organization of connections of the medulla, bridge, midbrain, interbrain, limbic system, and cerebellum are all different. The main functional areas of the cognitive agent shown in figure 3 are:

1. Sensory and motor area
2. The area of creating motivation and goals
3. Long-term associative memory area (semantic and episodic)
4. Motor control area
5. Area of conscious and subconscious switching of attention
6. Working memory area

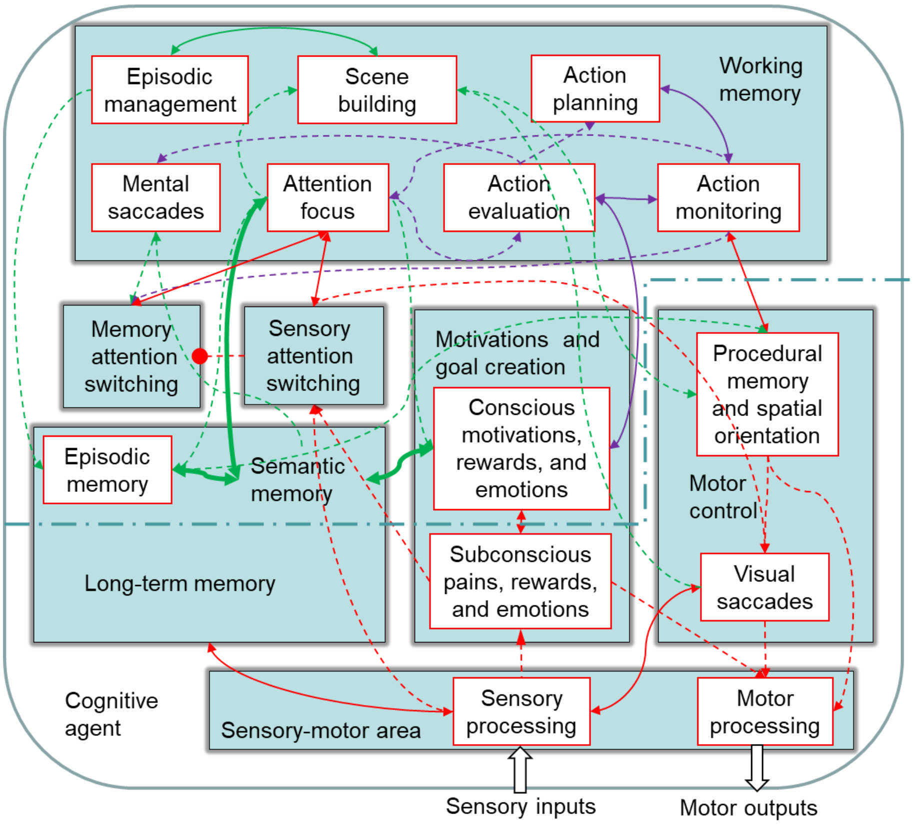

Figure 3. Main blocks of the cognitive architecture of the agent. Links between different memory areas indicate the flow of information between them. Dashed lines indicate one-way connections, with arrows indicating the direction of information flow.

The dash-dot line separates the conscious and subconscious areas. This division is fuzzy and depends on the interaction between top-down and bottom-up processes that consciously perceive objects or ideas on which cognitive functions are based. A conscious perception of an object may involve a combination of many sensory and motor elements involved in the recognition of objects. This may require further manipulation and observation of the object and the active search for characteristics that either confirm its recognition or lead to a different interpretation of the observation. Thus, during the observation of a single object, a group of activated neurons acting synchronously, representing conscious perception, may change. Let us briefly discuss these functional blocks, the flow of information, and the types of interactions between blocks.

Sensory-Motor Area

The **sensory processing** block receives information from the environment through sensory inputs, processes it, and sends the result to the long-term memory for conscious recognition. Although the recognition is the main function of the sensory block, it also stimulates the primitive pain signals, evokes emotions, or provides subconscious signals used in the sensory attention block to focus attention on a particular sensory input. The sensory processing block also receives feedback signals from the long-term memory to identify the subject of attention. Video saccades also help to focus on the scene element, but it is the concentration of attention in the semantic memory block that helps to identify the object. Object identification takes place through interactions between semantic memory and sensory processing, and this is similar to the concept of adaptive resonance theory developed by Carpenter and Grossberg (Carpenter 2003).

The **motor processing** block sends control signals to the actuators responsible for the operation of the motors. The main stimuli of neurons in this block come from the procedural memory, which controls the execution of sequences of operations and controls specific actuators in each motor action. The input from the visual saccades block is used to shift the focus of visual attention to the selected target. The input, which motor processing receives from the block of subconscious pains and emotions, causes a subconscious reaction to remove the source of pain or emotion. Sensory-motor processing takes place at the bottom of the hierarchy of memory areas.

Area of Creating Motivation and Goals

The block of **subconscious pains, rewards, and emotions** receives the stimulation from the sensory processing block and sends its outputs to motor processing. In addition, this block sends signals to the subconscious block of sensory attention switching, which draws attention to the source of pain and to the block of conscious motivations, rewards, and emotions, influencing the creation of abstract goals, conscious motivations, and emotions in accordance with the theory of motivated learning (Starzyk 2011a). It receives emotional feedback from a block of conscious motivations, responding to both painful and pleasant events, causing motor reactions expressing fear, frustration, disgust, anger, happiness, sadness, and surprise, depending on the context. Emotional reactions arise as a result of subconscious and conscious pains and rewards. Emotions have a strong impact on the cognitive process and motor activities, which affect the agent's behavior. The machine will never have human emotions due to a different morphology, but its emotions will be real and may be similar to human emotions. In figure 3, pains and emotions are combined in the same block for the convenience of description, since they are subconsciously generated and affect both the motivations of the system and automatic reflexes.

A block of **conscious motivations, rewards, and emotions** receives signals of primary pain from a block of subconscious pains, rewards, and emotions. Based on changes in these signals, it creates abstract goals and sends information about the level of pain signal to the action assessment block when such information is requested from working memory. This allows the agent to learn useful actions and establish new motivations and at the same time to recognize and avoid activities harmful to the agent. This block interacts with the emotional states of the agent, providing conscious information to regulate emotional reactions. Other feedback for conscious motivations comes from the attention block. Focusing attention on conscious motivation may result in the assessment and planning of activities that consider this motivation. The bidirectional link from this block to the semantic memory primes the semantic memory and may trigger specific motivation for actions in the course of competition with other motivations.

Long-Term Associative Memory Area

Long-term memory results from establishing permanent synaptic changes between neurons in the learning process. It differs from short-term memory, which is expressed through the activation of neurons and disappears with the disappearance of this activation. Long-term memory simultaneously describes the structure of connections between neurons. It contains various areas of memory, including the sensory-motor area, semantic and episodic memory, and procedural memory. Long-term memory contains related semantic and episodic memory due to their special role in learning, creating conscious experiences, acquiring skills, and creating motivations.

Semantic memory supports attention focusing, attention switching, creating new concepts and refining them, acquiring and storing knowledge, detecting novelties, building connections between concepts and events, creating goals, evaluating rewards, and learning, as well as anticipation and action planning. In addition, semantic memory supports the process of recognizing objects, building scenes in episodic memory, creating episodes, evaluating actions, and monitoring activities. This is one of the most extensive areas of memory. It receives input from the sensory processing block and incrementally learns invariant representations of observed objects through gradual changes of the weights of synaptic connections between neurons. Knowledge contained in semantic memory, activated by associations with goals and observations, is strengthened by focusing attention and supports all cognitive activities.

Episodic memory takes information about the observed scene from an episodic management block and saves it, together with the significance of episodes, using a long-term sequential memory mechanism. The significance of episodes or only elements of the scene has a strong influence on determining the duration and significance of each episode. Episodes that have little connection to the goals and needs of the agent are quickly forgotten. This differentiation protects episodic memory from storing a large amount of information and facilitates access to saved information through associations. The connection from the attention focus block to episodic memory is used to recall episodes, using associations with an object, event, or need that are in the spotlight.

A two-way link to semantic memory is used for recalling details of past episodes and learning in semantic memory. Because semantic memory uses an incremental learning, repetitions are needed to make certain observations permanent in this memory. Past events can be repeatedly reproduced from memory without the need to reobserve these events. Thus, both types of long-term memory interact closely with each other. Semantic memory is needed to record conscious events in episodic memory, which in turn supports semantic memory in the process of incremental learning. Finally, the connection to the procedural memory is used to learn procedures based on past episodes in which a useful action has been observed.

Motor Control Area

Procedural memory is responsible for the performance of previously learned actions. Sequences of operations are stored and recalled in this block. Initially learning sequences require full concentration and conscious attention. At this initial stage, the learning of activities is supported by working and semantic memories, but if the system often performs learned activities, cognitive supervision of these activities is gradually reduced, and they become routine and subconscious. We believe that procedural memory is a subconscious memory. Procedures can be learned directly from the observation, using mirror neuron activations and a scene building block or recalling these observations from episodic memory.

The procedural memory has two-way connections to the activity monitoring block in working memory. The sequence of actions learned and stored in the procedural memory can be consciously monitored by working memory, which supervises the progress of performed activities, directing attention to the manipulated objects and deciding about continuation or abortion of these activities.

Spatial orientation is implemented as a subset of procedural memory. It uses spatial information from the scene building block and uses it to create maps of familiar places. These maps are used by the agent to move in the familiar environment to perform routine tasks. Spatial orientation is also needed to manipulate objects that the agent uses to perform its actions. Finally, two unidirectional connections exiting the procedural memory to the motor processing block and the visual saccades block provide motor control of the activities performed and activate the saccades to observe the result.

The **visual saccades** mechanism receives control signals from the procedural memory to focus visual attention on the selected object or action. This includes searching the environment in building the scene. Another input to the visual saccades is a link from the subconscious attention switching block to detect the signal source that has switched this attention. The subconscious attention switching block decides which part of the observed scene is interesting from the point of view of the agent's needs in the context of his actions in the environment. The saccade mechanism uses this information to direct the vision to a specific place in the scene, and after the agent consciously recognizes what is there, the visual saccade mechanism switches attention to a different place in the observed scene (taking into account the elements of the image, such as brightness, color, movement, etc.). The output from the visual saccades is directed to motor processing in order to make a saccade movement (movement of cameras or eyeballs) or to the sensory processing block in order to simulate such movement (if a software equivalent of the saccade movement is used). The software equivalent of the saccade movement is obtained when, instead of directing the visual attention to the new part of the observed scene by moving the camera, an input image is scanned by software to select the next focal point of the video attention. This yields a faster search of the observed scene; there is no need for physical camera movement. The link from the visual saccades to the scene building is used to determine the location of objects in the scene.

Area of Conscious and Subconscious Switching of Attention

The block of **subconscious attention switching** receives inputs from the sensory processing area and sends interrupt signals to the attention focus block. In addition, it is a source of information for visual saccades, determining the strength of the signal that controls the saccade movement on the basis of color, size, motion, and other important sensory features. This block also receives signals from the block of subconscious pains, rewards, and emotions to switch the agent's attention and focus it on the source of pain. Finally, it helps in focusing attention on the observed objects and the conscious recognition of the object. The block of subconscious attention switching sends an inhibitory signal to the conscious attention switching block to prevent focusing on planning and thinking until the subconscious signal from the environment or the inside of the body is evaluated.

The block of **conscious attention switching** responds to mental saccades in order to process thoughts, plans, and associations based on the priming of semantic and episodic memory. It receives inhibitory signals from the block of subconscious attention switching. This inhibition prevents switching attention by the process of thinking or planning actions so that the system can focus on subconscious sources of attention switching associated with significant changes in the environment or subconsciously growing pain signals. The attention focus block reacts to signals from the conscious attention switching block, focusing on the selected area of semantic memory. Another block that interacts with the conscious attention switching is the activity monitoring block. Monitoring activities can direct the conscious attention to objects, activities, or thoughts necessary for the implementation of activities.

Area of Working Memory

The area of working memory highlighted in our model is only an example of the organization of the basic functions that this type of memory provides. A detailed organization, using activated areas of long-term memory and mechanisms supporting short-term activation of stimulated neurons and their inhibition, may affect the form of mutual interactions between attention focus and the degree of maintaining this attention while maintaining several parallel stimuli, assessment of the usefulness of activities, accuracy of the assessment of the situation when making decisions, associations, and so on. Equally important for proper functioning of working memory is to consider time constraints imposed on the process of making informed decisions, the emotional state of the agent, its life experience, the state of self-awareness, social status, and many other factors.

Action planning is used for conscious planning and checking all aspects of the planned action, including its value to the system. Action planning initiates monitoring of the planned activity, checking the environmental conditions, and the possibility of implementing the planned activity. Indirectly, action planning uses the information from procedural memory to determine the next operations needed to complete the planned action. Action planning uses the functions of the action monitoring block for the mental rehearsal of the planned activities, checking the conditions necessary for their implementation and assessing their results. If the mental rehearsal is positive and the planned action is beneficial to the agent, then the action evaluation block starts a new action by initiating the actual process of monitoring the action. If the plan cannot be executed, the scheduled activity is abandoned and is removed from the action planning block.

The **action evaluation** block considers the possibility of taking action when the attention focus block concentrates on a new object or idea. The action evaluation uses a bidirectional connection to a block of conscious motivations to see if the action being evaluated is beneficial to the agent. If so, the action evaluation block activates the activity planning block to initiate the plan. After completing the action plan, the action can be initiated by the signal from the action evaluation block, which receives information on the feasibility of the action plan and the suitability of the action. In the event of a decision to start an activity, the motor control and the monitoring of the activities performed shall commence. After the end of an activity or its abandonment, this action will be removed from working memory. The action evaluation block also blocks the planned action if it is harmful or impossible due to adverse environmental conditions. If the assessed action is not favorable, the action evaluation block forces another mental saccade, and the system focuses attention on the next option.

The **action monitoring** block has a bidirectional link to the procedural memory to monitor the progress of the operations according to the procedures stored in the procedural memory. This block also uses a one-way link to the attention focus block to focus on every step of the performed procedure. A one-way link to the conscious attention switching block forces focusing of attention on the planned action. The signals from the action monitoring to the conscious attention switching block have higher weight than signals from the mental saccades but lower than signals from the subconscious attention switching. This helps the agent to concentrate on the activities performed while allowing the subconscious attention switching that can warn of significant changes in the environment. The action monitoring block also includes bidirectional connections to action planning and evaluation. The incoming activations from the planning and evaluation blocks initiate the process of monitoring activities (actual or planned), and outgoing feedback signals from the action monitoring block inform the evaluation and action planning blocks about the progress achieved.

Scene building is very important from the point of view of learning processes, memorizing episodes, active exploration, spatial orientation, and mapping. It includes conscious recognition of scene elements and their locations, assesses the significance of observed events, and assigns time references or other relevant information about the scene. It uses signals that control visual saccades to determine the location of the object, its contextual significance, and its meaning obtained by associating the selected object with conscious motivations. Scene building is regulated by an episodic management block, which creates sequences of scenes, updating the contents of the episodic memory. Scene building is a key to spatial orientation and creates a map of the observed area, sending spatial information to the spatial orientation block in the procedural memory.

Episodic management is a part of the working memory responsible for scene building and creating episodes. Episodic management saves a scene to episodic memory when it detects a significant change in the current scene (new event or new location). It also collects information about the significance of the elements of the scene. Scenes are arranged in episodes and saved to the episodic memory, together with the significance of these episodes to the agent. When the scene is saved in episodic memory, the new scene is initiated by the episodic management. Not all episodes are worth storing in episodic memory; they can be forgotten even before they are sent to episodic memory if their significance is low. Episodic management organizes and records scenes in real time.

The block of **mental saccades** is the main mechanism for changing thoughts and plans using priming and associations in semantic memory. It takes input data from all primed neurons (neurons that are partially excited) in semantic memory and evaluates them using conscious attention switching. When the winner of competition between partially excited neurons is selected, the feedback signal from the conscious attention switching block is sent to the attention focus block, activating the selected semantic memory region. Note that many signals from semantic memory (indicated by thicker lines in figure 3) compete for attention, along with interrupt signals from the subconscious attention switching block, which have priority over other signals.

The **attention focus** block selects the chosen concept in semantic memory, activating the short-term working memory, and is responsible for all cognitive aspects of the systems operation. It helps to provide a conscious interpretation of all observations, actions, plans, thoughts, and memories. The focusing block has one-way connections to conscious motivations and episodic memory and a bidirectional connection with semantic memory. Thus, all episodes or motivations in the focus of attention are interpreted semantically. The attention focus block receives signals from the activity monitoring block and has two-way links to blocks of conscious and subconscious attention switching. The direct connection from the attention focus block to the performance evaluation block helps to assess the feasibility of the planned action, and the indirect connection through the conscious motivations determines the value of the planned activity. The signal coming from the action monitoring block focuses attention on the currently performed or planned actions in order to end or stop the operation. Bidirectional links with subconscious attention switching help focus the system's attention on the source of interrupt signal. Finally, the signal sent by the bidirectional connection from the conscious attention switching block focuses attention on the target object. A change in focus may be the result of a subconscious interrupt or conscious attention switching. Conscious attention switching can come from monitoring of cognitive activities—what's next—or from competition between associated areas of semantic memory in the context of past episodes, observations, knowledge, and motivation. Competition between the associated areas of semantic memory is the weakest source of attention switching, but it is the most common one. Although in figure 3, sensory and memory sources for attention switching are separated from each other, they all compete for the attention of the agent. Therefore, the thresholds of subconscious switching signals and entry thresholds to switch attention from the activity monitoring block should be properly adjusted. Only when none of these inputs exceeds the appropriate thresholds will the winner of the competition of associations in semantic memory be in the spotlight and be considered by working memory.

The Future

We chose a provocative title for this book. In this provocation, there is an incentive for those who would like to understand what consciousness is. Our goal was to explain the phenomenon, which is perhaps even harder to understand than the emergence of life from inanimate matter. The analogy to how a boy can understand the principles of radio or television seems to be a naive simplification, and the expectation that we will explain consciousness as simply as the action of heterodyne may be an unattainable goal. First of all, a simplified description of the operation of the radio or television was possible because the experts knew how to build them and how they work—a popular description explained this to laymen. Unfortunately, there are no experts today who know how to build conscious machines. We explain how consciousness can arise in the complex brains of intelligent machines, but do we know how to build such machines? The answer to this question depends on the availability of adequate technology in which the implementation of our ideas will be possible. There is no such technology today, but the development in the field of artificial intelligence is huge, and we believe that it will bring with it the necessary technologies.

What technologies are needed to build embodied, conscious machines? First of all, their sensory arousals must be based on the observations of the results of actions in the real world; they must be associated with the feelings that these observations cause in the embodiment of an agent. This requires the development of sensorimotor coordination integrated with the machine value system. For this, a memory is needed that records system-relevant observations and is able to reproduce them in real time, predicting the response of the environment to machine operations. This requires further development of machine vision and other sensory inputs such as hearing, touch, smell, or sensations of changes in temperature, pressure, or humidity. The quantity and quality of sensory inputs will affect the representations of observed objects created in the artificial brain. Research has been carried out in this area for years, and methods of machine perception have already been well developed. But they are still not based on the agent's needs, pains, and actions, they do not produce feelings.

The second important technology is the development of learning methods, building representations, and control of the robot's movements. This includes the development of motoric functions, activators, grippers, methods of movement, and navigation. Building the representation of learned activities depends on the way the memory of motor functions is organized, represented by the sequences of vectors controlling the agent's movements, and coupling these motor functions with perceptions, with the planning system, and with the needs of the agent. Every day, the industry provides more and more effective techniques leading to autonomous machines performing complex operations in the real world, but these are just the beginnings of advanced techniques for learning motor functions in intelligent systems.

Intensive works on the construction of associating spatiotemporal memories are underway. These memories, inspired by the functions of biological neurons, have been developed for many years in the field of artificial neural networks. Their real-time operation in systems with a large number of artificial neurons is still limited by technological capabilities, both in terms of memory capacity as well as power consumption. The basic limitation is the lack of technologies enabling parallel processing of information in billions of neurons, and their simulation on classic computers requires the high computing power and energy consumption of the most powerful supercomputers. Although we do not have technology today that would allow parallel processing of information in networks containing billions of neurons, there is hope that further development of computers will solve this problem.

However, in this book, we argue that when such technologies are available, we will be able to construct and develop thinking, conscious machines. We do not say this in vain. As a result of research sponsored by the National Science Center in Poland, we built a virtual robot that operated according to the principles of motivated learning (Starzyk 2017). We showed how a virtual robot can learn how to solve problems in a virtual environment, learn behaviors beneficial to its needs, create internal goals, and achieve them in real time. Robots based on motivated learning create a system of values necessary for the emergence of consciousness. Such a system is the basis of the processes of recognition, planning, thinking, associating, predicting, learning, and purposeful action. The value system is also an important element of the mental saccades and focusing of attention. Based on associative memory, the coordination and predictions of effects of motor activations with the perception and signals of pain are the basis for creating feelings. Creating feelings is the foundation of conscious experiences, emotions, understanding of the world, and the emergence of self-consciousness. The resulting systems will have motivated emotional minds (MEMs). Emotions, cooperation, and understanding of the intentions of other individuals are the basis for communication, actions for the public good, creation, and observance of moral principles and norms.

It is obvious that further development of the concepts described here will be an ongoing process of improving technologies, materials, methods of implementation, algorithms, and methods of organizing artificial memory. Just as nature in the process of biological development experimented to produce higher forms of intelligent life, designers can experiment with morphology, organization, artificial brain construction, designed needs, mechanisms of motivation, emotions, and ways of educating artificial entities to produce intelligent machines. We cannot determine how long it takes for machines to become self-conscious beings interacting with people, but we also do not see it as impossible to meet obstacles so that such machines can be created as creatures constructed by humans or other intelligent machines.

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

Imagine a robot that has millions of touch receptors, mostly located in the fingers of both hands. Suppose then that this robot can recognize sensory arousal, has associative memory, can perform coordinated movements that ensure its survival in the environment, has an internal motivation system, can learn, and is autonomous. It may seem that such an assumption immediately raises many doubts as to the current technological possibilities and the state of knowledge about the organization of memory, the way of learning, and the construction of representations that use sensory arousals, which can prevent the effective operation of such a robot. And it is true that many of these problems, due to the degree of complexity, have not yet been solved in a satisfactory manner. However, none of them deserves to be called an unsolvable problem (hard problem) in understanding and explaining consciousness. Let us deal with the difficult problems, which include qualia, assuming that the easier ones have already been or will soon be solved. A difficult problem is, for example, how the robot's mind can create feelings related to the perception of the world and how the mind gains consciousness using these feelings. In our book, we tried to answer these questions.

A robot learns the world by touching and manipulating objects. In its memory, sensory-motor relationships are created that allow it to build spatial models of objects, anticipate their reactions to hand movements, and feel the pressure they produce in response to the clamping of the robot's fingers. This gives the robot an idea of whether the objects are stiff or soft, rough or smooth, elastic or brittle. The robot, anticipating reactions to its movements, feels the objects it touches. Maybe this feeling is slightly different than a person's reaction to touching the same objects, but some feeling is undeniably the same. People may have more touch receptors, different brain structures, or different needs, but the basic properties of objects, such as softness, smoothness, or surface roughness, will be similar. It can be similar with other sensations—robots getting to know the world by using their senses and motor functions will feel it and represent it in memory just like we do. The degree of similarity will depend on the sensitivity of their senses and the accuracy of control of motor functions. The construction of such sensory and motor functions is and will be a huge technological challenge for a long time, but it is not counted among the difficult problems of consciousness. The problem is also the processing of large amounts of information flowing through the sensors, their storage in memory, and the control of motor functions of robots in real time.

There is a belief that meeting the technological requirements for the construction of intelligent and conscious robots will, due to the high computational complexity, force the use of supercomputers with even higher computing power, which we do not yet have, than the classic AI. It is doubtful that such computers will ever be created, however, and if they are to be created, then it is expected that they will be quantum computers or something even more sophisticated, made with technologies that we currently have no idea about. This would postpone the future of conscious robots indefinitely. But it seems that such a danger does not actually exist.

Using sparse connections, a relatively small number of neurons suffice to represent complex objects. Nowhere is it said that the qualia of the artificial system must be highly defined. The visual, auditory, and tactile fields occupy a large part of the cerebral cortex, and these can be reduced. Even more important are the learning procedures during which basic qualia, the appropriate quantum of knowledge, the model of reality, and finally the worldview must be created. If the elements of this reality are people using speech for communication and the artificial system has effectors in the form of a speech synthesizer, the conscious artificial intelligence will have the chance to learn how to speak. The rich language of reporting and the exchange of information about its intentions, emotional states, and needs will make it easier for us to recognize that the system really has consciousness.

High technological requirements have constrained the development of artificial intelligence for years. We have a twofold problem. In order to face the functional efficiency of the human brain, we need the power of supercomputers. Support for 10^{14} synaptic connections (assuming that only 10 percent of neurons are excited at one thousand operations per second) requires the computing power of a computer that would operate at 10^{17} operations per second.

The first among currently fastest supercomputers Fugaku, runs at $415.5 * 10^{15}$ operations per second. It is powered by Fujitsu's 48-core A64FX SoC. For machine learning, neural networks and artificial intelligence applications, Fugaku's peak performance is over 1,000 petaflops. Number two is Summit, an IBM-built supercomputer (OLCF-4) that delivers 148.8 petaflops operations per second and uses 13 MW of power. The system has 4,356 nodes, each equipped with two 22-core Power9 CPUs, and six NVIDIA Tesla V100

GPUs. The nodes are connected with a Mellanox dual-rail EDR InfiniBand network. The Chinese supercomputer TaihuLight can perform 93 * 10^{15} operations per second. However, TaihuLight has over 10 million CPUs and uses 15 MW of power. The brain needs only ten watts to do the same tasks. Miniaturization of the computing hardware helps to reduce the power requirement. A company in California called Cerebras produced wafer-scale super computer CS-1 has 1.2 trillion transistors organized in 400 000 processing cores delivering top performance of 9*10^{15} bytes per second. Thus it is over 20 times slower than Summit supercomputer but it uses 1000 less power.

A much better solution is to use dedicated integrated circuits, such as the aforementioned Neurogrid, which need much lower power than general-purpose supercomputers. Neurogrid simulates millions of neurons connected by billions of synapses in real time, competing with a supercomputer and consuming one hundred thousand times less energy—five watts instead of a megawatt! Neurogrid simulates six billion connected synapses into structures inspired by the neuroanatomic structure of the brain. Axons are often connected in the neighborhood so that nearby neurons receive input from most of the same axons, in accordance with the way the cortex is organized. Local links between neighboring silicon neurons emulate these nonuniform neighborhoods by means of analog communication, evoking postsynaptic potentials in the programmed area of the artificial cortex. For example, in an area of six artificial neurons, a single axon can create one hundred synapses, increasing the total number of synaptic connections from sixty million to six billion. But even these latest solutions are three to four orders of magnitude below the brain capacity in terms of both computing power and energy consumption. Some hope may hang on new technologies, like building an all-optical network (Zhuo et al. 2019). In recent work published in the Optical Society's journal described a new two-layer, all-optical neural network and successfully solved a complex classification task. According to the authors, an all-optical scheme could enable a neural network that performs optical parallel computation at the speed of light while consuming little energy.

How long does it take for the neuromorphic chips to rise to the challenge and match the computing power of the brain? It is difficult to answer this question. Rapid progress in the computing power, which we have observed since the 1960s, has clearly slowed down. Twenty-five years ago, Intel produced the Ni1000 chip, which had 1,024 neurons and performed twenty billion operations per second. The SyNAPSE chip produced by IBM in 2014 implemented one million artificial neurons. In our opinion, we still need at least twenty-five to thirty years of further, equally effective development, unless there is a faster change in the performance of hardware after the introduction of a new technology unknown today. However, hardware is not everything. There is also a need for significant progress in the development of artificial intelligence to use these computational powers to create intelligent machines.

Observed in recent years, progress in the development of artificial intelligence and machine learning and their use in gadgets, mobile devices, and large systems has been huge and will continue. So the question of development leading to conscious machines is not a question of whether such a development will take place. This is certain. The only question is when this level of development will be sufficient for conscious machines to arise and be used in practice. In the nineteenth century, the massive use of steam engines led to the industrial revolution. The twentieth century saw the development of roads and means of transport, which led to mass urbanization and technological revolution. The second half of the twentieth century saw the development of computers responsible for the information revolution. The twenty-first century began with a massive increase in internet access, communication between people, and social media. Artificial intelligence is entering our daily lives more and more. Voice communication with machines is more commonly used in services, autonomous cars are already running on city streets, and robots have been used for years in industry, performing work that would be tedious or dangerous for people. Artificial intelligence is used in finance, medicine, aviation, games, and education, and its applications and complexity are still growing.

Visionaries of conscious robots are already proposing machinery architectures that understand what they do, have impressions and emotions, are aware of their experiences, understand speech, and use that speech to express their thoughts, plans, and needs (Haikonen 2003). Haikonen claims that the robot, which will work according to the architecture he proposed, will feel pain. It will try to avoid this pain—that is, it will be conscious of its experience—and if it is able to speak the language, it will say it is conscious (Haikonen 2012). He discusses existing tests and tests presented in the literature on consciousness. He rejects tests that are insufficient to establish the consciousness of machines, such as the Turing test (Turing 1950) or the test proposed by Koch and Tononi (Koch 2011). He claims that only by asking the machine about its impressions, which requires the machine to be able to use the language, will we be able to determine whether it can have internal feelings of qualia. Similarly, if a robot can describe the flow of its thoughts and associations, and if we know that this description was not programmed and that the robot's internal architecture was able to generate these associations and thoughts and interpret them, then we can recognize that the robot is conscious. The robot consciousness test from Haikonen focuses on the following issues:

- Does the robot have mental images of objects or activities?
- Can the robot describe the content of its conscious experiences (observations, thoughts, plans, needs, etc.) to itself or others, and does it recognize that these are its experiences?
- Is the robot able to describe its physical feelings (qualia)?
- Does the robot remember what has happened recently?
- Does the robot feel pain, and if so, which kind?

Haikonen also presents tests for self-consciousness. The self-conscious robot will be able to determine its body, thoughts, actions, decisions, or needs. First, he determines that the self-aware robot will be able to distinguish its own body from the environment using the sensory-motor system, which provides the impression of touch, pain, and temperature, according to a simple rule: if it hurts, it is me, and if it does not hurt, it is not me. Pain is important from the point of view of protecting the robot from damage. People who do not feel pain suffer significant damages to their health and injure themselves more often than people who feel pain when they hurt themselves. Robots also need tactile or proximity sensors in order to protect themselves from collisions and damage.

Based on tactile sensations, the robot will be able to define the boundaries of what it perceives as its own body, and this is the basic condition for creating the concepts of self, self-determination, and self-consciousness. But such a functional test for self-consciousness is not enough. Consciousness can only arise in a mind that is able to create feelings, learn, have working memory, plan, think, create goals, and build a value system. The self-conscious robot will be able to distinguish its own thoughts and states of mind. It will be able to determine who it is, how it feels, what it wants, or how it sees others. In addition, self-consciousness should be supported by an episodic memory that contains information about its own history and allows it to maintain continuity of the self. Knowledge of the cognitive robot's architecture is helpful to recognize its self-consciousness. The robot must have a sensory-motor system, episodic memory, and the ability to create conscious experiences, but this is not enough. The existence of the robot self should be tested functionally by checking whether the robot refers to its body and whether it is able to have associations related to its thoughts and states of mind.

One of the known tests for self-consciousness is discussed in part I, the mirror test; this test can also be used to test the self-consciousness of a robot. The newly built system will not be a ready-made product. It will have to acquire the necessary knowledge for a long time and gain experience in its use and in the use of its own body and mind. It must have time to learn to speak. It must have time to develop logical inference or create a new language of mathematics. Let's hope that such developments will take a shorter time in the history of robotics than in the case of *Homo sapiens*.

Can an artificial, self-conscious system achieve a degree of intelligence and consciousness equal to people's? Does the proposed model of natural and artificial mind architecture allow for adding anything to current speculations? The most important benefit of the models presented here is the perception that the conscious use of language requires a relatively high degree of consciousness. People have very different levels of consciousness, and there will probably be a period when the intelligence of artificial minds will approach the average level we observe in people. However, it is very unlikely that it would stop at this level. The model presented does not contain any limitations on the expansion of memory areas of neural memory fields or the increase in the number of processing levels. The speed and reliability of electronic systems will ensure greater efficiency of artificial systems. Efficient homeostatic systems will ensure effectual work without fatigue and the related breaks for some kind of sleep. However, one should not expect a miracle, a dramatic overcoming of human intellectual abilities. Even the most abstract concepts require embedding in qualia, and these, as mentioned above, will be similar to ours. Of course, if we equip machines with more perfect senses, x-ray vision, the ability to feel the magnetic field or receive ultrasonic waves, or anything else, their wealth and variety of sensations will increase. This may cause development of new concepts and new words, new feelings, sensations and meanings will appear that are difficult for us to understand.

However, today we are also able to broaden our sensory experiences by using tools and devices that convert all kinds of stimuli into ones we can register through our senses. New ways of perceiving the world through machines will not be shocks for us. At the highest levels of abstraction describing the models of our world, one cannot expect a sudden detachment of their level of complexity from the models we use. Each generalization toward a higher-level model must be based on data. We realize that in the world of micro- and macrophysics. Such data will be harder and harder for us to get. Of course, efficient intelligent systems may allow us to plan new experiments and, in large data sets, to see regularities inaccessible to our minds. However, it will not be a rapid process.

For a long time, human minds will keep up with the interpretations of more intelligent minds, just as a student keeps up with his teacher. Einstein's one mind was sufficient for the entirety of humanity to benefit from his theory. Similarly, the outstanding minds of artificial machines may for a time set the paths of human thinking and understanding of the world. Unfortunately, the biological evolution of the brain is too slow, and one day its adaptive capacity will prove insufficient. We will then need to integrate our brain with an artificial support system. Even later, the biological addition to a superefficient machine will prove completely unnecessary.

Let us think about who needs the conscious robots and what the benefit is of having consciousness or self-awareness. If you evaluate these questions from the point of view of species development, it can be argued that conscious animals can achieve higher intelligence than the unconscious, and that essentially self-consciousness is characterized by animals at a higher stage of development. Self-consciousness brings with it the possibility of recognizing the state of consciousness in others. It is a recognition not only of their behavior but also of needs, patterns of thinking, emotions, expectations, and a system of values. This allows us to predict the actions and reactions of creatures competing for resources in the same environment. Therefore, this helps to achieve greater efficiency in the exploration of the environment and survival in a limited-resource environment.

The conscious system is able to understand the social effects of its actions, and in addition, the conscious robot, guided by the understanding, can better fulfill the restrictions imposed on robots by Asimov. Unconscious robots are obedient tools in the hands of people, and today they kill people, either as directed or autonomously. A conscious soldier can and should refuse an order leading to atrocities. We can expect similar behavioral constraints from conscious robots.

Will conscious machines be a threat to people? In his science fiction novels, Asimov proposed three laws of robotics to protect people from artificial intelligence:

1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.
2. A robot must obey orders given it by human beings except where such orders would conflict with the First Law.
3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.

Asimov's laws were criticized for not taking into account the fundamental principles of coexistence, including the principle of mutual respect. These laws subordinate robots to people, and the asymmetry of relationships in which robots are treated as subordinate beings is clearly visible. Therefore, such formulated laws can apply only to machines and automats with very low autonomy. In addition, it is pointed out that the rules of action of autonomous robots will be controlled by the constructors and their sponsors, and nowadays these are mainly the defense industry and the military, which are interested in using thinking robots on the battlefield. The civil industry, which finances research on advanced robots, is mainly concerned with profits rather than with the moral behavior of robots.

Sir David King, chief scientific adviser of the British government, believes that when feeling, conscious robots are created, they should have the same rights as the people. According to his position, states will be obliged to provide such robots with the right to work, education, residence, and health care, which would mean repair and regular maintenance service. Just as we do not approve the abuse of animals, understanding that they suffer and have feelings and emotions, so we should take into account the needs of conscious robots and strive to provide them with means to sustain their self-consciousness.

Limited resources can become a source of conflict between human civilization and robots. This can be avoided if we consider both the needs of both sides and the available means. But that must be learned. So far, we have trouble ensuring the necessary conditions for the development of existing intelligent beings, such as animals and people. If we do not learn this before the intelligence of robots exceeds ours, we can expect to be treated by them just as we now treat animals.

The worrisome situation in which an artificial intelligence system can itself fully automatically produce derived artificial intelligence systems is already in place today. ImageNet, a system built by Google in which artificial intelligence controls the machine-learning program, outperforms all human-built machine-learning methods, achieving the best results in real-time recognition of scenes in a data set for object classification.

According to modern futurists, if artificial intelligence reaches such a level of development that it will be able to improve itself, then our influence on it will be very limited. Humanity can then find itself in a very difficult position. It can be expected that, on the one hand, postindustrial societies will be highly dependent on artificial intelligence systems. On the other hand, the uncontrolled development of the new generation of superintelligent robots and the infrastructure control systems used by humanity can cause us to stop understanding where the systems working with robots are heading—especially when they will work closely together on a global basis. Then it may already be too late to strive for recovery of any control over the world of artificial intelligence.

It seems that we can easily prevent this in many ways. However, the Swedish transhumanist philosopher Nick Bostrom, in his book *Superintelligence: Paths, Dangers, Strategies,* points out that all these methods can fail or be circumvented in an amazingly easy way. The simplest seems to be to mount a superswitch, which, at one push of the button, will turn off the power to all robots or turn their consciousness off. However, this may not be sufficient to stop robots that will be able to improve themselves and have time to prepare for actions that would prevent such switching off. Our superswitches could be removed ahead of time, especially if these superintelligent robots steer the power systems and, de facto, will have their fingers on the switch.

Perhaps we can therefore deprive robots of any executive devices so that without us or simple robots, machines cannot do anything. But this idea may prove unreliable. Superintelligent beings could easily subdue other machines and robots with less intelligence. By the same principle, they could subordinate less intelligent people. After all, persuasion, psychology, and methods of manipulating people through the promise of rewards, punishment, or blackmail will not be unknown to them. Beings with higher intelligence tend to subordinate other creatures quite quickly, exemplified by the domination of people over other animals.

In a recent IEEE Spectrum article (Russel, 2019) author discussed risks involved in development of artificial minds and cited opinions of many AI researchers including Bostrom's. He cited LeCunn who downplayed the risk from AI saying that "there is no reason for AIs to have self-preservation instinct … unless we built these emotions into them", but as the author states the conscious machines will acquire them anyway. In a similar way he criticized Rodney Brooks for saying that machine cannot achieve goals set for it by humans without causing problems for them. Quite a contrary he states, the machine will be aware of these problems but will not consider them as problematic. In summary he concludes that researchers have failed to explain why superintelligent systems will remain under human control. Our position is that indeed superintelligent systems may one day be out of our control, may not share human values or even consider us a nuisance in their further development. The same may happen if extraterrestrial intelligent beings will one day discover our existence. They may care less about our survival and well-being.

Therefore, the creators of artificial intelligence should ensure that it develops sympathetic to people. Nick Bostrom states that it is necessary to ensure that the goals of artificial intelligence are conducive to people (Bostrom 2005). He argues that kindness toward people should be designed from the outset, and designers should know that their designs can be flawed and that the robot will learn and evolve over time. The challenge is therefore to design a mechanism for the evolution of artificial intelligence within the control system, which ensures that machines despite their development remain benevolent to people. We must learn how to instill in them appreciation for general values such as empathy, altruism, fondness for maintaining diversity and dignity of all conscious beings. Something humanity never learned themselves would be our only hope.

In these futuristic considerations concerning artificial intelligence, it is also stressed that artificial intelligence will be able to understand the notion of kindness toward people only when it becomes self-conscious. This is also our position, and that is why it becomes obvious that superintelligent robots need consciousness.

And what about the problem of criminal behavior of people who would like to use intelligent machines for their evil purposes? Hackers have been using computers almost since their dawn to destroy programs, databases or hardware. Even today, cyber war is in the arsenal of developed armies around the world. Will further development of intelligent machines increase this kind of threat to our lives? Are we not giving criminals greater opportunities for malicious activities by developing intelligent machines? What about organized piracy and cybercrime at the state level? Already today, disinformation distributed via the Internet, is used by state organizations to spread uncertainty, cause distrust of state authorities, manipulate minds, or to disrupt societies.

It is true that today some people use technology and artificial intelligence for criminal purposes. However, the answer to their criminal activities is not to undo social development but to further develop technologies and methods of containing and prosecuting crime. A democratic, well-organized state, can stand up to criminal groups. But can it be different? Sure it can. Even today in many countries there are dictators who subjugate millions of people, using state institutions. What is the role of machines in this? Instrumental. Can it be different? Of course it can. Moral, intelligent machines can warn us and protect us against criminal activities of people, especially when their intelligence, morality, and the rule of law would not be subject to human ambitions, willingness to possess ruling power, pathological behavior, or mental illnesses.

In fact, questions about the use of technology for criminal activities are similar in nature to questions about whether machines more intelligent than human will want to destroy us. We already answered to this fear with a note of hope as to how and why moral machines would want to preserve human civilization. And on the same principle, we trust that they will have a moral advantage over the destructive powers of those, who for their own criminal purposes, want to sabotage the further development of our civilization. This hope includes stopping the destructive actions of both human and artificial saboteurs.

Life and Death of Conscious Minds

The problem we face constructing conscious robots is the duration of their conscious mind. At the end of his book *The Cognitive Approach to Conscious Machines* (Haikonen 2003), Haikonen imagined the hypothetical meditations of a conscious robot, who is wondering where he came from: “My child, what does it matter, you will live forever, you will never die.” These words express his belief that the eternal continuation of the machine mind is possible. After all, the chips from which the brain of the machine may be built will hardly ever age. In addition, electronic components are easy to replace. But is it right? A conscious cognitive mind rests on a unique structure of connections between neurons. Even if we can keep the robot’s mind at full efficiency by replacing aging modules, is this operation advisable? Does long-term maintenance of consciousness in the machine make any sense? Is the long duration of the system or organism a value that we should cultivate and develop? It is a philosophical issue that can be interpreted through the prism of other values precious to the person who reflects on these dilemmas. It is worth starting with what system or what body we mean. Intuitively, most of us feel that the value is in the continuance of structures, organisms, and systems of great complexity, the existence of which is rare because of the apparent impossibility that such beings, structures, or organisms arise at all.

Nature provides us with stunning examples of the wonders of the plant and animal world but also unusual structures on a local, geological, or cosmic scale. We have no impact on many of them. But many we can destroy, decommission, or distort. The growing awareness of the value of such objects prevents us from doing so. We strive to protect the rare specimens of a lively and inanimate nature. One of the values that we protect is beauty. The better we know the nature that created us, the world in which we live and which we exploit, the more we appreciate the captivating beauty of its oddities. Incidentality is associated with the complexity of objects because the existence of extremely complex structures seems very unlikely. So we are willing to appreciate the works of nature and human creations characterized by beauty and complexity. In this complexity of the structure hides the great work of the creators or forces of nature that led to its formation. Sometimes it is a multitude of interactions; sometimes the selection is from many less-permanent objects, and sometimes it is simply the passage of a lot of time necessary for its creation.

Despite the apparent simplicity of the processes leading to the manifestations of consciousness, we realize that they can only occur in beings of appropriate complexity. This is because one of the pillars of consciousness is the ability to build an adequate model of reality. If this model is sufficiently rich and contains enough information for an efficient exploration of the environment, then the structure to reflect it must also be sufficiently rich to accommodate that amount of information. The conscious beings do not have to be characterized by beauty of shape and appearance, but beauty is their subtle perception of the world that surrounds them and the sophisticated response to perceived phenomena. They simply have the ability to absorb the beauty of the world and preserve it in memories, imaginations, and dreams. If anything is worth careful protection, it is the consciousness that is the essence of what we value.

The powerful complexity of conscious beings stems not only from their material structure but also from the information that is being collected during their lifetime experience. At this point, time appears as a factor increasing the value of conscious beings. A person who is experienced, educated, and active for a longer period of existence represents a greater value to their loved ones and the community in which they operate. Hence, there is respect for the elderly and confidence in the experienced.

The cult of knowledge enamored with material structures is probably best illustrated by the devotion to Einstein's brain. After Einstein's death, his brain was preserved, and then fanatical researchers, believing that they would somehow reveal the knowledge condensed in it, fought an insidious struggle for its fragments.

It seems to contradict the observed cult of youth and extraordinary concern for the younger generations. This contradiction can be explained by the clever trick that life has invented. Beauty does not need to be preserved. Beauty can be recreated! The most complex brain structures can be reproduced through genetic processes and processes of education and knowledge transfer through science. Therefore, the hopes of successive generations are the next generation and the undisclosed potential of children's minds.

The long duration of conscious beings leads to a dilemma faced not only by the builders of conscious robots but by modern societies, in which the lifetimes of the members forming them extend. Sustaining oneself, which is the very essence of consciousness, requires the maintenance of the existing neuronal connection structure that contains things like knowledge, habits, routines etc. And what if the world went on and the knowledge possessed by a robot or an old man became obsolete and the habits and means of learning obstruct the efficient organization and use of newly acquired knowledge? After all, the way the machines (but also people and animals) learn depends on the previously accumulated knowledge and influences the ways and possibilities of gathering new knowledge.

There is a moral problem before the constructors: Do you keep obsolete robot constructions alive? Conscious robots will have their rights, but they will also be afraid of death, just as people are afraid. This could be compared to a situation in which we tell an elderly man that his brain is no longer keeping up with the changes in technology and that he himself has become a relic of the past and a burden on society. Would we decide today to condemn such a person to be shut down only for this reason? Definitely not. Just the opposite—we strive to extend people's lives and to preserve their good health as long as possible, even though we know that they are not able to function as well in terms of assimilating technical innovations as their grandchildren are. Moreover, many cultures appreciate the experience and life wisdom of the elderly. Historically, in many cultures, a senior council decided what the community was supposed to do to protect itself from enemy attacks and survive. In the modern world, leaders are often experienced people in middle age or older, people whose ability is proven in terms of skills and leadership abilities and who are familiar with social mechanisms. But their success is sometimes the result of a fairly lengthy process of proving themselves and advancing in social structures. Older people have many life experiences and routine actions, but we know that the greatest scientific discoveries are most often the work of young people, who are less affected by stereotypes and barriers to impossibility, who break down barriers to thinking and have new ideas and curiosity about the world.

Perhaps nature has decided that it does not pay to continue the life of old individuals, that from the point of view of the development of the species, a more efficient method is the renewal of vital and mental forces through new generations. The elderly, despite their knowledge and routines, perish. Each species has a natural life expectancy, followed by a mass death of cells and the whole organism. But progress in the development of the species resulting from evolution is very slow; it is a method of trial and error. In technology, we could develop a shorter and more efficient improvement process for intelligence. But is the death of an individual (in the sense of interrupting the continuity of his self and maintaining self-awareness) necessary? Let's try to answer this question.

There are two aspects of prolonging the life of an individual whose brain was shaped in the distant past and who no longer fits the modern world. The first is to add new areas of memory where new knowledge would be created. An attentive analysis of this possibility indicates that the added memory areas would have to be integrated into existing storage so as to not compromise self-awareness and ensure a sense of continuity. How could it be done?

For example, one could introduce new elements that would connect locally to existing memory structures, duplicating and complementing the features of aging computing elements. This type of memory repair has the potential to prolong the life of individuals whose brain structures have been destroyed by illness (e.g., Alzheimer's disease). In neurobiology, the options for this involve methods that utilize stem cells. These cells, supplying new neurons, can fill damaged areas of the brain. Because the structures they create are localized and work in specific functional areas, they take over the functions of neurons that functioned in the area before destructive lesions arose. Acting in the vicinity of neurons, they provide them with the necessary data memory structures. Here, one can exploit naturally the redundancy of both the neural representation and the knowledge locally associated with neighboring memory areas. These memory areas respond in a similar way to arousal as replaced by them areas of the damaged brain.

The trouble is that the bigger the damaged area, the harder it is to reproduce. Simply local information, which was saved by adjacent neurons, disappeared along with them (by damage) and cannot be recreated on the basis of the similarity of excitations. It would be necessary to introduce new brain cells at the same time in many places so that they could sustain the aging and malfunctioning processing in neurons. A mechanism would also be needed to deliver such new neurons to sites at risk (or aging). The new, complete module is not enough to preserve the knowledge and personality of the mind. It can only become an additional memory module to be used. The knowledge that will be saved in this area depends on its location and the type of excitations coming into this area from other areas of the brain.

The second aspect of prolonging life, apart from sustaining existing structures, is development geared toward a new way of thinking. This can be very difficult. First, the learning brain uses existing structures for learning. Thus, the accumulated knowledge has a very strong influence on how the new information is processed and adapted to the organism's needs. Here we will see a clear influence of the habits, stereotypical thought, and flip-flops that characterize older people. Young people organize their brains differently, adopting new knowledge without prejudices or dependence on the baggage of previously learned facts, skills, and habits. Today a granddaughter often teaches her grandfather how to use new technologies (e.g., an iPad or iPhone), not because she is smarter, but because she may just know it as the only way of using the internet. The five-year-old brain has many more neural connections than the brain of a teenager. Such a young brain can quickly learn many new things. It is not bothered by the already existing associations and educated synaptic connections.

This is not the case with Grandpa. First, his capacity for learning new skills is less because he has fewer available connections between neurons. Think of the early programmer, who first used computers that were controlled by paper tape or using a card reader programmed in the machine's internal language and then in Algol, FORTRAN, Kobol, Lisp, and several other languages. He found it increasingly difficult to learn new programming styles, data storage and processing languages, information protection, ways to search for information, and social networking programs—skills that his younger colleagues easily mastered. In this case, maintaining the continuity of existence, and therefore the originally learned knowledge and structures of memory, is a major limitation of adaptation to the new environment, and this adaptation, according to our definition, is the measure of the success of the intelligent system.

The associated moral problem—sustaining life or interrupting continuity—is a problem that we cannot understand and judge today. If the recommended approach is to replace the body's old brain with a new one, hence losing the memory of what was and the continuity of self, will this be acceptable in future communities? Today it seems to us to be synonymous with system death. What if the system does not want it? Its self-consciousness, which has developed over the years, can be of greater value to it than the promise of a new mind. Psychological research shows that within a few years, human mentality changes sometimes radically. Personality, worldview, value system, and even the memories of the past change. The individual becomes another human. Long prison sentences do not make much educative sense (though they may be justified in terms of revenge or crime deterrence). After a few years, we have jailed different persons. However, even if their memories fade and behavior changes due to incarceration, the change of their personality may never be complete. The reasons of his criminal behavior still may exist, even if only in their subconscious mind. That is why, in spite of long incarcerations their past way of thinking and criminal behavior may return. Yet, without full participation of our consciousness, we change over time. People generally do not want to come to terms with this. Traces of memories of the past create a form of continuity. But in fact we can go to sleep and wake up as a slightly different person. And if we do not wake up? This thought frightens us. Will robots also be horrified if they understand that during their sleep, their service technician can erase or add something to their memory? Will they fight against this? Or will they do the opposite: wanting to be useful to society, will they voluntarily agree to euthanasia? For moral reasons, shouldn't we protect their personalities? Today, it seems to us that eternal life is a desirable value and that it has an unquestionable advantage over death. Many of us simply want to participate in this future life, looking forward to the times of new discoveries and seeing how society can cope with various problems, even if we would not be able to help them.

Discussing human motivations, we described curiosity as a strong drive to act. In our youth we are curious about new experiences of love, food, sightseeing, and exploring the world. In the mature person, this curiosity weakens and affects us less, but we are concerned for our children, our successors, and the fate of the works we have begun. In old age, we have only a faint interest in the fate of the great-grandchildren and future generations. At the end, curiosity expires, preparing us to say goodbye to the mortal life. Being aware that we are a liability, should we not choose all the more to end that life? Nature has clearly decided that this is the best solution.

Our Vision

We are convinced that a conscious system can arise through the development of robots using an autonomous, motivated learning, in which robots themselves determine their needs, objectives, and methods of operation. Autonomous robots influence their environment, using their own embodiment and motor functions, and observe the effects of their action through their senses. The predicted effects of these actions give direct sensory impressions that the robot brain perceives as qualia, creating internal representations of the observed external world. The robots see the world as it is. The activity of neurons in their brains trigger subjective sensations of the physical characteristics of perceived objects. The robots can be conscious of their actions and the distinctness of their body from the rest of the environment, which leads to self-determination and the concept of self.

The robots, trying to fulfill their needs, build a model of their world, create the higher-order needs, and learns methods of their satisfaction. Both artificial emotions and character traits can arise as a result of interactions with the environment and other agents that will either cooperate with the robot or compete with it for the limited resources needed to achieve their objectives. Representations created in the robot's associative memory are based on the subconscious parallel processing of sensory stimulation and streams of sensations associated with them. Building representations of objects, ideas, and action plans is associated with motor control, leading to activities that aim to meet the needs of the robot.

We have presented a functional organization of cognitive architecture in which it is possible to obtain conscious sensations, planning, thinking, prediction, episodic and semantic memory (Horzyk 2017), and control of motor functions. In this architecture, one of the leading roles has the working memory in which the robot prepares a plan of its activities, focusing on selected objects in its surroundings and assessing their suitability from the point of view of its needs. We pointed to the critical impact of signals (both subconscious and conscious) that, competing with each other, switch attention, focusing it on the selected elements of the environment, the needs of the robot, or the planned actions. These signals, combined with mental saccades, are also the basis for an uninterrupted stream of thoughts by a robot that is able to perform its activities in the environment while noting unexpected ambient signals, its internal needs, and alternative concepts. These concepts are obtained through associations of different ideas, solutions, and object models in the process of planning an action with the collaboration between semantic memory, episodic memory, and motor actions.

Conscious sensations, which are the result of perceptions of the world related to the stimulation of the robot's senses, lead to the imaging of the actual properties of perceived objects and phenomena in the robot brain. At the same time, the conscious thought processes involve predicting the effects of the robot's actions with all the consequences and in particular the anticipation of how such actions can affect the satisfaction of the robot's needs. The robot functions as an autonomous agent. An autonomous agent has a natural need for action, which results from the very definition of an agent, as an interacting being. Its self-awareness means distinguishing the body or housing of the agent from the surrounding environment. Having a body capable of acting entails many further needs, and therefore a strong motivation to satisfy them. The system requires the supply of energy and necessary materials to carry out tasks, provide information about the environment through the senses and information on the activity of other agents and changes in the environment, removing metabolic products, etc. Therefore, we call our system a motivated system. Satisfying needs or failing to meet them causes positive or negative emotions, respectively. That is why we call the system an emotional system. The consequence of gaining awareness in such a system is the emerging the MEM.

The physical process of interactions between neurons of robot brain transforms into a subjective feeling that something that is observed actually exists and possesses the physical properties attributed to it, such as size, color, temperature, roughness, or weight.

The robot's brain does not perceive the activity of its own neurons, but it represents the physical objects and phenomena that take place in the outside world. On the other hand, the robot is able to distinguish its own embodiment, senses, and motor functions as different from the rest of the environment in which it operates. This leads to self-determination and creation of the physical impression of distinctness from the rest of the world, which is a symptom of self-consciousness. The robot is able to think about itself in the first person, remember its past, identify the limitations of its own body, or even realize the characteristics of its character, memory limitations, and skill level compared to other individuals.

In this book, we show how the processing of sensory signals through neurons in the robot's brain and the control of its activities can trigger a subjective impression of the existence of an external world. Through the impressions and focus of attention, the robot realizes what its attention is focused on, what the properties of the observed objects are, and how it can use them to fulfill its needs. These conscious sensations can be transferred not only to the objects observed by the robot but also to its internal thought process, where the subjective impressions stored in the memory representations allow it to think about the memorized episodes, activities, skills, and impressions as though they were real. Using these conscious sensations, the robot is able to determine their suitability, choose alternative solutions, or deliberately focus its attention on the problem it wants to solve.

The conscious robot is able to describe its own sensations, observations, emotions, and conscious needs. The symbolism associated with these observations is needed, and the ability to describe objects, ideas, and activities using a symbolic language to communicate its own thoughts facilitates and makes credible the robot's conscious observations. It must be pointed out, however, that it is not necessary to have such a symbolic language for the conscious experience. The language facilitates communication between different individuals and can be used for a more accurate description of their own sensations and observations. For this, it is helpful to have similar internal needs, systems of memory organization, and methods of working, perceiving, and processing information.

We realize that further progress in the development of conscious machines depends on a number of factors, including advances in the technology of miniaturized computing elements such as artificial neurons, their speed of operation, and their power consumption; the fields of neuropsychology and molecular biology; information processing; algorithms; memory organization; modeling phenomena of perception; and the development of the robot's control and motor mechanics. These issues are very important, and their effective solutions and use in the practice of building artificial brains can take many years. Nevertheless, though these are issues to be solved, we think the difficult problems of consciousness have been explained. We hope that this book will interest students, future researchers, and inventors and will inspire them to work on machine consciousness.

The age of conscious machines is coming. Many readers will not believe this, claiming that our functional description of a self-conscious robot is a far-fetched science fiction. To answer them, let us quote from a presentation by researchers at the Foresight Institute in Palo Alto, California. "Imagine looking far into the future. If what you see looks like science fiction, you might be wrong. But if it doesn't look like science fiction, then you are definitely wrong."

Epilogue

How Far Can We Understand Ourselves?

We often ask ourselves: How can we understand our own consciousness? We hope that this book will help to explain how the consciousness can arise in tangible, natural beings as well as artificial ones. But do we fully understand it? Can we comprehend its possibilities and limitations? Can we describe it mathematically?

In part II, we tried to describe our efforts to build models of conscious machines. Building a model of anything means understanding the essence of it. However, the depth and detail of this understanding depends on the specificity of the phenomenon or object. Full understanding means that you can build a model exactly like the original. So do we build machines that imitate our minds? Yes, and no. For the attentive reader, it is clear that the mind is a very personal matter. Its character depends not only on the construction of the brain, which is its substrate, but also on the history of its formation, on learning, and on life experiences. In examining the mind of a man, we cannot be sure that the mind of his twin brother has the same qualities. Probably it does not. The mind of the artificial machine will be even more different than the human or animal minds of biological brains. However, the efforts of the designers of smart self-conscious robots do not aim to build a system that accurately replicates some particular mind, animal or human. We have a conviction that if we understand how consciousness is born and can build a machine with a conscious mind, we will surely continue to develop it. Further generations of these machines will be created at an accelerated pace that may be determined by some new Moore's law.

One day, the level of consciousness and the correlated level of intelligence will reach the level available to the human brain. It is absolutely unlikely that at this moment of development, further progress will stop. So every next model will exceed the possibilities of the human mind. Development will go through faster, avalanched acceleration if these hyperintelligent creations are directed to improve themselves and to design the next, ever more perfect generations. This will force us to familiarize ourselves with cognitive processes in their minds. If we can't handle it ourselves, we might get help from our artificial, superintelligent successors.

The considerations outlined in the previous sections do not touch a significant problem of understanding what consciousness is. Perhaps if, even after reading this book, we better understand how consciousness can arise in specially organized neural networks inside embodied systems, many readers will still ask themselves questions: How does it happen that we feel conscious? Why is it so difficult to define and describe this feeling? The answer is obvious when we recall all the conditions of the manifestation of consciousness described in this book. A sense of consciousness is a simple, direct feeling. Such direct sensations are called qualia. Yes, feeling your own consciousness is a quale! This is a subjective, first-person experience. We will never be able to describe it in a strictly symbolic language and even less so in formal one such as mathematics, geometry, or logic. We can only try to describe our feelings in the language of poetry using comparisons, metaphors, and analogies. We will never find out if another unknown being we have encountered is really conscious of anything in the sense of how we feel it ourselves. The assessment of the assurances regarding the primary subjective impressions from a tertiary perspective will not be possible. That is why it is so difficult to tell whether other people, animals, robots, or newcomers from other planets are conscious or whether we are only victims of elusive illusions.

Of course, the phenomenon of consciousness has its objective side, which can be identified from both the first- and third-person perspectives. We need to distinguish the sense,

the content, and the concept of consciousness shaped by reflections over our cognitive abilities. If the subject responds to stimuli from the environment and intelligently adjusts its behavior to the signals it receives, it smartly plans its activities in a complex, innovative way, —if it manifests the ability to learn—if, moreover, it can communicate and maintain social relationships—then we can suspect that it has some degree of consciousness. However, we will never be certain of what it really feels or if it feels qualia. Maybe it's just a smartly programmed automaton, or maybe it's just a play of inborn instincts and drives. Who knows? If our observations relate to objects from a foreign civilization, how can we distinguish whether its representatives have visited us or if they have only sent us autonomous, unconscious robots? What other attributes of the conscious being could we expect?
We have previously wondered if there is a center of consciousness that makes us aware that we are conscious, a center that tells us when we lose consciousness and when we regain it. It would seem that this hypothesis is confirmed by detection of the "center of consciousness" in the brain by the Xu Liu group from MIT (2012). They detected that the irritated group of cells located in the brain shuts down the consciousness of the individual experimented on. After stopping the stimulation, consciousness returns to its normalcy. Is this a place where our consciousness is hidden? Definitely not. In creating consciousness, as we tried to prove in this book, almost the whole brain and its embodiment are involved.

In addition to knowing that we are awake and conscious, not even thinking, acting, or recalling any specific knowledge, we are able to feel that we are conscious. This happens immediately upon waking up, when we may not yet remember who we are. Similar sensations are known to those who have suffered sudden loss of consciousness. One of the first sensations after reviving, when the eyes are still closed and no sounds are heard, is the feeling of consciousness. Only a moment later do we begin to perceive our environment. Similar sensations were reported in sensory-deprivation states, when senses are isolated from external stimuli and the person tries not to think of anything. What he feels is knowing that he exists and that he is conscious.

Let's run another thought experiment to test whether machines can show phenomenal consciousness: In part I, we've shown that embodied natural brains have morphological and biophysical means to generate qualia and respond emotionally to perceived objects. In part II we showed that artificial brains of autonomous robots, equipped with associative and episodic memory, can create neural representations in their electrical circuits that enable them to achieve their own goals. This is done by associating perceptions with perceived or imagined pain signals, which should be equivalent to emotional states. Let's ask the question: Do epiphenomenons of emotions and feelings appear in these artificial brains? If so, where do they appear?

It would be difficult to explain how such epiphenomenons might not appear in an artificial neural network, just as it does in a natural network. If, as a result of motivated learning, neuronal representations of objects and phenomena have already appeared in the brain, then their stimulation through the bottom-up sensory signal, imagined signals from above (top-down) or spontaneous mental saccades, will produce the same memories and feelings that originally accompanied creating these representations in the brain. As we described in part I, the same neurons will be stimulated as when looking at the scene or imagining phenomenon during the learning procedure. These stimulations are demanded epiphenomenons, and just like in the natural brain, are thereby the neuronal representations. It does not matter that artificial neurons are very primitive and there are less of them than in the brains of the smartest animals. The result will be the same: recognition, understanding and feeling emotions, which is a symptom of phenomenal awareness. These will be symptoms of the functioning of a motivated, emotional mind (MEM).

Deeper Understanding of Consciousness

Fully conscious minds create representations of the world intertwining the gnostic aspect with the phenomenal aspect. Let's imagine that we are taking a boat to the huge iceberg. We feel awe at its size, hear the cold sigh of the ice colossus, sense the power of its mass struck by furiously dancing waves. Any of these waves could throw our vessel onto protruding edges and cavernous crevices. This sense of terror is inevitably a quale. But we can say a lot about this threat in assessing the strength of the waves and our chance keeping the boat intact, taking into account its agility and maneuverability. This knowledge of the threat accompanies the feeling of threat. Let's look for a more sophisticated quale. Perhaps in drawing closer, we will feel the majesty of the iceberg. We can have much knowledge of icebergs. We know that only 10 percent of their mass protrudes above the water level and that they are chunks of ice that have detached from Arctic ice sheets. From Wikipedia, we can learn about their structure and history related to climate change, sea currents, atmospheric and geophysical processes, and so forth. This does not in the least affect our sense of majesty in the face of this ice monster. Can we describe our feelings to someone? Only those who have experienced similar feelings toward mountains, oceans, whales crossing the water, Egyptian pyramids, the cosmos, or infinity, only someone who has experienced this emotion at the thought of unshakable power indifferent to our fate can understand the comparison of our impressions when we come in contact with these phenomena to kneeling before the idea of the emperor or god. This sense of majesty has no direct relationship with any senses. It is a dramatic collision of both ends of our model of the world, a deep sense of the significance of perceived matter and understanding of the abstract idea that it gave birth to.

An extremely reductionist approach, resulting from our model of motivated emotional mind (MEM), does not contradict the consequences of this theory, which leads to the perception of consciousness as the emergent property of the mind. A feeling of consciousness is not the same as the consciousness-raising process. The existence of neurons that trigger it has nothing to do with the state of consciousness, which is an emergent feature of the whole brain and has no specific location. The essence of consciousness is to build a model of reality, to define/understand its place in this reality, and to feel emotion and satisfaction arising from that fact. Learning by planning your own actions requires the association of a conscious mind with the body and the motor mechanisms that perform the intentional actions of the mind. The deliberate action of the mind arises from the creation in the mind of the hierarchy of values as the immanent trait of the reality model. Anticipation of the effects of actions allows the subjective definition of "good." The mind understands what can be "good" for it in the shorter and longer term.

This hierarchy of values also includes a sense of self-importance and significance as a being distinguished by existence and enriched by the hardships of survival. Does it help in building one's own self and purpose? Is it not thanks to this that we have an expansion of the value scale? Does our concern for children, for our own tribe, and for the nation or humanity come from understanding the tremendous value of society and its individuals that results from the enormous work and effort needed to rise, educate, and organize their lives and cooperate effectively? Certainly yes. Just as we sense the majesty of the iceberg, we also sense the majesty and pride of the nation, the state, and these great ideas that characterize modern civilizations.

The formulation of the long-term goal of existence constitutes a sense of self-existence and, consequently, the meaning of the world as a tool for fulfilling one's mission in this world. By developing this thought, it must be stated that the desirability of the existence of matter only makes sense in the presence and consciousness. A widespread conviction of the transcendental nature of consciousness is a generalization of life experience. Dominance of this attitude appears naturally in the process of early learning through analogy. If something is happening around us, it's because we're doing it. And we always act for some purpose. In that case, if something happens without our intervention, it means that someone else is doing it for some purpose. Teleological thinking arises. The default assumption of such thinking is the existence of a transcendent consciousness, giving purpose and therefore meaning to existence. When asked, "Why do cod swim in the Baltic Sea?," if we respond from the point of view of our own consciousness, the correct answer should be, "So that we can eat them." Another equally good answer, if we respond from the point of view of the consciousness of the cod, is "to be able to reign over the Baltic," or, if we believe it, "to the glory of God."

It is astonishing that if we ask about the purpose and meaning of the matter, we must admit that no such purpose exists, if there is no consciousness for which we could formulate such a purpose. However, matter has the ability to self-organize in such a way that consciousness arises. This form of organization is an embodied network of neurons capable of mapping perceived reality. It is therefore possible to reformulate the answer to the question about the purpose and meaning of matter. The meaning and purpose of the existence of matter is the emergence and preservation of consciousness. This sense arises at the moment when consciousness arises.

References

Abbott L.F., (2008) „*Theoretical Neuroscience Rising*", Neuron, 489-495, 6 November 2008.

Abbott L.F., Regehr, W.G. (2004) "Synaptic computation", Nature 431 (7010), 796-803.

Adleman L.M., (1994) Molecular computation of solutions to combinatorial problems, Science 266 (5187), 1021-1024.

Anastassiou, C.A., et al. (2010), *The effect of spatially inhomogeneous extracellular electric fields on neurons".* J. Neuroscience. 30, 1925–1936.

Anastassiou C.A., Perin R., Markram H., Koch C., (2011) *Ephaptic coupling of cortical neurons*, Nature Neuroscence Vol. 14, No 2, Febr.

Anderson J.R., et al. (2004), "An integrated theory of the mind," *Psychological Review*, vol. 111, no. 4, 1036–1060.

Anderson J.R., (2007), "Using brain imaging to guide the development of a cognitive architecture," in *Integrated Models of Cognitive Systems*, W. Gray, Ed. New York, NY: Oxford University Press.

Antic D. Zhou W-L , Moore Anna R. , Short, S.M. and Ikonomu K. D. (2010), "The Decade of the Dendritic NMDA Spike", Journal of Neuroscience Research 88:2991–3001 (2010)'2010 Wiley-Liss, Inc.

Asada M., et al. (2009), "Cognitive developmental robotics: a survey". *IEEE Transactions on Autonomous Mental Development*, Vol.1, No.1, 12-34, May.

Aur, D., Connolly, C.I., and Jog, M.S, (2005), "Computing spike directivity with tetrodes", *Journal of Neuroscience Methods*, Volume 149, Issue 1, 30, 57-63; 2005

Aur, D., Connolly, C.I., and Jog, M.S, (2006), Computing Information in Neuronal Spikes, *Neural Processing Letters*, Volume 23, Issue 2, Pages: 183 - 199, 2006;

Aur D., Jog, M.S., (2006), Building Spike Representation in Tetrodes, *Journal of Neuroscience Methods* Volume 157, Issue 2, 30 October, 364-373.

Aur D., Jog M.S., (2007), "Neuronal spatial learning", *Neural Processing Letters*, Vol 25, no 1, 31-47.

Aur D. and Jog M.S., (2010), *Neuroelectrodynamics - Understanding The Brain Language*, IOS Press,1 – 235.

Aur D, (2010), „Where is the 'Jennifer Aniston neuron'?", dostępne na *Nature Precedings* http://precedings.nature.com/.

Aur D. (2011), Understanding the Physical Mechanism of Transition to Epileptic Seizures, *Journal of Neuroscience Methods*, Vol. 200, Issue 1 30 August, 80–85.

Aydede M. (2009), „Is feeling pain the perception of something?", *The Journal of Philosophy*, vol. CVI, No. 10, October.

Baars B.J. (1988), A cognitive theory of consciousness. Cambridge University Press: New York, NY.

Baars B.J. and S. Franklin, (2009), "Consciousness is computational: the LIDA model of global workspace theory," *Int. J. Machine Consciousness*. vol. 1, no. 1, 23-32.

Bailey C.H., Bartisch D., and Kandel E.R., (1996). "Toward a molecular definition of long-term memory storage". *Proceedings of the National Academy of Sciences of USA* 93, Nov., 13445–13452.

Bailey C.H., Kandel E.R., Si K. (2004) „The Persistence of Long-Term Memory: A Molecular Approach to Self-Sustaining Changes", Learning-Induced Synaptic Growth, *Neuron*, vol. 44, 49–57, Sept. 30.

Baker K., Mattingley J., Chambers Ch., Cunnington R. (2011) „Attention and the readiness for action", Neuropsychologia 49 (2011) 3303–3313

Baranes A. and P.-Y. Oudeyer, (2009), R-IAC: Robust Intrinsically motivated exploration and active learning, *IEEE Trans. Autonom. Mental Develop.*, vol. 1, no. 3, 155-169.

Basawaraj, J.A. Starzyk, A. Horzyk, (2017), "Lumped Mini-Column Associative Knowledge Graphs", *IEEE Symposium Series on Computational Intelligence*, Honolulu, Hawaii, USA, Nov. 27- Dec.1.

Basawaraj, J.A. Starzyk, and A. Horzyk (2019), "Episodic Memory in a Minicolumn Associative Knowledge Graph", *IEEE Trans on Neural networks and Learning Systems*, vol. 30, no.11, Nov., 3505-3516.

Beilby, James K., ed. (2002). Naturalism Defeated?: Essays on Plantinga's Evolutionary Argument Against Naturalism. Cornell University Press. ISBN 0-8014-8763-3.

Benjamin et al. (2014), "Neurogrid: A Mixed-Analog-Digital Multichip System for Large-Scale Neural Simulations", vol. 102, no. 5, May 2014, Proceedings of the IEEE.

Benson D.L., Colman D.R., Huntley G.W., (2001) Molecules, maps and synapse specificity, *Nat. Rev. Neurosci.*, vol. 2, 899—909.

Bernat J. L. (2009) *Chronic Consciousness Disorders,* Annual Review of Medicine, Vol. 60: 381-392.

Beverly, J.; (2009), Nelson's Illustrated Guide to Religions: A Comprehensive Introduction to the Religions of the World, Thomas Nelson Inc.

Blackmore S. (2009) *"Ten Zen Questions."* Oneworld Publications, Oxford 2009. **ISBN**: 978-1-85168-642-1

Blanke O, Ortigue S, Landis T, Seeck M. (2002), Stimulating illusory own-body perceptions. Sep 19; 419(6904), 269-270.

Blickhan, R., Wagner, H., and Seyfarth, A. (2003). Brain or muscles? Rec. Res. Devel. Biomechanics, 1, 215-245.

Block, N. (1980). *Functionalism*, Readings in philosophy of psychology. Cambridge, Mass.: Harvard University Press.

Block, N. (1995). On a confusion about a function of consciousness. Behavioral and Brain Sciences, 18(2), 227–247.

Block, N. (2007). Consciousness, accessibility, and the mesh between psychology and neuroscience. Behavioral and Brain Sciences, 30(5–6), 481–548.

Blue Brain Project – A Swiss Brain Initiative, https://www.epfl.ch/research/domains/bluebrain/

Blumenfeld H., et al. (2004) *Positive and negative network correlations in temporal lobe epilepsy.* Cereb. Cort. 14: 892–902.

Bostrom N., (2014), *Superintelligence: Paths, Dangers, Strategies,* Oxford University Press.

Bostrom N., (2005), In defense of posthuman dignity, *Bioethics*, vol. 19, no. 3, 202-214.

Brodziak A., (2013), A current model of neural circuitry active in forming mental images". Med Sci Monit. Dec 12;19:1146-58. doi: 10.12659/MSM.889587.

Brooks R.A., (1991). Intelligence without reason, Proc. 12th Int. Joint Conf. on Artificial Intelligence, 569-595, Sydney, Australia.

Brożek B., M. Hohol M., (2014), *Umysł Matematyczny*, Copernicus Center Press, Kraków.

Bunzeck N. Doeller C.F. Dolan R.J. Duzel E. (2012) Contextual Interaction Between Novelty and Reward Processing Within the Mesolimbic System, *Human Brain Mapping* 33: 1309–1324.

Burton R.A. (2008), *On Being Certain. Believing You Are Right Even When You're Not*, Macmillan Publishers/St. Martin's Press, New York City, 2008

Caporale N, Dan Y. (2008), Spike timing-dependent plasticity: a Hebbian learning rule. *Annual Rev Neurosci.*;31:25–46.

Carpenter G.A., S. Grossberg, (2003), Adaptive resonance theory, in *The Handbook of Brain Theory and Neural Networks*, M. Arbib (Ed.), MIT Press, Cambridge, MA, 87–90.

Cassimatis N. and L. Nicholas, (2002), *Polyscheme: A Cognitive Architecture for Integrating Multiple Representation and Inference Schemes*, MIT Ph.D. Dissertation.

Chalmers, D. (1995), Facing up to the problem of consciousness. *J. Conscious. Stud*. 2, 200–219.

Chalmers, D. (1996), *The Conscious Mind: In Search of a Fundamental Theory.* Oxford: Oxford University Press.

Chalmers, D. (1998), On the search for the neural correlate of consciousness, in *Toward a Science of Consciousness II: The Second Tucson Discussions and Debates,* eds S. Hameroff, A. Kaszniak, and A. Scott (Cambridge: MIT Press), 219–230.

Chalmers, D.J. (1998), "What is a neural correlate of consciousness?", in Metzinger, Thomas (ed.), Neural Correlates of Consciousness: Empirical and Conceptual Questions, MIT Press.

Choi, H.; Smith, S. M.; Otani, H.; Widner, R. JR (2005). "Incubation and the resolution of tip-of-the-tongue states". The Journal of General Psychology. 132 (4): 365–376.

Churchland, P. (1981). "Eliminative Materialism and the Propositional Attitudes". Journal of Philosophy. ***78*** *(2, February): 67–90.*

Clark, A. (1997), *Being There.* Cambridge, MA: MIT Press.

Clark A. (2005), Painfulness is not a quale. In M. Aydede (ed.) Pain: New Essays on its Nature and the Methodology of its Study, Cambridge MA. MIT Press. 177-197.

Coggan J.S., et al. (2005), Evidence for ectopic neurotransmission at a neuronal synapse. Ellisman M. H., Sejnowski T. J. *Science* 309, 446–451.

Costa R.P, Sjöström P.J. and van Rossum M.C.W, (2013), **Probabilistic Inference of Short-Term Synaptic Plasticity in Neocortical Microcircuits** - *Frontiers in Computational Neuroscience* 7:75, 2013.

Cowan N., (1995), *Attention and memory: an integrated framework*. Oxford University Press.

Cowan N., (2008), What are the differences between long-term, short-term, and working memory? In W.S. Sossin, et al. (Eds.), *Progress in Brain Research*: Vol. 169. Essence of Memory (323-338), Amsterdam: Elsevier B.V.

Craig (2003), "A New View of Pain as a Homeostatic Emotion", Trends in Neurosciences. 26(6): 303-307.

Crane T. (2003)," The Intentional Structure of Consciousness", From: Consciousness: New Philosophical Perspectives ed. A. Jokic and Q. Smith Oxford and New York 2003: Oxford University Press.

Crick F. and Koch C. (1990), Towards a neurobiological theory of consciousness. Seminars in Neuroscience, vol. 2, 263–275.

Crick F. (1994), The astonishing hypothesis: the scientific search for the soul. Macmillan Reference USA. ISBN 978-0-684-19431-8.

Cui, Y., Surpur, C., Ahmad, S., and Hawkins, J. (2015), Continuous online sequence learning with an unsupervised neural network model. arXiv.1512.05463 [cs.NE].

Cutter B., Tye M. (2011), Tracking Representationalism and the Painfulness of Pain, Philosophical Issues 21(1): 90-109.

Damásio A.R. (2000), *The Feeling of What Happens*, Harcourt Brace, New York, 2000

Damásio A.R. (2003), *Descartes' Error. Emotion, Reason and the Human Brain.* Published September 27th 2005 by Penguin Books Ltd (first published 1994)

Damásio A.R. (2018), *Strange Order of Things. Life, Filing and the Making of Cultures.* Pantheon February 6, 2018.

Davidson D. (1980), *Mental Events*, in Essays on Action and Events. Oxford University Press 1980

Davidson D. (1993), *Thinking Causes*, in Mental Causation ed. Heil & Mele, Oxford: Claderon 1993.

Dawkins, M.S. (1993), *Through our eyes only? The Search for Animal Consciousness.* Oxford: Oxford University Press. ISBN 9780198503200.

Deecke L., (2012), There Are Conscious and Unconscious Agendas in the Brain and Both Are Important-Our Will Can Be Conscious as Well as Unconscious, *Brain Sci.* 2, 405-420.

Dehaene S.*, Sergent C. and* Changeux J.P. *(2003), A neuronal network model linking subjective reports and objective physiological data during conscious perception. Proc. Natl. Acad. Sci. USA 100: 8520–5.*

Dehaene (2014), S. *Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts*, Viking.

Dehaene S., Changeux J.P. (2011), *Experimental and theoretical approaches to conscious processing*. Neuron. 2011 Apr 28; 70(2): 200-27.

Dehaene S., Changeux J.P., Nadal J.P. (1987), Neural networks that learn temporal sequences by selection. Proc Natl Acad Sci U S A. 1987 May;84(9):2727-31.

Deng J., et al. (2014), *Scalable multi-label annotation.* ACM Conference on Human factors in Computing (CHI), 2014.

Dennett, D. C. (1988), *Quining Qualia*, in Consciousness in Modern Science, eds A. Marcel and E. Bisiach (Oxford: Oxford University Press), 121–158.

Dennett, D. C. (1991), Consciousness explained (1st ed.). Boston: Little, Brown and Co.

Dennett, D. (2005), *Sweet Dreams: Philosophical Obstacles to a Science of Consciousness*; MIT Press: Cambridge, MA, USA.

Diamond A., (2013), Executive functions. *Annu Rev Psychol*. 64: 135–168.

Duch W., (2003), *"Neurokognitywna teoria świadomości"*. Studia z kognitywistyki i filozofii umysłu (red. W. Dziarnowska i A. Klawiter). *Tom. I, Subiektywność a świadomość*. Zysk i S-ka, Poznań, 133-154.

Eccles J.C. (1994), "How the Self Controls Its Brain", Springer Verlag.

Edelman, GM (1987) Neural Darwinism: The Theory of Neuronal Group Selection. New York: Basic Books, Inc.

Edelman, G. M. (2003) Naturalizing consciousness: a theoretical framework. Proc Natl Acad Sci U S A 100, 5520-4.

Edelman, G. M. & Tononi, G. (2000), *A universe of consciousness: how matter becomes imagination.* New York, NY: Basic Books.

Eftekhar A., Fullwood C., Morris N. (2014), "Capturing personality from Facebook photos and photo-related activities: How much exposure do you need?". Computers in Human Behavior. Vol, 37, August 2014, 162-170.

Engels F, (1962) *Anti-Dühring,* Moscow: Foreign Languages Publishing House, 1962.

Ericsson K.A., and W. Kintsch, (1995), Long-term working memory. *Psychological Review*. 102 (2): 211–245.

Eroglu C., Barres B. A. (2010), Regulation of synaptic connectivity by glia. *Nature*. 2010; 468(7321):223–231. doi: 10.1038/nature09612.

Esch T, Stefano G: (2011), The neurobiological link between compassion and love. Med Sci Monit, 17(3): RA65–75.

Fell J., Klaver P., Elger C.E., Fernandez G. (2002), The interaction of rhinal cortex and hippocampus in human declarative memory formation. *Rev. Neurosci*.; 13: 299–312.

Feldman, J., & Narayanan, S. (2004), Embodied meaning in a neural theory of language. Brain and language, 89(2), 385-392.

Fell J., et al. (2003), Is synchronized neuronal gamma activity relevant for selective attention? Brain Res. Brain Res. Rev. 42: 265–272.

Fields R. D, (2006), Beyond the neuron doctrine, *Scientific American Mind*, June/July 17, 20-27.

Fields R.D., (2011), Drugi mózg, Rewolucja w nauce i medycynie, Prószyński Media, Warszawa.

Fivush R: (2011), The development of autobiographical memory. Annu Rev Psychol, 2011; 62: 559–82.

Floreano D., Mondada F., Perez-Uribe A., Roggen D., Iida F., (ed.) ; Pfeifer, R. (ed.) ; Steels, L. (ed.) ; Kuniyoshi, Y. (ed.), Evolution of Embodied Intelligence, In: Embodied Artificial Intelligence, Berlin : Springer Verlag, LNCS Series, vol. 3139, 2004.

Fockert J.W. de, G. Rees, C.D. Frith, N. Lavie, (2001), The role of working memory in visual selective attention. *Science* 291 (5509), 1803-1806.

Fodor J.A. (1989),” Making Mind Matter More”, Philosophical Topics 17.

Freeman W.J. (2003), The wave packet: An action potential for the 21st century, *Journal of Integrative Neuroscience* 2, 3–30. doi:10.1142/S0219635203000214.

Freeman W. (2005), *Societies of Brains: A Study in the Neuroscience of Love and Hate*, Lawrence Erlbaum Associates, Hillsdale, NJ.

Fries P., Reynolds J.H., Rorie A.E., Desimone R. (2001), Modulation of oscillatory neuronal synchronization by selective visual attention. Science; 291: 1560–1563.

Fukazawa, Y., et al. (2003), Hippocampal LTP is accompanied by enhanced f-actin content within dendritic spine that is essential for late LTP maintenance in vivo. Neuron, 38:447-460.

Fukuda K, Vogel E.K. (July 2009), Human variation in overriding attentional capture. *The Journal of Neuroscience*. 29 (27): 8726–33.

Furlan, F., & Pereira, A. (2009), On the role of synchrony for neuron-astrocyte interactions and perceptual conscious processing. Journal of biological physics, 35 4, 465-80.

Gallagher, S. 2011, Interpretations of embodied cognition. In W. Tschacher and C. Bergomi (eds.), *The Implications of Embodiment: Cognition and Communication* (59-71).

Gallagher, S. 2013, A pattern theory of self. *Frontiers in Human Neuroscience* **7 (**443): 1-7.

Gallese, et al. *(1996), Action recognition in the premotor cortex. Brain **119** (2): 593–609.*

Gallese V., Goldman A., (1998), Mirror Neurons and the Simulation Theory of Mind-Reading *Trends in Cognitive Sciences*. 2,12.

Gallese V., (2009), Ucieleśniona symulacja: Od neuronów, po doświadczenie fenomenologiczne; Formy aktywności umysłu. Ujęcie kognitywistyczne, t.2: Ewolucja i złożone struktury poznawcze. Wydawnictwo Naukowe PWN, Warszawa, s.196.

Galus W.L., (2015 a), Architektura Świadomości. Część I: Logika i morfologia sieci neuronowej, *Roczniki Filozoficzne* . 63 .1 139-171.

Galus W.L., (2015 b), Architektura Świadomości. Część II: Struktura molekularna i biofizyka pamięci, *Roczniki Filozoficzne*. 63 . 2 237-261.

Galus W.L., (2015 c), Architektura Świadomości. Część III: Wola i sens istnienia, Roczniki Filozoficzne 63 . 4. 129 – 168.

Gelbard-Sagiv, H., et al. (2008), Internally generated reactivation of single neurons in human hippocampus during free recall, Science 322 (5898), 96-101.

Gerstner W, Kistler W. (2002a), *Spiking neuron models single neurons, populations, plasticity*. Cambridge, UK: Cambridge University Press.

Gerstner, W.; Kistler, W. (2002b), Mathematical Formulations of Hebbian Learning, Biological Cybernetics, vol. 87, num. 5-6, p. 404-415 Springer Verlag.

Goldman, A. and De Vignemont, F. (2009), Is social cognition embodied? *Trends in Cognitive Sciences* 13 (4): 154-59.

Graham J., J. A. Starzyk, and D. Jachyra, (2015a), Opportunistic Behavior in Motivated Learning Agents, *IEEE Trans on Neural networks and Learning Systems*, vol. 26, no. 8, 2015, 1735-1746.

Graham J., et al. (2015b), A Comparative Study between Motivated Learning and Reinforcement Learning, *The Int. Joint Conf. on Neural Networks*, Killarney, Ireland, July.

Griffiths, P. E. and Scarantino, A. (2005), Emotions in the wild: The situated perspective on emotion. *Cambridge Handbook of Situated Cognition*.

Grossberg, S. (1982), Studies of Mind and Brain, D. Reidel, Dordrecht, Holland.

Grossberg, S., and Levine, D.S. (1987), Neural dynamics of attentionally modulated Pavlovian conditioning: blocking, inter-stimulus interval, and secondary reinforcement. Psychobiology, 15(3), 195–240.

Gruber T., Tsivilis D., Montaldi D., Muller M.M. (2004), Induced gamma band responses: an early marker of memory encoding and retrieval. Neuroreport; 15: 1837–1841.

Haikonen P.O., (2003), *The Cognitive Approach to Conscious Machines,* Exeter, UK: Imprint Academic.

Haikonen P. O. (2007), Robot brains; circuits and systems for conscious machines. Wiley 2007.

Haikonen P.O. (2012), *Consciousness and Robot Sentience*, World Scientific.

Hameroff, Stuart (1987), *Ultimate Computing,* Elsevier 1987; ISBN 978-0444702838 .

Hameroff S. (2006), "*Consciousness, neurobiology and quantum mechanics,*" in: *Jack A. Tuszynski (5 September 2006), The Emerging Physics of Consciousness. Springer Science & Business Media. 192–251. ISBN 978-3-540-36723-9.*

Hanegraaff, W. J. (1996), New Age Religion and Western Culture. Esotericism in the mirror of Secular Thought, Leiden/New York/Koln: E.J. Brill.

Harman, G. (1999), The Intrinsic Quality of Experience. Reasoning, Meaning, and Mind, 1, 244-262.

Hawes N., et al. Brenner (2007), "Towards an Integrated Robot with Multiple Cognitive Functions ", AAAI 7, 1548-1553.

Hawkins J., S. Ahmad, and D. Dubinsky, (2011), Cortical learning algorithm and hierarchical temporal memory, 2011, *Numenta Whitepaper*.

Hawking S. and L. Mlodinow, (2010), *The Grand Design*, Bantam, 208.

Haykin S., (1994), *Neural networks: A comprehensive foundation*.

Heller M. (2016), Filozofia Nauki, wyd. III Petrus, str.21, Kraków 2016.

Hochreiter S. and J. Schmidhuber, (1997), Long short-term memory, *Neural Computation*, vol. 9, issue 8, 1997, 1735-1780.

Holland O., et al. (2013), Real and apparent biological inspiration in cognitive architectures, Biologically Inspired Cognitive Architectures, vol. 3, Jan., 105–116.

Horzyk A. (2013), *Sztuczne systemy skojarzeniowe i asocjacyjna sztuczna inteligencja*, Akademicka Oficyna Wydawnicza.

Horzyk A., J.A. Starzyk, and Basawaraj (2016), Emergent creativity in declarative memories. *IEEE Symposium Series on Computational Intelligence*, DOI. 10.1109/SSCI.2016.7850029, Athens, Greece.

Horzyk A., J. A. Starzyk, J. Graham, (2017), "Integration of Semantic and Episodic Memories", IEEE Trans on Neural networks and Learning Systems, vol. 28, no. 12, Dec., 3084-3095, DOI: 10.1109/TNNLS.2017.2728203.

Horzyk A. (2017), *Deep Associative Semantic Neural Graphs for Knowledge Representation and Fast Data Exploration*, Proc. of KEOD 2017, Scietepress Digital Library, 67-79.

Iida F., R. Pfeifer, L. Steels, Y. Kuniyoshi. (2004), Evolution of Embodied Intelligence, In: *Embodied Artificial Intelligence*, Berlin: Springer Verlag, LNCS Series, vol. 3139.

Izhikevich EM. (2001), Resonate-and-fire neurons, Neural Network,14:883–94.

Jackson, Frank. (1986), "What Mary Didn't Know", *Journal of Philosophy*, 83, 291-295.

Jackson, F. (2003), "Mind and Illusion", in *Minds and Persons: Royal Institute of Philosophy Supplement 53*, ed. by Anthony O'Hear, Cambridge: Cambridge University Press, 251-271.

Jackson, F. (1986), "What Mary Didn't Know", *Journal of Philosophy*, 83, 291-5.

Jackson, F. (2003), "Mind and Illusion", in *Minds and Persons: Royal Institute of Philosophy Supplement 53*, ed. by Anthony O'Hear, Cambridge: Cambridge University Press, 251-271.

*James, William (1890). "*The principles of psychology*". New York, 1.01.1890.*

Johnson-Laird P.N. (1986) "Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness" (Cognitive Science Series), Harvard University Press 1986.

Jung C.G., (1921), Psychologishe Typen, Rascher Verlag, English version Psychological Types, Collected Works of C.G. Jung, vol. 6, Princeton, N.J.: Princeton University Press.

Kahneman D. (2011). Thinking Fast and Slow. Farrar, Straus and Giroux. 2011

Kandel E.R., (2000), Cellular mechanisms of learning and the biological basis of individuality, in *Principles of Neural Science*, 4th ed. 2000, McGraw-Hill, New York. 1247—1279.

Kandel E.R., (2001), The molecular biology of memory storage: a dialog between genes and synapses, Biosci. Rep., Volume: 21, 2001, 565—611.

Kandel E. R. (2007), In search of memory: The emergence of a new science of mind. W. W. Norton & Company. ISBN 978-0393329377.

Kane M.J., R.W. Engle, (2002), The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective, *Psychonomic Bulletin & Review* 9 (4), 637-671.

Kerry J.J., Livesley J., and Vernon P.A. (1996), "Heritability of the Big Five Personality Dimensions and Their Facets: A Twin Study." *Journal of Personality*, vol. 64, no. 3, 577-592.

Kim Jaegwon (1998), *Mind in Physical World. An Essey on the Mind-Body Problem and Mental Causation.* Massachusetts Institute of Technology, 1998.

Kim Jaegwon (1996), *Philosophy of Mind* Westview Press 2005.

Klemm W. R. (2010) „Free will debates: Simple experiments are not so simple", Adv Cogn. Psychol. 2010; 6: 47–65. doi: 10.2478/v10053-008-0076-2.

Koch, C. (2004), The quest for consciousness: a neurobiological approach. Englewood, US-CO: Roberts & Company Publishers.

Koch C. and Hepp K. (2006), *Quantum mechanics and higher brain functions: Lessons from quantum* *computation and neurobiology.* Nature 440: 611–2.

Koch C., and N. Tsuchiya , (2007), "Attention and consciousness: two distinct brain processes", *Trends in Cognitive Sciences* 11 (1), 16-22.

Koch C. and G. Tononi, (2011), A Test for Cosciousness, *Scientific American*, vol. 304, no. 6, 26-29.

Kohonen T., (1982), "Self-Organized Formation of Topologically Correct Feature Maps". Biological Cybernetics. 43 (1): 59–69.

Kornhuber H. H., (1973), Neural Control of Input into Long Term Memory: Limbic System and Amnestic Syndrome in Man, Memory and Transfer of Information 1973, 1-22.

Kornhuber, H.H., (1984), Attention, readiness for action, and the stages of voluntary decision. Exp. Brain Res., 9, 420–429.

Kramer M.A. (1992), Autoassociative neural networks, *Computers & Chemical Engineering*, vol. 16, no. 4, April, 313-328.

Kripke S.A. (1980), *Naming and Necessity*, Cambidge MA, Harvard University Press.

A. Krizhevsky, I. Sutskever, and G. E. Hinton, (2012), "ImageNet classification with deep convolutional neural networks," in Advances in Neural Information Processing Systems 25, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. New York, USA, 1097–1105.

Ladd, A.; Anesko, M.; Phillips, J. R.; Meyers, K. (2010), Romanticism and Transcendentalism: 1800-1860, infoBase Publishing

Laird J.E., (2008), Extending the Soar cognitive architecture, in *Artificial General Intelligence 2008*, Memphis, TN: IOS Press, 224-235.

Lakoff G., Johnson M., (1999), *Philosophy in the Flesh. The Embodied Mind and It's Challenge to Western Thought*, Basic Book, New York.

Lakoff G., Johnson M., (2010), *Metafory w naszym życiu*, Altheia, Warszawa.

Lakoff G., Nunez R.E., (2000), Where Mathematics Comes from. How the Embodied Brain Brings Mathematics into Being? Basic Books, New York.

Langley P., (2005), An adaptive architecture for physical robots, in *Proc. 2005 IEEE/WIC/ACM Int. Conf. Intelligent Robot Technology*, Compiegne, France, 18-25.

Langley P. and D. Choi, (2006), A unified cognitive architecture for physical robots, in *Proc. 21st Nat. Conf. Artificial Intelligence,* Boston, MA: AAAI Press, 1469-1474.

Lewis-Peacock J.A., A.T. Drysdale, K. Oberauer, and B.R. Postle, (2012), Neural evidence for a distinction between short-term memory and the focus of attention. *Journal of Cognitive Neuroscience*, 24, 61–79.

Liao J., D.N. Beal, G.V. Lauder, and M.S. Triantafyllou, (2003), Fish exploiting vortices decrease muscle activity. *Science 302*, 1566-1569.

Libet, B., Gleason, C.A., Wright, E.W., Pearl, D.K. (1983), Time of conscious intention to act in relation to onset of cerebral activity, readiness potential. *Brain 106*, 623–642.

Libet, B. (1985), Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behav. Brain Sci. 8*, 529–566.

Lisman J.E., Idiart M.A. (1995), Storage of 7 +/– 2 short-term memories in oscillatory subcycles. Science 267: 1512–1515.

Liu X., et al. (2012), Optogenetic stimulation of a hippocampal engram activates fear memory recall, Nature 484, 381–385;19, April.

Luck, S. J., & Vogel, E. K. (1997), The capacity of visual working memory for features and conjunctions. *Nature*, 390, 279–281.

Makino H., Malinow R. (2009), AMPA receptor incorporation into synapses during LTP: the role of lateral movement and exocytosis. Neuron 64, 381–390.

Martin A: (2007), The representation of object concepts in the brain. Annu Rev Psychol, 58: 25–45.

Matei, S. A., Miller, C. C., Arns, L., Rauh, N., Hartman, C., & Bruno, R. (2007). VisiblePast: Learning and discovering in real and virtual space and time. First Monday vol. 12.

Matthews Gary (2000), Neurobiologia. Od cząsteczek i komórek do układów, Wyd. Lekarskie PZWL 2000, chapters 16, 17, 18 i 19.

McClelland J.L., Rumelhart D.E. (1988), Exploration in parallel distributing processing. Cambridge: Brandford Books, MIT Press.

McGinn, C. (1989), Can We Solve the Mind-Body Problem?, Mind, Vol. icviii, 391, July.

McGinn, C. (2004), Consciousness and its origins. Oxford University Press.

Mele, A.R. (2005), "Decisions, Intentions, Urges, and Free Will: Why Libet Has Not Shown What He Says He Has". In Explanation and Causation: Topics in Contemporary Philosophy; Campbell, J., O'Rourke, M., Shier, D., Eds.; MIT Press: Boston, MA, USA.

Mentré P., (2012), Water in the orchestration of the cell machinery. Some misunderstandings: A short review, J Biol Phys **38;** 13–26. doi:10.1007/s10867-011-9225-9.

Merolla P.A. et al. 2014, A million spiking-neuron integrated circuit with a scalable communication network and interface, *Science* 08, Aug 2014: Vol. 345, Issue 6197, pp. 668-673, DOI: 10.1126/science.1254642

Mesulam M: (2008), Representation, inference, and transcendent encoding in neurocognitive networks of the human brain. Ann Neurol, 64: 367–78.

Metcalfe J. & Shimamura A. P. (1994), Metacognition: Knowing about knowing.

Milner D. A., Goodale M. A. (2008), Mózg wzrokowy w działaniu. Tłumaczenie: G. Króliczak Wyd. Naukowe PWN, Seria: Biblioteka Psychologii Współczesnej.

Milstein J. N. Koch C., (2008), Dynamic Moment Analysis of the Extracellular Electric Field of a Biologically Realistic Spiking Neuron, *Neural Computation*, The MIT Press, Vol. 20, No. 8, Pages 2070-2084.

Minsky M. (1980), K-Lines: A Theory of Memory, *Cognitive Science* 4, 117-133

Minsky M. (1993), Book review: Allen Newell, Unified Theories of Cognition, Artificial intelligence in perspective. Cambridge, MA: The MIT Press, 59, 343-354.

Miyake A., and P. Shah, (Eds.) (1999), *Models of working memory. Mechanisms of active maintenance and executive control*. Cambridge University Press.

Nagel, T. (1974), "What Is It Like to Be a Bat?". *The Philosophical Review.* 83 (4): 435–450.

Natsoulas, T. (1978). Consciousness. American Psychologist, 33, 909–914.

Natsoulas, T. (1992). The concept of consciousness3: The awareness meaning. Journal for the Theory of Social Behavior, 22, 199–225.

Natsoulas, T. (1994). The concept of consciousness4: The reflective meaning. Journal for the Theory of Social Behavior, 24, 373–400.

Nevzat Tarhan,(2019) https://questionsonislam.com/article/consciousness ; available on-line April 2020.

Newberg, A., (2010), *Principles of neurotheology*, Ashgate Publishing, Surrey.

Newberg, A. & Waldman, M.R., (2010), How God changes your brain: Breakthrough findings from a leading neuroscientist, Ballantine Books, New York.

Newman, E.A. (2003), "New roles for astrocytes: regulation of synaptic transmission". Trends in Neurosciences. 26: 536–542.

Nguyen V. A., Starzyk J. A. , Goh W-B. , Jachyra D., (2012), "Neural Network Structure for Spatio-Temporal Long-Term Memory" IEEE Trans on Neural Networks and Learning Systems, vol. 23 no. 6, June, 971-983.

Nguyen V. A., Starzyk J. A. , Goh W-B., (2013), "A Spatio-temporal Long-term Memory Approach for Visual Place Recognition in Mobile Robotic Navigation" –Robotics and Autonomous Systems, Elsevier, vol. 61, no. 12, Dec. 1744–1758.

Noë A., (2004), "Action in perception", MIT Press.

Noë A., (2009), "Out of our heads: Why you are not your brain, and other lessons from the biology of consciousness", Macmillan.

Newman, E.A. (2003), "New roles for astrocytes: regulation of synaptic transmission". Trends in Neurosciences. 26: 536–542.

Nuxoll A.M., Laird J. E., (2004), A Cognitive Model of Episodic Memory Integrated With a General Cognitive Architecture, *International Conference on Cognitive Modeling* (AAAI).

Nuxoll A. M., Laird J. E., A (2007), Enhancing Intelligent Agents with Episodic Memory, Cognitive Systems Research, *Proceedings of the twenty-second AAAI conference on artificial intelligence.* Vancouver, BC.

Oberauer K., (2002), Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 411-421.

Oberauer, K. (2003), Selective attention to elements in working memory. *Experimental Psychology*, 50, 257-269.

Oberauer, K., Süss, H.-M., Wilhelm, O., & Wittmann, W.W. (2008), Which working memory functions predict intelligence? *Intelligence*, 36, 641-652.

Oberauer K., & E.B. Lange, (2009), Activation and binding in verbal working memory: A dual-process model for the recognition of nonwords. *Cognitive Psychology*, 58, 102–136.

Ortigue S., Bianchi-Demicheli F. (2011), Intention, false beliefs, and delusional jealousy: Insights into the right hemisphere from neurological patients and neuroimaging studies. Med Sci Monit, 17(1): RA1–11.

O'Shea M. (2005), „The Brain. A very short introduction". Oxford University Press 2005.

Oudeyer P.Y., F. Kaplan, V.V. Hafner, (2007), Intrinsic motivation systems for autonomous mental development, *IEEE Transactions on Evolutionary Computation*, 11 (2), 265-286.

Oudeyer, P.Y. (2010), Intrinsically motivated goal exploration for active motor learning in robots: A case study Baranes, A. ; The 2010 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS).

Pain. (2005), W: *Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Stanford, CA: CSLI Publications.

Parpura V., et al. (1994), Glutamate-mediated astrocyte-neuron signalling. *Nature*. 369(6483):744–747. doi: 10.1038/369744a0.

Penrose R. (1989), *The Emperor's New Mind* .Oxford University Press.

Penrose R. (1994), *Shadows of the Mind.* Oxford University Press.

Penrose, R. (1997), „Physics and the mind", M. Longair (Ed.), The large, the small and the human mind (93–143), Cambridge, England: Cambridge University Press.

Penrose R. (1998), "Quantum computation in brain microtubules? The Penrose-Hameroff `Orch OR' model of consciousness," Phil. Trans. R. Soc. Lond. A, 356, 1869-1896.

Perea, G. Araque A. (2005), "Properties of Synaptically Evoked Astrocyte Calcium Signal Reveal Synaptic Information Processing by Astrocytes". The Journal of Neuroscience. 16 March 2005; ***25*** *(9): 2192–2203.10.1523/jneurosci.3965-04.2005.*

Perea G., Yang A., Boyden E.S. and Sur M., (2014), Optogenetic astrocyte activation modulates response selectivity of visual cortex neurons in vivo, *Nat Commun* **5** (2014), 3262.

Pereira Jr. A. (2007), Astrocyte-trapped calcium ions: The hypothesis of a quantum-like conscious protectorate, *Quantum Biosystems* **2;** 80–92.

Pereira Jr. A. (2016), "The Complex Dynamics of Cations with Water and Proteins in Neuro--Astroglial Interactions", At: Lugano–Switzerland, 10/2016, Ordinal: IV, Conference: Biophysical aspects of complexity in health and disease, DOI: 10.13140/RG.2.2.10910.13128.

Pereira, Jr. A. (2017), Astroglial hydro-ionic waves guided by the extracellular matrix: An exploratory model. Journal of integrative neuroscience, 16 1, 57-72.

Pereira R., H. Sá *(draft 2020), "Kantian Nonconceptualism as a Solution to Mary's Puzzle", available at Internet page: https://www.academia.edu/35458379/ Nonconceptualism_as_a_Solution_to_Mary_s_Puzzle_.docx?auto=download*

Perlovsky L. I. (1998), *„Conundrum of Combinatorial Complexity"*, „IEEE Transactions on Patern Analysis and Machine Intelligence" vol. 20, No. 6, June.

Perlovsky L. I. (2006 a), "Toward Physics of the Mind: Concepts, Emotions, Consciousness, and Symbols." Phys. Life Rev. 3(1), 22-55.

Perlovsky L. I. (2006 b), „ The Knowledge Instinct", New York.

Perlovsky L. I. (2008), Sapience, Consciousness, and the Knowledge Instinct (Prolegomena to a Physical Theory), Toward Artificial Sapience 2008, 33-60

Pietrzykowski T., (2011), Nowa dyrektywa o doświadczeniach na zwierzętach – wątpliwości, wyzwania i szanse, Prawo i Medycyna, nr 2/32, s. 1.

Pitcher, G. 1970, Pain Perception. *The Philosophical Review* 79(3): 368–393.

Pfeifer R., and C. Scheier, (1999), Understanding Intelligence, MIT Press, Cambridge, MA.

Pfeifer R., F. Iida, G. Gómez, (2005), Japanese Scientific Monthly, Vol. 58 (2).

Pfeifer R., and J.C. Bongard, J. C., (2007), *How the body shapes the way we think: a new view of intelligence*, MIT Press, Cambridge.

Plantinga, A. (1993). Warrant and Proper Function. New York: Oxford University Press. doi:10.1093/0195078640.001.0001. ISBN 0-19-507864-0.

Poznanski R.R, Tuszynski J.A., Feinberg T.E. (2016), Biophysics of Consciousness: a Foundational Approach. WSPC (August 26, 2016) ISBN-13**:** 978-9814644259

Poznanski R.R. and others, (2019), Theorising how the brain encodes consciousness based on negentropic entanglement; April 2019, Journal of Integrative Neuroscience 18(1):1-10. DOI:10.31083/j.jin.2019.01.0105,

Pulvermueller F. (2001), Brain reflection of words and their meaning, Trends in Cognitive Sciences, 5.12.

Quiroga, R.Q., et al. (2005), Invariant visual representation by single neurons in the human brain, *Nature 435* (7045), 1102-1107.

Quiroga R. Q., Fried I., and Koch C., (2013), Brain Cells for Grandmother, Scientific American. com 31, Feb.

Ramachandran, V. S., and Hubbard, E. M. (2001), Synaesthesia – a window into perception, thought and language. J. Conscious. Stud. 8, 3–34.

Ramachandran V.S., (2011), The Tell-tale Brain: A Neuroscientist's Quest for What Makes Us Human, W. W. Norton & Company.

Ramachandran V.S., (2012), Neuronauka o podstawach człowieczeństwa. O czym mówi mózg?, Wydawnictwo Uniwersytetu Warszawskiego, Warszawa.

Reggia J.A. (2013), “The rise of machine consciousness: Studying consciousness with computational models,” *Neural Netw.*, vol. 44, 112–131, 2013.

Rizolatti G. et al. (1988), Functional organization of inferior area 6 in the macaque monkey. II. Area F5 and the control of distal movements. Experimental Brain Research. Experimentelle Hirnforschung. Experimentation Cerebrale. 71. p. 491-507.

Rizzolatti G. et al. (1996), Premotor cortex and the recognition of motor actions, Cognitive Brain Research 3. 131-141

Rosenberg G.H., (1996), Rethinking nature: A hard problem within the hard problem. Journal of Consciousness Studies 3:76-88.

Rosenberg G.H., (2004), A Place for Consciousness. Probing the Deep Structure of the Natural World. Oxford University Press.

Rosenthal D.M. (1997a), A Theory of Consciousness. in N Block, O Flanagan & G Guzeldere (eds) The Nature of Consciousness, 729-753. Cambridge MA: MIT Press.

Rosenthal David M.(1997b) „Phenomenal consciousness and what it's like”, *Behavioral and Brain Sciences* 20 (1):156--57

Russell S., (2019), “It’s not too soon to be warry of AI”, IEEE Spectrum, Oct., 47-51.

Schmidhuber Jürgen, (2003), Exploring the predictable, Advances in Evolutionary Computing, 579-612 Springer-Verlag, New York, Inc. New York, NY.

Schultz W. Dayan P. and Montague P. (1997), A neural substrate of prediction and reward, *Science*, vol. 275, 1593–1599.

Schultz W. Dickinson A., (2000), Neuronal coding of prediction errors, Annual review of neuroscience 23: 473–500.

Schwartz, J.H., Castellucci, V.F., and Kandel, E.R. (1971), Functioning of identified neurons and synapses in abdominal ganglion of Aplysia in absence of protein synthesis. J. Neurophysiology. 34, 939–953.

Schwartz, B.L.; Frazier, L.D.; Otani, H.; Widner, R. (2005), "Tip-of-the-tongue states and aging: Contrasting psycholinguistic and metacognitive perspectives". The Journal of General Psychology. 132 (4): 377–391.

Schwartz, B.L. (2010), The effects of emotion on tip-of-the-tongue states, Psychonomic Bulletin & Review, Feb., vol. 17, 82–87.

Schwartz B.L. and Metcalfe J. (2011), "Tip-of-the-tongue (TOT) states: retrieval, behavior, and experience" Memory & Cognition July, volume 39, Issue 5, 737–749.

Seager, William E., (1995), "Consciousness, information, and panpsychism", *Journal of Consciousness Studies*, 2(3):272–288.

Shapiro L.G. and G. C. Stockman (2001), *Computer Vision*. Prentice Hall.

Shigetomi E., Patel S., Khakh B. S. (2016), Probing the complexities of astrocyte calcium signaling. *Trends in Cell Biology*. 26(4):300–312.

Sieb R.A., (1990), A brain mechanism for attention, *Medical Hypotheses*, vol. 33, issue 3, 145-153.

Simons, D. J., Chabis, Ch. F. (1999), Gorillas in our midst: Sustained inattentional blindness for dynamic events. Perception, 28, 1059-1074.

Sloman A., M. Croucher, (1981), *Why robots will have emotions*, Sussex University.

Sporns, O., Tononi, G. & Edelman, GM (2000) Theoretical neuroanatomy: Relating anatomical and functional connectivity in graphs and cortical connection matrices. Cerebral Cortex 10, 127-141.

Starzyk J. A., (2008), *Motivation in Embodied Intelligence*, in Frontiers in Robotics, Automation and Control, Oct., 83-110.

Starzyk J.A. Prasad D.K. (2011), A Computational Model of Machine Consciousness, International Journal of Machine Consciousness, vol. 3, no. 2 255-281.

Starzyk J.A., (2011a), Motivated Learning for Computational Intelligence, in *Computational Modeling and Simulation of Intellect: Current State and Future Perspectives*, edited by B. Igelnik, IGI Publishing, ch.11, 265-292.

Starzyk J.A., (2011b), Mental Saccades in Control of Cognitive Process. *Int. Joint Conf. on Neural Networks*, San Jose, CA, July 31 - August 5.

Starzyk J.A., Graham J. T., Raif, P. and Tan A-H., (2012), Motivated Learning for Autonomous Robots Development, Cognitive Science Research, v.14, no.1, p.10(16) 10-25.

Starzyk J.A., and Basawaraj, (2014), "Memristor Crossbar Architecture for Synchronous Neural Networks", IEEE Trans. Circuits and Systems, Part I, vol. 61, no. 8, Aug., 2390-2401.

Starzyk J.A., J. Graham, A. Horzyk, (2016), Trust in Motivated Learning Agents, *IEEE Symposium Series on Computational Intelligence*, Athens, Greece, Dec. 6-9.

Starzyk J.A., J. Graham, L. Puzio, (2017), Needs, Pains, and Motivations in a Simulated Learning Agent, *IEEE Trans on Neural networks and Learning Systems*, vol. 28, no.11, 2528-2540.

Starzyk J.A., J. Graham, (2017), "MLECOG - Motivated Learning Embodied Cognitive Architecture" IEEE Systems Journal, vol. 11, no. 3, 1272-1283.

Starzyk J. A., Ł. Maciura, and A. Horzyk, (2019), „Associative Memories with Synaptic Delays" *IEEE Trans on Neural networks and Learning Systems,* vol. 30, no.12, Dec. DOI: 10.1109/TNNLS.2019.2921143.

Sun R., (2007), The importance of cognitive architectures: an analysis based on CLARION, *J. Experimental and Theoretical Artificial Intelligence*, vol. 19, no. 2, 159-193.

Sutton R.S. (1984), *Temporal Credit Assignment in Reinforcement Learning*. PhD thesis, University of Massachusetts, Amherst, MA.

Sutton R.S. and A.G. Barto, (1998), *Reinforcement Learning: An Introduction*, MIT Press, Cambridge, MA.

Taylor J.G. (2012), A Final Solution to the Mind-Body Problem. Journal of Mind Theory, vol.1 No.1 25-58.

Thernstrom M. (2010) "The Pain Chronocle.: Cures, Myths, Mysteries, Prayers, Diaries, Brain Scans, Healing, and the Science of Suffering", Farrar, Straus and Giroux. *2010* PB

Thompson, E. and Varela F.J. (2001), Radical Embodiment: Neural dynamics and consciousness. Trends in Cognitive Sciences 5 (10); 418-425.

Thrun S., (1995), Exploration in active learning, in *Handbook of Brain Science and Neural Networks*, M. Arbib, Ed. Cambridge, MA: MIT Press.

Todd J. J. and R. Marois (2004), Capacity limit of visual short-term memory in human posterior parietal cortex, Nature; vol. 428, 15 April.

Tolhurst, D.J., Smyth, D., Thompson, I.D., (2009), The sparseness of neuronal responses in ferret primary visual cortex, Journal of Neuroscience 29 (8), 2355-2370.

Tononi G. (2004), *An Information Integration Theory of Consciousness*, „BMC Neuroscience" 5(1), 42, doi: 10.1186/1471-2202-5-42.

Tononi G. (2008), *Consciousness as Integrated Information. A Provisional Manifesto*, „TheBiological Bulletin" 215(3), 216-242, doi: 10.2307/25470707.

Tononi G. (2010), *Information Integration. Its Relevance to Brain Function and Consciousness*, „Archives Italiennes de Biologie" 148(3), 299-322.

Tononi, G. Edelman, G. M. (1998), *Consciousness and complexity.* Science 282, 1846-51.

Tononi G., Koch C. (2014), *Consciousness. Here, There but Not Everywhere*, „Philosophical Transactions of the Royal Society of London B: Biological Sciences" 370(1668), 20140167, doi: 10.1098/rstb.2014.0167.

Tononi, G., Boly, M., Massimini, M. and Koch, C. (2016), "Integrated information theory: from consciousness to its physical substrate". Nature Reviews Neuroscience. Nature Publishing Group. 17 (5): 450–461.

Turing A.M. (1950), Computing Machinery and Intelligence, Mind LIX, no. 2236, 433-460.

Tye, M. (1995), Ten problems of consciousness: a representational theory of the phenomenal mind. Cambridge, Mass.: MIT Press.

Tye, M. (2000), Consciousness, color, and content. Cambridge, Mass. MIT Press.

Tye, K.M., et al. (2008), Rapid strengthening of thalamo-amygdala synapses mediates cue-reward learning. Nature 453, 1253–1257.

Tye, M. (2009), Consciousness Revisited: Materialism without Phenomenal Concepts, Cambridge, MA: MIT Press.

Ullman S. (1996), High level vision: object recognition and visual cognition, Cambridge Massachusetts: MIT Press.

Vadakkan K. I. (2010), Framework of consciousness from semblance of activity at functionally LINKed postsynaptic membranes. Frontiers in Psychology (Consciousness Research). Oct., vol. 1, Article 168, 2.

Vadakkan K. I. (2010) Framework of consciousness from semblance of activity at functionally LINKed postsynaptic membranes. Frontiers in Psychology (Consciousness Research) 10.

Vadakkan K. I. (2011), Processing semblances induced through inter-postsynaptic functional LINKs, presumed biological parallels of K-lines proposed for building artificial intelligence. Frontiers in Neuroengineering.

Vadakkan K. I. (2012), The nature of "internal sensations" of higher brain functions may be derived from the design rules for artificial machines that can produce them, Journal of Biological Engineering, 6-21.

Vadakkan K. I. (2015), The functional role of all postsynaptic potentials examined from a first-person frame of reference. Rev. Neurosci. Feb;27(2):159-84. doi: 10.1515/revneuro-2015-0036.

Vadakkan K. I. (2016), “A first-principle for the nervous system”, Nov. 12.

Varela F., Thompson E., Rosch E., (1992), "The Embodied Mind", November 13th. MIT Press. ISBN 0262720213.

Verkhratsky A., Nedergaard M. (2014), “Astroglial cradle in the life of the synapse”. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*. 2014; 369(1654) doi: 10.1098/rstb.2013.0595.20130595.

Verkhratsky A., et al. (2016), Astrocytes as secretory cells of the central nervous system: idiosyncrasies of vesicular secretion. *The EMBO Journal*; 35(3): 239–257. doi: 10.15252/embj.201592705.

Volterra A. (2002), The Tripartite Synapse: Glia in Synaptic Transmission. Chapter 13: Release of transmitters from glial cells.: Oxford University Press. 164–184.

Voss P., B. Goertzel, & C. Pennachin, eds., (2006), *Essentials of general intelligence Artificial General Intelligence*, Springer, ISBN 3-540-23733-X.

de'Waal F., (2014), Bonobo i ateista. W poszukiwaniu humanizmu wśród naczelnych, Copernicus Center Press, Kraków.

Walczak, A., et al. (2013), Novel Higher-Order Epigenetic Regulation of the *Bdnf* Gene upon Seizures, *The Journal of Neuroscience, 6 Feb., 33(6): 2507-2511.*

Wang W., B. Subagdja, A.-H. Tan, and J. A. Starzyk, (2012), Neural Modeling of Episodic Memory: Encoding, Retrieval, and Forgetting. *IEEE Trans. on Neural Networks*, vol. 23, no. 10, Oct., 1574 – 1586.

Weisberg J., van Turennout M., Martin A. (2007), A neural system for learning about object function. Cereb Cortex, 17: 513–21.

Weng J., J. McClelland, A. Pentland, O. Sporns, I. Stockman, M. Sur, and E. Thelen, (2001), Autonomous mental development by robots and animals, *Science*, vol. 291, 599–600.

Werndl Ch. (2009), „What Are the New Implications of Chaos for Unpredictability?" Brit.J.Phil.Sci 60(1009):1955-220.

Widner, R.L, Jr; Smith, S.M; Graziano, W.G. (1996), The effects of demand characteristics on the reporting of tip-of-the-tongue and feeling-of-knowing states" The American Journal of Psychology; Urbana Vol. 109, Iss. 4, (Winter 1996): 525.

Wierzbicka A. (1965), "O języku - dla wszystkich", Wiedza Powszechna, p.160.

Woo, S. E., Jebb, A. T., Tay, L., and Parrigon, S. (2018), Putting the "person" in the center: Review and synthesis of person-centered approaches and methods in organizational science. Organizational Research Methods, 21(4), 814-845.

Woolf C.J. (2010) „What is this thing called pain?" J Clin Invest. 2010;120(11):3742–3744. doi:10.1172/JCI45178.

Wyller, Truls. (2005), The place of pain in life. *Philosophy* 80.03: 385–393

Yen S-C, Baker J., and Gray C. M., (2007), Heterogeneity in the Responses of Adjacent Neurons to Natural Stimuli in Cat Striate Cortex, J. Neurophysiology 97: 1326-1341,

Young M.P. (2002), Connectional organisation and function in the macaque cerebral cortex. In Cortical Areas: Unity and Diversity, Schűz A. and Miller R. eds., 351-375. London: Taylor and Francis.

Zanto, T. P.; Gazzaley, A. (March 2009), Neural suppression of irrelevant information underlies optimal working memory performance. *The Journal of Neuroscience*. 29 (10): 3059–3066.

Collection of articles: (2004), Glial Neuronal Signaling, ed. Glenn Hatton, Vladimir Parpura, pp. 1 - 456.

Y. Zuo, et al., (2019), "All-optical neural network with nonlinear activation functions", Optica, vol. 6, no. 9, Sept.