# PHILOSOPHY OF INFORMATION, MIND, CONSCIOUSNESS AND WILL

*Lukasz Lazarz*

***Keywords: Philosophy, philosophy of mind, cognitive science, philosophy of artificial intelligence, philosophy of psychology, philosophy of economics, ethics, aesthetics, information, mind, consciousness, knowledge, emotions, impressions, free will, happiness, beauty, goodness, value.***

## *Abstract:*

*At the outset, I would like to explicitly state that the philosophical style of this work is inspired by the words of Stephen Hawking, spoken at the 2011 Google conference, which could serve as its motto: "Philosophy is dead." Of course, the author of these words likely didn't mean that the ability to ask questions, even the most general ones, has died. Rather, it probably meant that science, within its domains, has reached certain limits of knowledge that compel us to ask questions about what lies beyond them, using a more precise language to describe the world. "Classical" philosophy, in the face of progress in the exact and natural sciences, serves merely as a tool and sometimes as a conceptual framework for problems considered within the realm of metascience. One could say that the philosophy of nature replaces ontology, the philosophy of mind especially draws from cognitive science, and psychology or economics replaces eudaimonia and axiology. However, one could also argue that the reflection on science and philosophy has been intertwined since their beginnings, and good philosophy must simply be practiced in close connection with science.* ***In this work, in addition to preliminary ontological remarks related to the fundamental concept of information for this study, I would like to primarily discuss the problem of Mind, Consciousness, Will (and other causes of human behavior). In other words, in simple terms, apart from initial remarks about (I) what this World is (what it is made of), I would like to indicate (II) who we are from the perspective of contemporary cognitive science, how we perceive and experience the World, and (III) how we act and how we should act (or how it should be).***

## 0. Introduction:

*The emergence of sciences enabling a more scientific analysis of these fundamental problems was closely linked to the first industrial revolution, which began in the late 18th century. The energy crisis caused by a lack of wood forced the inhabitants of England to search for an alternative source of heat energy. Considerable reserves of easily accessible coal, located near rivers, turned out to be that source, with transportation proving to be cheap and easy. The so-called second industrial revolution occurred in the second half of the 19th and early 20th centuries. It was triggered by rapid scientific development, accompanied by the emergence of new technological solutions (such as the light bulb in 1879). In the field of biological sciences, the groundbreaking significance was attributed to the theories formulated by Charles R. Darwin. In his published scientific works, "On the Origin of Species" (1859) and "The Descent of Man" (1871), he argued that the plant and animal world emerged through the gradual development and transformation of simple organisms into increasingly complex and diverse forms. Albert Einstein (1905) presented the theory of relativity, unprecedentedly modifying the understanding of concepts such as space and time. This gave rise to the fields of quantum physics and quantum mechanics. In the realm of psychology and psychiatry, there was an increasing focus on studying humans using experimental methods, detached from the concept of the soul. Physiological psychology and behaviorism emerged. In the field of metamathematics, the works of Russell, Hilbert, Gödel, Church, and Turing highlighted the limitations of formalization in description. Alan Turing (1936), while in Cambridge, wrote his arguably most significant mathematical work, "On Computable Numbers," in which he introduced an abstract machine capable of executing programmed mathematical operations, or algorithms. Parallel to the logical problems, mathematicians also delved into information processing from a more practical standpoint, involving data transmission and compression. Claude Shannon and Warren Weaver developed the mathematical theory of information.*

*Simultaneously, engineering and computer technology were developing. The works of Babbage, Ada Lovelace, and Konrad Zuse led to the creation of the first computer prototypes. In the United States, the Atanasoff-Berry Computer (abbreviated as ABC) and the ENIAC (Electronic Numerical Integrator and Computer) developed by Eckert and Mauchly emerged. Instead of ordinary relays, they utilized electron tubes.*

*All these changes facilitated the emergence of cybernetics, which focused on describing biological systems as computational systems. This, in turn, paved the way for disciplines such as cognitive science, cognitive psychology, and artificial intelligence. In 1956, workshops on information theory took place at MIT, and workshops at Dartmouth were intended to foster the development of formalized forms of thinking (led by Claude Shannon, Marvin Minsky, and John McCarthy). Subsequently, advances in neurobiology and research on neuron models (Warren McCulloch, Walter Pitts) and parallel information processing systems (Donald Hebb, Frank Rosenblatt, John Hopfield) ultimately shaped the modern understanding of the science of the mind. In recent years, unprecedented progress in artificial intelligence technology has led to the widespread dissemination of these scientific findings, particularly through initiatives like DeepMind, AlphaGo, and ChatGPT, reaching a broad audience.*

# CONTENTS

# 1. *Information, i.e., what the World is made of?*

## 1.1. Matter

The concept of "matter" is one of the oldest categories used to rationally explain the world. The understanding of matter as the substance from which individual things are "built" is already indicated by the Latin etymology of the term "materia," which in one of its basic meanings referred to the material from which something can be made, the building material, the substance—in the physical sense of the word. This term is considered to be closest to the Greek term "hyle," whose usage in the meaning of "material" of material things was established in philosophy thanks to Aristotle.

The dispute over the scope of the concept of "matter" arose among physicists already in the era of the formation of physics as a science, i.e., in the 17th century. According to Descartes, matter is an extended entity in space, filling it entirely, in the absence of absolute vacuum, infinitely divisible, not only in the form of bodies (particles) but also as a subtle material that carries interactions between them.[1] Due to its completely speculative character for that time, the concept presented by Isaac Newton prevailed: the world is composed of bodies built of hard and rigid, indivisible particles, impenetrable (only one particle can occupy one place), extended in absolute space, and enduring in absolute time, capable of motion, endowed with inertia (inertial mass) and weight (ponderable mass), separated by absolute vacuum; interactions are transmitted at a distance (through vacuum, without the mediation of material bodies) at an infinite speed. [2] In this context, mass is the primary attribute of matter and the measure of its quantity (and in physics, possessing measurable characteristics determines the recognition of a given entity as real). This concept seemed to have a strong foundation in the strict laws of Newtonian physics (dynamics and the theory of gravity) and became the philosophical basis for the further development of classical mechanics and later, after heat was recognized as the result of particle motion (Herapath, Carnot), kinetic theory of gases and thermodynamics. The development of physics seemed to demonstrate the nonexistence of imponderables—objectively existing, extended entities without mass, such as caloric.

## 1.2. Energy

The concept of energy generally expresses activity. Originally, it did not only apply to physics but also to the human spirit and human will. Philosophically understood (from Greek "ἐνέργεια"), the term had long been used to denote action, act, entelechy, or force. In medieval philosophy, it was expressed with Latin terms: virtus or vis. Nowadays, "energy" is a concept associated with physics. It refers to the ability of a force to perform work. While the concepts of mass and force appear in Newton's Principia in a completely correct sense, synonymous with their modern meanings, the term "energy" does not appear there. However, the creators of mechanics were aware of the conservation of total energy in mechanical processes (Huygens' theory of the pendulum is implicitly based on the concept of energy integration). Descartes expressed the idea that in natural processes, a certain quantity should remain constant over time. He considered the product of mass and velocity as such and named it "quantity of motion." Leibniz expressed the same idea, but he considered the product of mass and the square of velocity as the quantity that remains constant over time, which he called "living force" (vis viva). This gave rise to the famous debate about the true measure of "quantity of

[1] Descartes, Meteorology, 1637, Principles of Philosophy, 1644

[2] Izaak Newton, Philosophiae naturalis principia mathematica 1687 , Optics 1704.

motion." Alongside the idea of constancy over time for a certain quantity, thanks to Leibniz, the idea that physical processes occur according to the principle of least action also developed. Difficulties with formulating the modern understanding of the principle of energy conservation could have been resolved with the development of the science of heat.

Mayer, Helmholtz, and Joule are widely regarded as the creators of the modern understanding of the conservation of energy. Joule determined the mechanical equivalent of heat, i.e., the amount of heat that can be obtained from a given amount of mechanical energy. Under the influence of philosophical currents seeking unified theories in the 19th century, the concept of energy began to play a fundamental role in the emerging science. It was demonstrated how one form of energy could be converted into another, and more and more forms of energy were recognized, including mechanical, thermal, chemical, acoustic, light, magnetic, electrical, etc. However, efforts were made to reduce this expanding classification to two basic forms: potential and kinetic energy. In the mid-19th century, the principle of energy conservation began to take two forms of expression: "in an isolated system, the total energy is constant" or "energy cannot be created or destroyed." In this way, physics took the form of a science not only about matter but also about energy. Furthermore, in Albert Einstein's Special Theory of Relativity, he demonstrated certain relationships between mass and energy in the form of the famous equation $E=mc^2$.

### 1.3. Information

The theory of information has its roots in statistical physics, specifically in the concept of entropy. Entropy is a measure of the degree of disorder in a system, and also a measure of our ignorance about it. On the other hand, information is always associated with reducing our ignorance. In 1948, Claude Shannon published the article "Mathematical Theory of Communication," in which he presented a mathematical model for transmitting information using communication means such as telegraphs or radio transmitters. He also introduced a measure of information: the entropy function H, which satisfies the mentioned conditions and is defined by the formula:

$$S = -k \sum p(i) \ln(p(i)).$$

The mathematical theory of communication does not address the structural features of information, issues related to the meaning of information, nor does it refer to any clearly defined definition of information. Its only connection to information arises from the belief that when we know the probability distribution of selecting elements from a certain set (referred to as the alphabet in the original context), we can determine a measure of something that can be interpreted as "produced" information (Shannon's terminology) by such a selection. In the case of entropy within the theory of information, we are solely concerned with quantitative properties.

The mathematical theory of communication (MTC) is not the only theory that quantitatively estimates the amount of information. Drawing on achievements in computer science, some have proposed alternative approaches to the problem of information quantity. Their research forms the basis of algorithmic information theory.

The key idea of this concept is the assumption that a theory that explains phenomenon X is akin to a computer program that computes X. Therefore, such a program must be somewhat smaller than the phenomenon it explains. The concept of information is based on this assertion and the results obtained in the study of complexity in computer science.

However, information is not merely "something in some quantity," but also "something about something."

In cybernetics, alongside the quantitative characteristic of information that relates to reducing uncertainty about predicted states of a system, there is also an attempt at a qualitative

description of information. According to Marian Mazur, information is the transverse transformation of messages in a control path, meaning the change from one distinguished state to another distinguished state at the same point in the control path; assigning these states to each other. Thus, information is formally defined as a relation between these states. The highlighted aspect in the definition of information is the relationship between states, the relational element between "something and something."

However, nither Shannon's quantitative approach nor Mazur's purely formal and qualitative approach resolve the semantic issue concerning the meaning of information, that is, the understanding of information as "something about something." The common characteristic of semantic information theories is defining the semantic content of a message within the framework of probability theory concepts. In the theories of Yehoshua Bar-Hillel and Rudolf Carnap, as well as Fred Dretske and the theory developed by Luciano Floridi, the key aspect in defining the concept of semantic information is the assessment of the cognitive value associated with information, which is inversely proportional to the probability of a specific state of affairs occurring. According to this principle, the informational value is greater when the probability of a specific state of affairs occurring from the set of possibilities is smaller. In Bar-Hillel and Carnap's theory, this value is closely related to linguistic communication and its position within a specific closed linguistic system. If information is "something" about another "something," then semantic information of sentence A will be defined as that other "something," which is the set of possible worlds excluded by A. Some argue that Bar-Hillel and Carnap's theory is not more "semantic" than quantitative theories or Mazur's theory. Instead of emphasizing the quantity of information or the syntactic/formal relationship between states, their concept of information describes the syntactic relationship with a formally defined set of possible worlds excluded by A.

In Dretske's conception and that of his successors, the cognitive value is modified by the subjective propert’ies of the recipient, and although the information carrier does not have to be linguistic in nature, the concept of information itself is narrowed down to true sentences. The informational content of a signal is defined as follows: For a subject x possessing basic knowledge K (about the possibilities existing at the source of information), the signal r carries information that s is F if and only if the conditional probability of s being F given r (and K) is equal to 1 (but given K alone, it is less than 1). According to Dretske, false statements are not considered information.

In addition to semantic theories of information that refer to qualitative representation, similar theories with a pragmatic character, based on decision theory, have also emerged. Information appears there as a stimulus for a specific decision ("something" is a stimulus for "something").

Unfortunately, the programmatic and symbolic constraints of these classical concepts (which define requirements regarding the relationship between two "somethings" or the nature of a specific "something"), as well as their close association with classical data processing methods, result in their very limited application.

In light of the remarkable achievements and development of cognitive sciences and artificial intelligence (artificial neural networks), information is increasingly viewed in the context of processing by such artificial systems. Information simply becomes "something" in relation to "something else," akin to the relationships between neuro-like elements. At such a rudimentary level, conversations about the semantic or pragmatic nature of information become quite sterile, as semantics and pragmatics become mere products of syntax, the relationships between successive network elements. It can be said that the **metaphysics of the contemporary concept of information places greater emphasis on the relationship between elements than on the elements themselves**.

Similar intuitions are provided by modern quantum physics. Quantum information is encoded in non-local connections between individual elements of a system, which have no counterpart in classical information theory. We say then that the system is in an entangled state.

## 2. Mind and consciousness, *that is who we are, how we perceive, and how we feel?*

### 2.1. Intelligent Mind, knowledge, thought, thinking, reasoning

***Soul***

Fundamentally, until the mid-20th century, the philosophy of mind had a transcendental, fundamental, and phenomenal character, and both in the realm of religion and philosophy, it referred to the concept of the soul.

Characteristic philosophical views on the relationship between the soul and the body include Descartes' views. Although Descartes already assumes that "the body is merely a statue and a machine," he also finds a place for the soul in the pineal gland. This particularly delicate form of subtle matter is set into motion by the vibrations of the soul. Bodily fluids take on these vibrations, which then move the organs, glands, and muscles. On the other hand, Baruch Spinoza proclaimed that there is only one substance (God), and the mind and body are two manifestations (properties) of that substance. Gottfried Leibniz, in turn, proclaimed that monads are endowed with two attributes: extension and mentality. The synchronization of monad actions is made possible by pre-established harmony. Everything is composed of monads, which is why everything is endowed with hproto-mentality. The phenomenal or transcendental nature of the mind was also a central theme in later philosophy (e.g., phenomenology) and psychology (e.g., phenomenology, analytical psychology of Jung).

***Calculation***

The development of cybernetics, informatics, artificial intelligence, cognitive science (neurobiology, cognitive psychology), and computer technology in the mid-20th century led to the emergence of many new disciplines and subsequently to a remarkably rapid development of our knowledge about the human mind. Particularly in the areas of its intelligence, knowledge representation, and information processing.

In 1960, Hilary Putnam published the work "Minds and Machines," defending machine functionalism, a philosophical concept that became the dominant trend in the philosophy of mind. Standard computational functionalism considers computational roles to be identified with specific causal roles in the cognitive system, and stimuli and behavior are referred to as "input" and "output." Over time, computational functionalism evolved into connectionism, which still views the phenomenon of an intelligent mind as a function of a computational system but one that operates in a way similar to how the human brain functions. In both cases, the reasoning/functioning of the mind also has a mechanical/physical nature (due to its computational/physical character).

Although the argument against the thesis of computational nature or the mechanistic nature of human reasoning is attributed to Lucas, many people, upon familiarizing themselves with Gödel's results, had a prior sense of a limitation concerning broadly understood machines, i.e., computers, robots, and their networks (such remarks were made by Kurt Gödel, Alan Turing, and Alonso Church). However, it was only in 1961, when technological progress consistently opened up hopes for its unlimited development, and pragmatically oriented fields of mathematics such as the mentioned

mathematical theory of information, cybernetics, cognitive science, linguistics, cognitive psychology, and artificial intelligence increasingly approximated the vision of a "digital human" as if "in defense of human uniqueness," that this philosophical argument was first described in somewhat detailed terms. Therefore, despite the fact that the text containing this argument is not very good and that the author is by no means one of the most prominent experts on Gödel's theorem, people usually refer to it as Lucas's argument.[3]
Lucas's argument can ultimately be reduced to an attempt to demonstrate the difference between a machine (a deterministic and symbolic system) and a human by indicating that the former is limited in its understanding while the latter is not. A rather disconcerting fact is that essentially this argument assumes what should actually be proven. Firstly, one can have doubts about the assumption that the current knowledge or experience of humans is in no way limited (the proven superiority of humans has been assumed). Secondly, if we already acknowledge the potential infinity of human knowledge, why not look more favorably upon the machine itself, which can also be subject to the laws of evolution by utilizing environmental diversity or "benefit" from quantum effects to avoid accusations of a finite nature of its existence/reasoning?

It must be honestly stated that among mathematical logicians, the view of the invalidity of Lucas's argument predominates, allegedly resulting from the works of Kurt Gödel, Alan Turing, and Alonso Church. This situation has been the case since 1960 when Hilary Putnam presented his critical remark, recognized by Boolos as "classic," even before Lucas's publication. Similar views can be found in the opinions of Quine, the analyses of Benacerraf and Wang, or in Putnam's review of the version of the argument developed by Penrose.[4]

***Embodiment***

Although arguments like Lucas's have been rebutted, the initial optimism among artificial intelligence researchers has diminished. The reason behind this was the increasing technical difficulties that hindered the reconstruction of human cognitive abilities using symbols.

The first cognitive architectures (simulating aspects of human cognition) were classical rule-based systems. They were based on the paradigm of transforming input symbols into output symbols, but these symbols were primarily associated with natural language sentences (hence the close ties between cognitive science and linguistics) or formal language, and the information itself was processed serially. In 1955-1956, Herbert Simon and Allen Newell developed a program called Logic Theorist for proving logical theorems, which searched for non-trivial proofs within the space of possible syntactic transformations. It is considered the first program in the field of artificial intelligence. The program discovered 38 theorems out of the first 52 theorems of Bertrand Russell's Principia Mathematica. The program also utilized heuristics in the process of effective proof search, limiting the potential search space. Other classical architectures in the field of artificial intelligence included the General Problem Solver (1957), developed under Newell's direction, SOAR (State, Operator, and Result) by John R. Anderson. Despite their undeniable contribution to understanding our cognitive processes, these architectures fell short of the lofty expectations of their most fervent enthusiasts. Problems in achieving the ultimate goal of building artificial intelligence equal to humans (General Artificial Intelligence) continued to multiply. Increasingly, it was argued that the vision of the nature of the human mind within the first classical paradigm of cognitive science (Good Old Fashioned Artificial Intelligence) was incomplete. For intelligence to exist, more than a single brain is certainly needed. Intelligence currently emerges from the interaction of living organisms with the environment. Therefore, Hubert Dreyfus introduced two philosophers into the discussion who had previously seemed to be in a very cognitive perspective: Martin Heidegger and Maurice

[3] John R. Lucas:// Mind, Machines and Gödel , Hybris Nr 8 (2009) Issn: 1689-4286
[4] S. Krajewski 2003: *Twierdzenie Gödla i jego interpretacje filozoficzne: od mechanicyzmu do postmodernizmu* (Wyd. Instytutu Filozofii i Socjologii PAN, ISBN 83-7388-017-8)

Merleau-Ponty (phenomenology). Dreyfus's main argument against the possibility of artificial thinking is the essential non-formalizability of the entirety of human experience. Furthermore, in 1969, a specific problem called the "frame problem" [5]emerged within the logical stream of research on artificial intelligence. In this stream, deductive systems are created to provide tools for representing the world, and through adequate representation, machines can make intelligent decisions. Knowledge representation is usually propositional and is based on some version of first-order logic. Facts in the world are expressed through theses constructed using axiomatic predicates. Based on a set of axioms describing fundamental regularities in the world (e.g., laws of mechanics) and types of actions of an intelligent agent (e.g., object manipulation), the "artificial intelligent" agent, with facts and action rules introduced into the database, is supposed to plan rational and intelligent decisions, and subsequently act based on them. However, it turns out that a significant amount of knowledge needs to be encoded in this database (e.g., ironing a plaid shirt does not make its sleeve longer). Specifically, in order for inferences to be made from an activated situation, the subsequent situations must be described in detail, especially stating what does not hold in a given situation. In normal situations, much is known about a given situation, what is normal, ordinary, or the same. Although it is widely accepted that the frame problem is not a logical or philosophical argument against computational functionalism for the same reasons that Lucas's theorem is not such an argument (just imagine an Eskimo in New York or a New Yorker at the North Pole), the technical difficulties associated with the efficiency of early artificial intelligence remained a fact.

To eliminate the frame problem, the development of logics that were closer to common-sense reasoning [6] began, while rapidly advancing neurophysiological research led to a new approach to data processing based on parallel computing. One of the foundations of early artificial intelligence development was the work from 1943 by neurophysiologist Warren McCulloch and mathematician Walter Pitts, who pointed out the similarities between the functioning of neurons and simple electrically realized logic circuits. McCulloch believed that the human brain is a computer that combines input data and compares them to a given threshold value. The mathematical structure described in that work can be seen as a direct precursor to today's artificial neural networks. In 1949, Donald Olding Hebb developed a training strategy for natural neurons, which became known as "Hebb's Learning Rule." Hebb suggested that learning and memory exercise are correlated with changes in the strength of connections between individual synapses. According to this rule, simultaneous activation of neurons leads to the reinforcement of synaptic connections between them ("cells that fire together, wire together").

Another important stage in the development of artificial neural networks was the work of Frank Rosenblatt on perceptrons. He and his colleagues managed to construct the first neural computer by implementing neural assumptions into appropriate "hardware" components in MIT laboratories. Although Rosenblatt's research on neural networks was sharply criticized by Marvin Minsky and Seymour Papert due to the flatness of these networks, it quickly became evident that the critics did not fully appreciate the significance of the new approach. [7] It was not until the 1980s that an effective learning method for multilayer perceptrons, such as backpropagation[8], was finally developed. The return to research on neural-like networks was influenced, among others, by the work of John Hopfield, who created a feedback network capable of solving fairly complex tasks (e.g., the traveling salesman problem with high computational complexity). Many other models of neural-like networks with various applications were also developed, including networks using fuzzy

[5] John McCarthy and Patrick J. Hayes SOME PHILOSOPHICAL PROBLEMS FROM THE STANDPOINT OF ARTIFICIAL INTELLIGENCE Computer Science Department Stanford University Stanford, CA 94305 jmc@cs.stanford.edu http://www-formal.stanford.edu/jmc/ 1969

[6] McCarthy himself developed a version of non-monotonic logic, being, within the calculus of situations model of common sense reasoning

[7] Minsky and Papet showed that a single-layer perceptron cannot perform disjunction operations separate. Although it was already known at that time that a two-layer perceptron could perform such an operation; they described the approach as sterile and stopped development on neural networks for a long time.

[8] Currently, everyone agrees that the strength of a network depends on how the nodes of the network are connected

logic and networks modeling associative memory. Various algorithms for their learning were also developed.

With the progress of science and technology, advocates of the new connectionist approach in cognitive science, based on neural-like networks, began to believe that they were ultimately capable of matching human intelligence. Computational functionalism started to be displaced from dominant positions by connectionist functionalism, also known as connectionism. With the successive spectacular successes of artificial intelligence and machine learning (particularly deep learning), such as the victory over Gary Kasparov in chess in 1997, the victory over Lee Sedol in the game of Go in 2016, and the presentation of ChatGPT in 2022, the thought that nothing stands in the way of generating intelligence equal to human intelligence, albeit not necessarily in the same way as nature did, is increasingly prevalent.

However, the prevailing belief in contemporary mind science that the mind is fundamentally determined by its computational function did not leave many thinkers at peace. As early as 1980, John Searle, a representative of naturalistic dualism, presented the Chinese Room Argument in his work titled "Minds, Brains, and Programs." The essence of the argument is to demonstrate that effectively simulating the mind by a machine is not equivalent to possessing it because performing specific tasks does not necessarily involve understanding them by the performer. It is commonly believed that Searle, who identified consciousness and mental events as biologically existing properties of the brain (and as causally effective properties), seemed to not fully grasp the complexity and possible consequences of the functioning of complex neural-like systems. Similar to the case of information, the **relationships between elements prove to be crucial for understanding both the semantics and the understanding itself.**

In light of the successes that artificial intelligence has achieved in the realm of thinking, the phenomenal aspects of the Mind have increasingly been discussed in the context of other phenomena of human psyche, such as emotions, and above all, experiences and finally **Consciousness**.

## 2.2. Emotions

***Embodied soul***

The problem of emotions in philosophy has never played a primary role. To some extent, reason, its activities, and its products were the privileged subject of philosophy (as well as its tool). "Knowledge rather than action, belief rather than emotions, intellect rather than will were the central topics of philosophical research."[9] The Greek word πάθος (pathos), derived from the verb pati, meaning "to experience," was associated with emotions. Similarly, the Latin word 'Passiō' is often translated as "feeling," "passion," but also "suffering" or "anguish." Another Latin word associated with emotions is 'e movere,' which means "mind in motion." Throughout the centuries, emotions were identified on one hand as an irrational element, a variable experience of the soul, contrasted with reason, and on the other hand, as a component strongly linked to the body. It was only with the emergence of empirical psychology in the 20th century that the status of emotions as an "embodied soul" was rationalized, with particular emphasis on embodiment.

The Standard Characterization of Emotions simplifies the typical emotion as: (i) very rapid, (ii) correlated with changes in the body state, (iii) involuntary behavior, (iv) a reaction, (v) to a perceptually identified, (vi) valenced (positively or negatively) object (state of affairs) occurring in the

[9] Kenny A. (1963), Action, Emotion and Will, London: Routledge and Kegan Paul.,

environment. Emotions are accompanied by a phenomenal aspect associated with the experience of an affective state.[10]

Classical theories of emotions primarily focused on physiological arousal combined with phenomenal experience rather than abstract knowledge or evaluation. The dispute between representatives of classical theories mainly revolved around establishing the weight or order of occurrence of the individual defining elements of emotions (James-Lange Theory, Cannon-Bard Theory, Activation Theory, Two-Factor Theory by Stanley Schachter and Jerome Singer).

***Calculation***

Cognitive theories of emotions encompass a rich range not only in terms of psychological theories but also philosophical ones (Errol Bedford, William Lyons, Martha Nussbaum, Robert Solomon, and many others). In contrast to earlier concepts focused on physiological arousal and related sensations, cognitive theories generally equate emotions with thoughts, beliefs, judgments, or their basic components. Supporters of different cognitive theories share three hypotheses: 1) conceptualization; 2) disembodiment; and 3) evaluation. [11]Consequently, emotions become, in a sense, cognitive-conceptual structures like thoughts (described in the previous chapter), while also involving judgment (positive or negative) in terms of fulfilling needs. Particularly significant within cognitive conceptions is the propositional solution developed within analytical philosophy, credited to Bertrand Russell. Propositional attitudes (also known as "judgmental attitudes") are intentional states of the subject, simple thoughts, including beliefs, desires, emotions, etc., expressed by the subject in language. "P claims that x," "P fears that y," "P loves Z," etc. In each of these attitudes, a verb, the clause "that," and a sentence of a certain content x can be distinguished. Applying this solution to emotions, it is assumed that emotional attitudes are primarily consciously and intentionally directed toward a certain object. In this conception, emotions are treated as essential elements of rational thinking (cognition), whose fundamental components, as mentioned earlier, are concepts and beliefs.[12] Therefore, there are no major obstacles for emotions to be accessible to machines in the process of information processing. Cognitive theories are criticized for the fact that the experiencer of emotions does not always need to possess knowledge of a particular object or hold specific beliefs to react emotionally. Secondly, they may experience emotions before recognizing something.

Individual cognitive theories (Richard Lazarus's theory, Nico-Frijda's theory, Keith Oatley and Philip Johnson-Laird's communicative theory) distribute the importance differently regarding specific elements of evaluation or cognitive processes.

***Embodiment***

Jesse Prinz's concept is an attempt to unify previous approaches, but it is based on the belief in the need to maintain the standard characteristics of emotions. On one hand, Prinz's concept sympathizes with James-Lange's physiological theory, while on the other hand, it also emphasizes the cognitive nature of emotions. Prinz assumes that emotions arise in the body – they are bodily changes. The primary function of emotions is to register (perceive) the aroused states of the body and evaluate them: emotions are embodied evaluations. To evaluate x is to represent x in the form of representation. Therefore, emotions are not only evaluations but also representations: "By saying that emotions are perceptions of bodily changes, I mean only that they are states in our sensory-motor systems that register changes in our bodies. [...] However, it will be helpful to distinguish 'registration' from 'representation'. [...] The mental state registers what reliably activates

[10] D. Wiener, Wprowadzenie do Emocje: domysły, pomysły i fakty w formy aktywności umysłu ,w emocje, percepcja, świadomość cz. 1., str. 1-11

[11] J. Prinz, Gut Reactions. cyt., s. 21-26

[12] W: Dąbrowski A. CZYM SĄ EMOCJE? PREZENTACJA WIELOSKŁADNIKOWEJ TEORII EMOCJI, „Analiza i egzystencja" 27 (2014) ISSN 1734-9923

it. Emotions clearly 'register' changes in the body, but there is still another question about what these states represent."[13]

Jaak Panksepp further develops Prinz's concept towards further embodiment. Panksepp's concept remains one of the most significant attempts to address emotional issues based on neurobiological, inspired by psychology, cognitive science, and philosophy. Panksepp believes in the autonomy of cognitive and emotional spheres. He believes that emotions constitute a distinct neurobiological kind in the mammalian brain. [14]Panksepp also proposed the distinction of the following basic emotional brain systems: SEEKING, RAGE, FEAR, LUST, CARE, PANIC/SEPARATION, and PLAY (Panksepp deliberately uses capital letters).

In the case of the cognitive aspect of emotions, similar to the process of thinking, it seems that the construction of appropriate neuro-like networks (specific embodiment of the computational system based on relationships between neuro-like elements) is sufficient for its effective reproduction. However, the question of the content of accompanying experiences (including affective ones) and the specific physiology accompanying this process remains open. In the case of emotions, which are closely connected to the body, this question becomes even more crucial. Would completely different reactions, different experiences (if we assume for a moment that such experiences are possible at all), and different experiences of affective states, for example, in artificial systems, allow us to perceive any emotional qualities at all? Do we limit the definition of emotional qualities only to those similar to ours? Embodied exactly the same as ours? Similarly? How similarly? Therefore, even more than in the case of thinking or intelligence, besides the shared cognitive aspect (where the matter is limited to performing specific computations), reflection on the importance of the qualitative element (and its embodiment) becomes particularly significant. These questions naturally lead to the discussion of impressions and consciousness, as concepts that are most entangled in phenomenological or embodied approaches (due to their elusive nature or that of their object).

## 2.3. Consciousness, sensations

### *Soul*

As mentioned earlier, until the mid-20th century, the philosophy of mind and philosophy of consciousness had a transcendental, fundamental, and phenomenal character, drawing on the concept of the soul in both religion and philosophy. However, in the mid-20th century, due to the development of cognitive science and artificial intelligence, the problem of thinking and its more closely related aspect, the mind, dominated the scientific debate, and the problem of consciousness as a strictly metaphysical problem ceased to capture the attention of its participants.

Sensory experience is the primary locus of consciousness. Some argue that without sensory experience, we would have no concept of consciousness.[15] Regardless of whether this is true, experiencing phenomena, appearance, sound, taste, and touch dominates our mental life. As Fred Dretske writes, "Completely removing this experience would make us what? Zombies?"[16]

---

[13] J. Prinz, Gut Reactions. cyt., s. 58

[14] Panksepp J, Emocje jako twory naturalne w mózgu ssaków. W: M. Lewis, J.M. Haviland-Jones. Psychologia emocji (str. 185-210), Panksepp J. Affective Neuroscience: the Foundations of Human and Animal Emotion. New York, NY: Oxford University Press. 1998 str. 46

[15] Marcel A.J.E. BISIACH, Phenomenal experience and functionalism. W: Marcel, Bisiach 1988, Consiusness in Conteporary Science. Oxford: Clarendon Press, str. 121-1581

[16] F. Dreske, Naturalizowanie umysłu, str. 19

Perception is one of the most important and well-known forms of mental activity. The study of this activity not only provides knowledge about the functioning of specialized systems for gathering information about the environment but also reveals the operation of the mind in its most fundamental connections to the world. The study of perceptual processes is a cornerstone of cognitive science. The goal of perception is for the perceiver to acquire information about objects, states, or detailed environmental features. This information is essential for the perceiver's survival and enables them to respond to changing environmental conditions.

Perception is a complex process. The initial stage involves gathering material for perceptual processing. The material is then subjected to further multi-stage processing. At subsequent stages of data processing, an additional component, namely the previously acquired knowledge of the perceiver, also comes into play. The involvement of this knowledge poses a challenge for contemporary perception researchers. This additional feedback between what is provided and what is already known disrupts the perception process as a unidirectional process. However, it can be assumed that the ultimate result of this process is a percept, a state of mind through which the mind "reflects" a specific fragment of reality. It is, therefore, its representation. The belief in its informational nature is widely accepted.

In addition to the problem of representation, the most difficult issue is the problem of subjective sensations or, more precisely, the problem of specific subjective qualities associated with sensations. This is an age-old problem presented in a new light by David Chalmers and discussed in the "Journal of Consciousness Studies."[17] Thus, the problem of consciousness has entered the domain of science. The Latin word "qualis" denotes a property detached from things possessing that property, hence the colloquial reference to the "problem of qualia."

When we imagine a large strawberry in front of us, we focus on its color. We recognize that it is red, and the sensory experience of redness accompanies the identification of this characteristic. We intuitively feel that the recognition and conscious experience of this color are two different processes. The former arises from the operation of a measuring device for recognizing colors in the form of the visual subsystem of the sensory system, starting from the cones in the retina of the eye and ending in the V4 field of the visual cortex. The experience, on the other hand, can only be said to coexist with the end result of these processes, namely the recognition of the red color. This subjective experience is associated with the second, phenomenological aspect of consciousness.

The existence of this subjective aspect of sensations (and also of emotions and even thinking itself) for many also constitutes a serious argument against the possibility of capturing the phenomenon of human consciousness solely in computational terms, even if embodied in neuro-like networks.

***Calculations***

For proponents of computational functionalism and connectionism, impressions were either an insignificant problem or impressions remained in a functional relationship to the computations themselves, primarily implemented in complex and scalable neuro-like networks. Embodiment in a neuro-like form becomes a crucial condition for establishing a qualitative identity between the impressions of the system and human impressions (including affective experiences). The secondary issue becomes both the complexity and specific structure of such a network, as well as its further embodiment related to physical relationships between organs and indirectly with the environment.

***Embodiment***

As I mentioned earlier (when discussing the issues of Mind, Knowledge, Thought, and Thinking, specifically their embodiment), in the 1970s and 1980s, the optimism among artificial

[17] David Chalmers,(1996), *The Conscious Mind*, New York: Oxford University Press, ISBN 9780195117899 ,

intelligence researchers diminished due to increasing technical difficulties that hindered the reconstruction of human cognitive abilities using symbols. Philosophers such as Hubert Dreyfus and John Searle argued that the main argument against the possibility of artificial thinking is the essential non-formalizability of the entirety of human experience.

Despite the emerging technical challenges, the classical cognitive approach to mind and consciousness evolved towards a connectionist perspective, considering them as computational machines (even with a specific structure). However, many still attempted to salvage the mystery of non-machinelike aspects within our minds and consciousness. Some, embracing scientific naturalism, tried to reduce the issue to a discussion of physical properties (Paul M. Churchland). Some naturalists referred to dualistic and phenomenalist models, where the subjective (phenomenal) aspect exists alongside the objective element (John Searle, David Chalmers). Roger Penrose, on the other hand, sought the essence of mind and consciousness in quantum effects, which sparked the imagination of many proponents of various fundamental theories. [18]It is fair to acknowledge, though, that this view is widely criticized by most representatives of cognitive sciences, as quantum effects do not seem to have significant relevance to the discussed brain functions.

Concurrently, within the field of cognitive science, a new direction inspired by Heidegger's concept of "being in the world" emerged. The search for the essence of mind and consciousness in the specific embodiment of the nervous system, particularly its interaction with the body and the environment, became a milestone in embodied cognition. Advocates of this approach included linguist George Lakoff and philosopher Mark Johnson, who opposed the ideas of one of the founding fathers of cognitive science, Noam Chomsky. Chomsky claimed that there is a special computational module in the human mind, operating according to abstract rules and responsible for language acquisition and linguistic competence. According to Lakoff and Johnson, "there is no such thing as a computational person (...). Real people have embodied minds, and their conceptual systems arise from and are shaped by their living bodies, and they have meaning because of them." In other words, the identity of embodiment was defined by rejecting the computational view of the mind.

While functionalism evolved from a purely computational standpoint to a connectionist one, incorporating the principle of embodiment in terms of specific elements of a neuro-like network and the network itself, it still maintained the metaphorical property of a computer. Embodied cognitive scientists, on the other hand, aimed to go much further, seeking the essence of mind and consciousness in the relationships between the nervous system and the body/environment (embodiment) or in more fundamental laws of nature (quantum effects).

Although it is difficult to deny that the body and the environment (what exists and is governed by laws) influence our mind and consciousness, metaphorically becoming part of them, it is also challenging to deny that if our brain could be temporarily removed from our body and kept alive in a vacuum, it would likely continue to have thoughts, emotions, and sensations similar to those experienced by all of us. Therefore, excluding the brain from the realm of thoughts, emotions, and sensations, even if it had a completely different body and/or environment, does not seem justified. Even in such a case, such a mind or consciousness would simply be unrecognizable to us.

Finally, I would like to emphasize that both functionalism (connectionism) and embodied naturalistic cognitive science ultimately reduce the phenomenon of Mind or Consciousness to **dynamic relationships between neuro-like elements in the former case and relationships with the body or the environment in the latter**. In both cases, it leads us to a **vision of a World where our Mind and Consciousness appear more as parts of a Whole rather than completely separate and autonomous entities. The world is deterministic or nondeterministic, and there is no room for any other autonomy within it.**

[18] R. Penrose, Wielki Umysł Cesarza,

# 3. Wartość, szczęście, dobro i piękno, *czyli jak postępujemy lub powinniśmy postępować (lub jaki świat powinien być)?*

## 3.1. Free will, self, instinct and intention

***Soul, free will***

The problem of free will is fundamental to determining the rules of human behavior. Undoubtedly, the problem of will also depends directly on an even more fundamental question concerning causality in physics as a whole.

The description of the world within classical, Newtonian physics is inherently deterministic, even if it includes suggestions of significant randomness (as proposed by Maxwell in his studies on the motion of particles in gases). Quantum mechanics, on the other hand, at first glance, appears to indicate something entirely different. In particular, the so-called Bell inequalities unambiguously demonstrate that at its foundations, the world is random. However, the fundamental determinism of the world can still be defended in non-local theories, i.e., those in which the possibility of faster-than-light information transfer exists. An example of such a theory is David Bohm's theory of hidden variables.

Regardless of whether the world is completely deterministic or indeterministic at its core, it is difficult to find justification for free will within the realm of physics. Determinism directly excludes free will. Despite attempts to link the existence of consciousness and free will to indeterministic quantum effects (by Roger Penrose, for example), it is widely believed that indeterminism (i.e., complete randomness) not only has nothing to do with free will but actually excludes it just like determinism does.

However, prior to the reflection on the results of the physical sciences, the existence of free will was widely accepted. Even René Descartes, who considered the body to be nothing more than a statue and a machine, ultimately finds a place for the soul, as well as its active aspect: free will, even in the pineal gland. This particularly delicate type of subtle matter is set into motion by the vibrations of the soul. The bodily fluids absorb these vibrations, which then move the organs, glands, and muscles.

***Calculations***

It was only with the advent of the computer metaphor in contemporary mind science that the existence of free will was widely questioned. Despite rather desperate attempts to ground it, including in quantum effects as mentioned earlier by Roger Penrose, it is widely acknowledged that proving its existence is nearly impossible, apart from directly appealing to the sense of it (as John Searle does). Furthermore, additional compelling arguments against its existence have emerged.

One of the most famous arguments against free will is the so-called Libet experiment.[19] In Benjamin Libet's original experiment, five participants were tasked with making spontaneous hand movements. The moment of movement was supposed to depend solely on them - Libet instructed the participants to move their wrists in the most spontaneous way possible. During the experiment, the participants stared at a clock consisting of a dial and a green dot that circled it every 2.5 seconds. Their additional task was to remember the position of the green dot when they decided to make the movement. Libet recorded the brain activity of the participants using EEG, aiming to demonstrate that conscious decision-making occurs before the neuronal activity associated with the movement. As expected, it turned out that the moment of conscious decision preceded the hand movement by 200 milliseconds. However, Libet failed to confirm the existence of free will. The readiness potential (RP), representing neuronal activity in both hemispheres of the brain (mainly in the supplementary motor area, primary motor cortex, and sensory cortex), appeared on average 550 milliseconds before the hand movement, and thus 350 milliseconds before the conscious decision. Therefore, the conscious decision required for the existence of free will was preceded by unconscious brain activity, leading to the seemingly straightforward conclusion: free will does not exist. In 2008, John-Dylan Haynes and his team conducted a similar experiment, but it was updated in practically every aspect. According to many researchers, the results of this experiment confirmed the conclusions of the Libet.

***Embodiment: self, instinct, intention***

Regardless of the debate about the reality of free will, there is a subjective sense of existence and agency, which can be described as an active aspect of consciousness. Without consciousness, there is no sense of subjectivity or agency, no structural features of self, no phenomenal or personal dimension. It seems that the authors who argue that there is no consciousness without someone who is conscious are correct. The egotic orientation is a structural characteristic of all varieties and developmental forms of consciousness (equipped with self-referential elements). Galen Strawson refers to this thesis as The Experience/Experiencer Thesis and expresses it through three logically equivalent statements: a) experience is impossible without an experiencer, b) experience is necessarily someone's experience, c) the subject of experience exists only when experience exists. Antonio Damasio writes in a similar vein: 'If the process of self were to collapse and cease entirely, the mind would lose its orientation, the ability to gather its parts. Thoughts would flow on their own, without an owner to acknowledge them. Our effectiveness in the real world would drop to zero, and external observers would consider us lost. How would we appear? Well, we would appear unconscious.' Developmental forms of self correspond to developmental phases of consciousness: a) proto-consciousness corresponds to proto-self, b) core consciousness corresponds to core self, and c) extended consciousness corresponds to extended self (autobiographical). Extensive empirical and theoretical research is focused on determining at what point in the development of individuals and species we can speak of a minimal subject or minimal self. It seems that wherever we encounter even the most elementary form of consciousness, it is consciousness assigned to a specific subject (I). Simple forms of non-reflective self-awareness and self correspondences accompany us from the first days of our existence. Bodily self, ecological self, proto-self, core self, and covert self are different designations for the primary, simple, and elementary forms of experiential subjectivity. On the other hand, the development of higher forms of self-awareness and self requires social interactions, experiencing other people's mental states, language, autobiographical memory, etc. As Damasio writes: 'The self is definitely not a thing; it is a dynamic process, maintained at a fairly stable level for most of the waking period but subject to smaller or larger changes throughout that time... The simplest level (proto-self) is the work of that part of the brain that represents the organism and consists of a set of images that describe relatively stable aspects of the body and generate spontaneous feelings of a living body (primary feelings). The second stage is the result of establishing a linkage between the organism (represented by the proto-self) and any part of the brain

[19] Libet B. (1985). „Unconscious cerebral initiative and the role of conscious will in voluntary action", The Behavioral and Brain Sciences, t. 8, s. 529-566.

representing an object of cognition. The result is the core self. At the third stage... the autobiographical self emerges, which is defined in terms of biographical knowledge encompassing both the past and the anticipated future. All three stages are carried out in distinct but coordinated workspaces in the brain.'

When a subject appears (describing its behavior), questions arise about the causes, motivations, or purposes/goals of such behavior. The most basic mechanism responsible for the direction of human behavior is instinct. Instinct (drive, urge) is a biopsychic disposition that causes an individual equipped with it to behave (or desire to behave) in a specific way, as well as "an innate ability to perform certain stereotypical, unlearned, more or less complex actions that are specific to a given species and essential for its survival." This concept is defined differently in the science of behavior and is commonly associated with intuition and behaviors that are not the result of rational and conscious deliberation.

In the case of a conscious mind, the concept of intention is used. Nowadays, when referring to intention, it is assumed to refer to the theory of objects of consciousness, which deals with the relationship between acts of consciousness[20] (or the subjects of those acts) and the objects of those acts. However, the issue of intentionality was already formulated in the Middle Ages, and the emergence of contemporary theories of intentionality is associated with the views of Franz Brentano. According to Brentano, a phenomenon is intentional if it is "directed towards something" or "relates to something." Of course, spatial and linguistic metaphors can be eliminated in favor of a literal use of the term "about," which leads to the following formulation: a phenomenon is intentional if it is "about" something.

Since the direction of our actions is conditioned by the evaluation of a specific state of affairs in a positive or negative dimension, considerations of positively or negatively "directed towards something" or "relating to something" intentions lead us to even deeper reflections on the evaluative (cognitive) aspect of emotions, related to happiness or psychological well-being.

## 3.2. Happiness, wellbeing, utility

### *Soul*

Quo vadis domine? Almost everyone would choose one of the following answers: (i) I am heading towards joy, happiness, or (ii) a given value. Undoubtedly, a person's fulfillment on their chosen path contributes to their psychological well-being. Throughout centuries, especially in the realm of religion, the concept of "the good life" has been associated with the most fundamentally understood values or even a transcendental dimension. Whether we are dealing with Hinduism, Buddhism, Islam, or Christianity in its monistic approach, ultimately boiling down to the notion of "Unity," or in a dualistic approach, to another specific value/entity with a different ontology (such as the soul) than the rest.

Ancient philosophy also linked psychological well-being to phenomenal and fundamental concepts. For Plato, Aristotle, and especially the Stoics, happiness (eudaimonia) had little to do with pleasure or any emotional experiences. What becomes important is what is worth our efforts, that which is in accordance with our nature (with the daimon or the "true Self"), and thereby allows us to

[20]Usually, we talk about intentional and unintentional behaviors. However, this distinction is not exclusive. For example, O'Shaughnessy (1998) refers to subintentional behavior, which, although not preceded by conscious intention, can be voluntarily controlled unlike a reflex. Examples of subintentional behavior include tapping one's foot to the rhythm of a song or moving one's lips while reading. On the other hand, Searle (1983) introduced the concept of unintentional actions, which align with intention but are carried out somewhat accidentally.

make the best use of our inherent potential. Happiness lies in an authentic life that emphasizes essential virtues and ensures self-realization, not necessarily a life of pleasure.

***Calculations***

In contrast, however, to the commonly prevailing belief until the mid-20th century regarding the existence of an individual soul representing the spiritual dimension of human beings, a different, more "earthly" view of the nature of happiness began to take shape quite early, even in ancient times. The hedonistic tradition (Greek: ἡδονή, hedone, "pleasure," "delight") represented the position that human action is motivated solely by the desire to obtain pleasure and avoid unpleasantness. According to this view, pleasure and delight are considered the highest good and the purpose of life, the primary motive for human behavior. Avoiding suffering and pain is the main condition for achieving happiness. The hedonistic approach has a long tradition. Reducing well-being to sensual pleasure was the fundamental thesis of Aristippus of Cyrene, a disciple of Socrates. This rather primitive, purely sensual hedonism was refined over time.

For utilitarians, the measure of happiness was the satisfaction of human desires, not limited to sensual desires. Similarly, for Jeremy Bentham and later utilitarians who saw the maximization of individual pleasures and the satisfaction of other individual needs as the primary condition for building a good society ("the greatest happiness for the greatest number"), the concept held true. Contemporary researchers in the field of quality of life, following in the footsteps of utilitarians, adopt a broad measure of subjective utility that takes into account both the balance of pleasure and pain and the satisfaction derived from pursuing various goals in any field.[21]

The realization of these goals was closely related to their utility. Therefore, the emergence of economics was of great importance for scientific reflection on hedonistic-utilitarian ideas. Utility is a fundamental concept in economics. Economics as a science emerged only in the second half of the 18th century, dealing with the issues of management (decisions, choices)[22], specifically the optimal allocation of limited resources in conditions typically characterized by incomplete information. In a situation of limited resources, participants in the economic game of the market must make specific choices. Needs are the factors that determine these decisions or at least influence them to a greater or lesser extent. The measure that determines a particular choice is its utility.

The development of utility theory contributed to a better understanding of decision-making, risk perception, price levels, and many other economic issues. At the same time, it is an area where many still-unresolved research problems have emerged, crucial for understanding the phenomena occurring in the world of economics and finance, such as the difficult-to-explain excess profits from the stock market compared to the bond market.[23]

To a large extent, the development of utility theory was influenced by subsequent paradoxes related to the understanding of this concept. Traces of reflection on the concept of utility can even be found in ancient times. Aristotle pointed out the paradox of water and diamonds. This paradox suggests that water, despite being essential for life, is inexpensive, while diamonds, which we can easily do without, are expensive. Adam Smith, one of the founders of economics, attempted to resolve this paradox by stating that it arises from the fact that value can be understood as utility, that is, the usefulness of a particular good, as well as its purchasing power or exchange value. Goods with high exchange value have low utility, and vice versa. Therefore, water, characterized by high utility,

[21] Diener E., Suh, E.M.Lucas, R.E. Smith, H.L. (1999). Subjective well being: three decades of progress. Psychological Biulettin, 125, 276-302, Czapliński J., Spotkanie dwóch tradycji ejudamonizmu i hedonizmu, Psychologia pozytywna, Warszawa 2008

[22] This understanding of economics originates from ancient Greece (5th-4th century BCE) with Xenophon and Aristotle (see, for example, [Marciniak, 1999]).

[23] Maryniak P. Ewolucja teorii użyteczności,

simultaneously has low exchange value, while diamonds, with low utility, have high exchange value.[24] However, the classical paradox most frequently mentioned in the literature is the so-called St. Petersburg paradox. According to the criterion formulated by Pascal, which is important for decision-making under risk, the principle of maximizing expected value is crucial. Expected value is defined as:

**EV = Σ pi V (oi)**

*where: EV - expected value, pi - the probability of the i-th event occurring, V (oi) - the value of the i-th event.*

However, this principle cannot be considered a rational criterion for decision-making under risk. This was pointed out by N. Bernoulli. In the aforementioned St. Petersburg paradox, he posed a significant question: how much should a potential participant pay for the opportunity to participate in a game of flipping a coin until tails appear (the game ends when tails appear)? The participant receives payouts according to the following rule: if the game ends on the (n+1)-th flip, the player receives 2^n ducats. As we can easily calculate, the expected value of this game is infinity. Therefore, according to Pascal's principle, the potential participant should be willing to pay an infinite amount to participate. However, it is known that no player would do so, and no one would accuse them of behaving irrationally. Nicolas' relative, Daniel Bernoulli, pointed out that maximizing expected utility cannot be a rational criterion for decision-making under risk. Instead, he proposed the principle of maximizing expected utility. In Pascal's formula, when evaluating the attractiveness of an action under conditions of uncertainty, one should multiply the probability not by the value of the consequences, but by their utility. This principle allows for an individual's attitude towards risk. However, this principle requires the decision-maker to maintain a high level of consistency in their preferences.

The formal solution to this problem was proposed by Daniel Bernoulli in his work published in 1738 titled "Specimen Theoriae novae de Mensura Sortis" (Exposition of a New Theory on the Measurement of Risk). In this work, Bernoulli postulates that a person faced with such a problem assesses its value not based on price but on utility. Bernoulli's work not only solved an important problem in probability theory but also laid the foundations for today's understanding of concepts such as utility and risk. However, during those times, Bernoulli's work had more to do with gambling than with emerging economics. Therefore, it was only in the 20th century that Bernoulli's concepts found their place in mainstream economics. Scholars such as Gossen, Jevons, Edgeworth, and Fisher further developed the reflection on the concept of utility, pointing out additional problems related to its measurement.

Lack of effective measurement of total utility, coupled with the absence of a need for its measurement in economic analysis, led economists to shift their focus from cardinal utility to ordinal utility. The principle of utility became the basis of economics, albeit in a different way than advocated by utilitarians. As a result of the ordinalist revolution, utility ceased to be a normative value. It was still assumed that individuals intuitively maximize utility in their actions. However, the inability to objectively measure utility in a cardinal sense meant that it was impossible to base the functioning of the economic system or economic policy on it.

It was only several decades later that the return to total utility occurred. In their work "Theory of Games and Economic Behavior," John von Neumann and Oskar Morgenstern laid the foundations for the branch of mathematics dealing with strategic action in conditions of conflict and cooperation,

[24] Jak słusznie jednak zauważa Stiegler (1950) argumenty Smitha miały bardziej charakter osądu niż logicznego dowodu. Smith nie uzasadnienił odwrotnej proporcjonalności wartości użytkowej i wymiennej dóbr uznając to po prostu za fakt. Wyjaśnienie tego paradoksu zostało przeprowadzone dopiero w drugiej połowie XIX wieku za sprawą tzw. rewolucji marginalistycznej.

namely game theory. They also presented the theory of expected utility. Von Neumann and Morgenstern observed that the banishment of cardinal utility by economists was likely premature, and the measurability problem might be similar in nature to the measurement of heat, which initially relied on intuitive sensations but was later resolved with the introduction of temperature as a specific measure. They demonstrated that if an individual's preferences satisfy certain conditions (completeness, transitivity, and continuity), then the measurement of utility is possible. Utility measurability was based on comparing differences in value associated with the choice of a second-order preferred good to the choice in a lottery between two extremely preferred goods (as the first and last in order). This way, the concept of expected value, introduced centuries earlier, was finally systematized. The introduction of mathematical rigor into economics was also crucial, permanently transforming its image.

It seemed natural, therefore, to expand the models of expected utility by incorporating into the analysis not only the subjective evaluation of the value of potential choice outcomes but also the subjective treatment of probabilities associated with those outcomes. The classic model that developed this idea is the subjective expected utility (SEU) model, the foundations of which can be found in Frank Ramsey's original work from 1929. Similarly, Bruno de Finetti independently pursued this line of thought in 1937. Both argued that subjective probabilities can be inferred from observations of human behavior. They proposed deriving the degree of certainty about the truth of a statement from the analysis of decision-makers' behavior when making bets and calculating this degree as a probability.

A comprehensive model of decision-making under uncertainty that takes into account subjective probabilities was presented in L. Savage's famous work, "The Foundation of Statistics" (1954). Savage synthesized the achievements of Ramsey and Finetti, as well as the expected utility model of von Neumann and Morgenstern. He derived a new analytical structure and necessary and sufficient conditions for the existence of subjective probabilities and their unique combination with utilities of available choice options. He characterized individual choice as maximizing subjectively expected utility.

It might have seemed that economists had achieved a complete model of human behavior. However, in reality, this theory proved to be inconsistent with empirical findings. In 1953 (a year before the first edition of "The Foundation of Statistics"), French economist Maurice Allais published a famous paradox in the journal Econometrica, presenting a decision situation that deviated from the von Neumann-Morgenstern model. This paradox demonstrated that decision-makers, when faced with choices between options with exactly the same expected utility, were inclined to select different options, which directly contradicted the assumed rationality. In the 1950s and beyond, psychologists described numerous deviations from the expected utility theory, reflecting human behaviors. However, they were unable to propose an alternative theory, and it became evident that the model needed to be flexible. They did it only Kahneman and Tversky (1979), formulating the theory of prospect. They first showed several main effects indicating how the EU model systematically fails as a descriptive model. Two of these effects concerning the nonlinear treatment of probabilities are a modification of the famous Allais paradox (1953).

It is worth noting that the issue of perspective has a lot in common with the problem of framing, which was mentioned in relation to the Mind and knowledge. Just as it is difficult to symbolically describe the functioning of the mind, it is also difficult to describe expected and cardinal utility. However, just as with mental representations (Mind and knowledge), these problems stem from the complexity of the mind and similarly do not exclude the possibility of increasingly accurate reproduction, including modeling the performance of specific choices.

***Embodiment (and spiritualization?)***

The development of economics and trade was obviously related to the birth of the industrial revolution. The accompanying rapid development of technology and science caused changes in how people perceived the world. Just as the vision of a computational human was unacceptable to many, the vision of a human whose sole goal was their own benefit was downright repulsive to them. These changes and the accompanying anxiety increasingly favored introspection. These changes eventually led to the emergence of a new science - psychology. The emergence of psychology can be seen in a broader perspective as historically progressing emancipation of the natural sciences from philosophy. The scientific methods that psychology applied, more or less rigorously, began to transcend the boundaries previously reserved for philosophy. Wilhelm Maximilian Wundt (1832-1920) and William James (1842-1910) are considered the fathers of psychology as a separate scientific discipline and academic field, distinct from philosophy. On the other hand, Sigmund Freud is the father of analytic psychology (psychoanalysis). Psychoanalysis is both a theory of the human mind and a therapeutic practice. It was initiated by Sigmund Freud between 1885 and 1939 and is still being developed by psychoanalysts worldwide. Psychoanalysis has four main areas of application: as a theory of mind functioning, as a method for treating psychological problems, as a research method, and as a way of looking at cultural and social phenomena. Freud clearly outlines the dynamics of the psychic apparatus from the perspective of the ID through the pleasure principle. Therefore, ego preservation can only occur under the sign of individual economy of libido. Psychoanalysis is evidently based on a hedonistic paradigm. However, this hedonism is marked by pessimism. The pleasure program is constantly threatened by failure, and if it is achieved at all, it is only in a very limited way.

To counter this rather pessimistic image, also in the context of emerging psychology passively supported by behaviorism and later cognitivism[25], analytical psychology of Carl Jung, character psychology, existential psychology, and humanistic psychology tried to find a solution. These approaches were focused on more fundamental and enduring values rather than simple pleasures and changing goals.

Positive psychology emerged at the end of the 20th century in the United States as both a protest against negative psychology and a reaction to the worsening social indicators in the most economically developed countries, especially in the United States. However, the scientific ambitions of positive psychology are also important. While positive psychology on one hand relates to humanistic psychology, showing connections with phenomenology, on the other hand, positive psychology completely dissociates itself from metaphysics and fundamentalism. The teleology of positive psychology is primarily the science of goals, not the science of values. The goals of positive psychology have nothing to do with the phenomenal description of some more fundamental layer of the mind (or reality). Positive psychology seeks to integrate two approaches to the issue of well-being: hedonistic, where each of us becomes the sole and ultimate judge in our own affairs, relying on personal feelings of pleasure-pain, fulfillment and dissatisfaction, good and evil; and eudaimonic, where humans do not always know what is good or bad for them in the longer term, lacking full knowledge of their nature, needs, and capabilities, and there are pleasures that contribute to a fuller utilization of human potential. Prominent representatives of positive psychology were previously mentioned David Kahneman and Amos Tversky.

The history of Kahneman and Tversky's collaboration dates back to the early 1970s. The first research program that brought them fame was a series of experimental studies on probabilistic reasoning known as heuristics and biases. This program was inspired by several surprising effects observed in the psychological laboratory. One of the most famous results of such experiments, which sparked many heated discussions and inspired a range of studies in cognitive science, was the so-called Linda problem. After hearing a description of a young woman with liberal views, participants claimed that it was more likely for her to be a feminist bank teller than a bank teller. This is an obvious violation of formal logic and set theory, as the set of feminist bank tellers is a subset of the set of bank tellers.

[25] more concerned with the processes of learning than with discovering or justifying fundamental values - which, however, I consistently do in the later part of this work from those perspectives.

Another surprising result was the opinion of student participants that in families with six children, consisting of three boys (B) and three girls (G), the birth sequence BBBGGG is less likely than BGBGBG, although according to probability theory, all such sequences are equally probable. Imperfect principles of probabilistic reasoning were called heuristics, and systematic errors were referred to as biases. An example of a heuristic is availability, where people estimate the probability of an event based on how easily instances of it come to mind. For example, people tend to believe that the number of homicides is higher than the number of suicides (homicide cases receive more media coverage and are therefore more memorable), even though the opposite is true. In 1982, Kahneman and Tversky, together with Paul Slovic, published a collection of articles titled "Judgment under Uncertainty: Heuristics and Biases," summarizing several years of research on heuristics and biases. "Heuristics and Biases" had a significant impact on our awareness of the existence of strong and systematic cognitive biases. Since 1982, it has been reissued unchanged multiple times by Cambridge University Press (in some years, even with two reprints). In 2002, edited by D. Kahneman, T. Gilovich, and D. Griffin, "Heuristics and Biases: The Psychology of Intuitive Judgments" was published, presenting a collection of studies inspired by the heuristics and biases program from the last 20 years.

The second famous research program that was likely the main reason for awarding the Nobel Prize to Kahneman was prospect theory. In 1979, an article by both scholars presenting a new theory of human behavior in risky situations was published in Econometrica. Prospect theory explained in a straightforward way many economic anomalies, which are behaviors inconsistent with economically understood rationality (economists consider behavior rational if an individual's choices are in line with the axioms of expected utility theory). Kahneman and Tversky's model assumed that decision-makers evaluate the potential outcomes of risky financial decisions not in terms of their impact on the decision-maker's total wealth but in relation to a reference point, which is typically zero gain/loss. People, following the principle of "a bird in the hand is worth two in the bush," are reluctant to take risks in the domain of gains but often take risks in the domain of losses—to avoid a smaller loss, they risk incurring a larger one, such as arguing with a police officer to avoid a fine. It was also observed that the sadness resulting from a loss of amount X appears to be more than twice as strong as the joy from receiving the same amount X, which was termed loss aversion. Highly improbable events are given disproportionate weight in the decision-making process, as well as events that are certain compared to those that are almost certain. A significant implication of prospect theory is that the formulation of the decision problem has a strong influence on the decision. For example, saying "this rescue plan allows us to save 100 out of 500 people" instead of "with this rescue plan, 400 out of 500 people will not survive." Kahneman and Tversky's model allows predicting and explaining the behavior of ordinary consumers, as well as experts (e.g., a doctor making a diagnosis), stock market investors, or government officials. In 1999, after Tversky's death, a collection of articles titled "Choices, Values and Frames" was published, which included over thirty works inspired by prospect theory.

The last comprehensive research project that Kahneman undertook after Tversky's death concerns the processes of evaluating pleasant and unpleasant experiences and is known as hedonic psychology. The way people remember pleasurable and unpleasurable experiences proves to be another challenge to common beliefs and economic postulates of rationality. It turns out, for example, that of two painful medical procedures, the one that lasts longer may, under certain conditions, be remembered as less unpleasant than a similar procedure that is shorter.

Thanks to Daniel Kahneman and Amos Tversky, we have systematic knowledge about the imperfections of our statistical inferences, both in everyday life and in the form of serious expertise. We can better explain the peculiar behaviors of ordinary consumers and large corporations. We have a better understanding of why people buy lottery tickets and why the risk premium in financial markets is so high. The research stream initiated by them will undoubtedly yield a series of intriguing discoveries of significant practical importance.

The dominant direction in the psychology of happiness is likely the hedonistic-utilitarian direction, as it is difficult not to perceive an obvious connection between happiness and pleasure or the fulfillment of goals in most cases. This direction has practical significance, especially for sciences like economics, because pleasure and goals can be measured somehow (according to a given criterion of utility), which is much more challenging when it comes to associating happiness with general values formulated in the most general way.

However, it is also difficult not to notice that these values do not necessarily oppose simple pleasures and goals; they can directly result from them. Evolutionary game theory, which emerged based on the game theory created by Morgenstern and von Neumann mentioned earlier, or the existence of mirror neurons[26], provides such insights. It is also challenging not to notice that the level of happiness ultimately depends simply on the conditioning of associations that trigger specific biochemical processes leading to positive experiences. Proper training of neural networks in the spirit of behavioral-cognitive psychotherapy is also a way to improve well-being. The same applies to meditation or contemplating specific values, often undertaken by religions and philosophies. Such inspiration to change certain associations can stem from a vision of the world and ourselves that is quite unambiguously derived from our already known knowledge: for example, "that **we are more a part of the world than autonomous beings, that the world is a unified whole, and what is common and unifies us is much more important than ourselves".**

### 3.3. Value, good, normativity i utility

***Soul***

Present in religions and philosophical systems, the world of values and goodness, even more than happiness, dissociated itself from bodily pleasure in an attempt to establish an ontology different from matter and the common world. As discussed earlier in a previous chapter, the ancient Greeks noticed that humans are composed of something more than just the body. Heraclitus observed that humans possess a soul, which is the cause of constant change. Following him, Plato claimed that fundamentally, humans are entirely souls that have been sent to the body as a punishment for their transgressions during their existence in the world of ideas (the real, true world; Plato considered everything associated with the body and the material world as a mere shadow of the world of ideas). He referred to the body as the soul's tomb, from which it can only escape through virtuous living. Initially, Plato's views on the separation of good and psychological happiness were quite radical, and the distinction between worldly good and ideal good was fundamental to him. However, he later acknowledged that real goods are a necessary step towards achieving ideal goods. Ultimately, Plato's ethics implied that worldly goods serve as a step towards ideal goods.

Plato's theory was refined by his student Aristotle, who argued that the soul and body are inseparable, together forming the essence of a human being. It is the soul that shapes the body, serving as its material form and giving life to it. The soul is the primary constitutive act and organizer of human bodily life. Aristotle, in turn, sought happiness and values (goodness) in eudaimonia. However, Aristotle's eudaimonia did not exhibit anti-naturalism. While moving happiness closer to values and distancing it from hedonistic pleasure, he still regarded both matter and soul as properties of one world, as well as the goals inherent in their nature. For humans, eudaimonia, or a rational life, represented such a goal. For ancient philosophers, a fundamentally happy life was synonymous with a valuable life, but even then, a direction emerged that sought happiness and a valuable life in much simpler pleasures.

[26] I will return to these examples in the section concerning Values, Goodness, and Beauty.

***Calculations***

Interestingly, unlike with Mind, Consciousness, or Happiness, before the 20th century, when science began to question our perception of the world and ourselves, in the case of Values and Goodness, science, particularly propositional logic (as a branch of mathematics), initially sided with the "traditional" view. The co-founders of the analytic philosophy movement, George Edward Moore and Bertrand Russell, opposed the idealism represented in the UK by neo-Hegelians Francis Herbert Bradley and John Ellis McTaggart. Analytic philosophy was later pursued by Ludwig Wittgenstein, the Vienna Circle (R. Carnap), ordinary language philosophers (Austin, Ryle, the later Wittgenstein), and American philosophers (Kripke, Putnam, Davidson, Searle). The source of analytic philosophy was the so-called linguistic turn, which occurred under the influence of the breakthrough in logic (the publication of "Principia Mathematica" by B. Russell and A. N. Whitehead). Analytic philosophy was characterized by treating language as a tool and subject of philosophy, precision in argumentation, aversion to hasty syntheses, and the abandonment of constructing a philosophical system in favor of detailed analysis of specific issues, intersubjective meaningfulness and controllability of results, and purely cognitive goals of philosophy. The key methods of analytic philosophy were classical conceptual analysis, logical analysis, linguistic analysis, and systematic analysis.

In 1903, in Principia Ethica, George Edward Moore put forward a consequential thesis: moral concepts, such as "good," cannot be defined in terms of natural concepts. He criticized some ethical theories for what he called the "naturalistic fallacy." As one of the founders of analytical philosophy, he believed that ethical problems could be solved through the analysis of language. According to him, understanding the meaning of a predicate is equivalent to knowing the property to which it refers. To define a given term, one must provide an elaborate, synonymous expression that can truly predicate the property referred to by that term. However, it turns out that there are concepts that cannot be defined. "Good" is one such concept. As evidence for this thesis, Moore presented the so-called "open-question test." This test can be reconstructed as follows: Let's assume that: 1) The predicate "good" is an analytical counterpart of the natural predicate "N." Therefore, by virtue of the definition of analyticity, we know that: 2) The meaning of the proposition "x is N" is identical to the meaning of the proposition "x is good." But if that were the case, then: 3) Someone who seriously asks the question "Is x, which is N, also good?" would commit a conceptual mistake. However, it does not seem that (3) is true because: 4) The question "Is x, which is N, also good?" is an open question - meaningful and not a confusion of concepts (e.g., one can sensibly ask, "Is an action x, which is desired by the majority of society, also good?"). Therefore: 5) It is not the case that the predicate "good" is an analytical counterpart of the predicate "N." Therefore: 6) By conceptual necessity, the property of being good cannot be identical to the property of being N. Consequently, Moore concluded that the property of goodness is self-existent, simple, and indefinable because it cannot be reduced to any complex of composite properties. It is also non-natural because it exists beyond time and space. Since scientific inquiry can only be concerned with natural properties, naturalists err by replacing the subject of ethics with the subject of one of the sciences (most commonly psychology or sociology). This critique opened a new chapter in the history of metaethics.

***Embodiment***

However, the development of technology and science in the 20th century, particularly the failures and limitations of classical rule-based systems, make us increasingly aware that such an approach is simply too simplistic. Moore, in fact, arbitrarily stated that the property of goodness is self-existent, simple, and indefinable because it cannot be reduced to any complex of composite properties. However, it may not be self-existent and simple, and it may result from specific processes of learning based on more rudimentary phenomena; it may be complex and defined in that way. Furthermore, it may be natural and exist within time and space (as long as we understand reality within their framework). Due to his analytical approach, Moore made significant simplifications and assumptions about the nature of the property of goodness. Even common understanding allows us to state that

good exists and is good because sometimes things are bad, and it is good (as we know, good precedes evil, and we specifically emphasize the importance of how things will be rather than how they are currently). Value, goodness, just like happiness or consciousness or mind, can be equally well reduced to the function of a specific brain. We can argue, of course, that these functions, in a sense, represent some real Values, like Plato's shadows, but whatever we want to say about them, we can only see them through the lens of these functions. We cannot escape the cognitive lens.

It is very interesting how game theory explains altruism, which is behavior that lowers an organism's own fitness for the sake of increasing the fitness of another player. In 1959, Robert Aumann showed that in an infinite series of such games, cooperation can be an equilibrium state. Then, in 1984, Robert Axelrod invited scholars from around the world to participate in a tournament for computer programs playing the iterated prisoner's dilemma. The submitted programs differed in complexity, initial behavior, reaction to the opponent's actions, etc. The results showed that in multiple rounds of play, selfish strategies yielded only very small average gains compared to more altruistic ones. Axelrod thus demonstrated the possibility of evolutionarily emerging altruistic behaviors from initially self-interested ones through natural selection.

A fashionable source of inspiration for researchers dealing with happiness or values is the discovery of so-called mirror neurons. The existence of these neurons is clearly associated with empathy and sympathy, which in turn contrast egoism with altruism. They allow the interpretation of observed behavior to take place in the recipient's brain through a specific simulation. Our brain enables us to automatically reproduce the behavior of another person, thereby enabling us to replicate (map) their behavior and feelings. Mirror neurons allow for mental "reading" and "re-enactment" of the behaviors of another brain. In humans, they likely also account for the ability to recognize others' nonverbal expressions of emotions and intentions. Originally a "tool" (serving to predict the behavior of another individual for one's own benefit), it ultimately became the source of empathetic behaviors related to many Values.

**Contemporary cognitive science is consistently monistic and naturalistic. One phenomenon emerges from another through fundamentally deterministic processes, or at least deterministic in a way that deterministic understanding is sufficient to explain the processes occurring in our minds. Perhaps for many, it means that individuals are essentially devoid of will and autonomy in a world. However, the sense that we are primarily parts of a connected and intertwined Whole seems to be even more fascinating.**