Studia Humana
Volume 8:2 (2019)

Contents

Studia Humana
Volume 8:2 (2019), pp. 1—2
DOI: 10.2478/sh-2019-0010

Preface: Philosophical Basis for Making Decisions (on the 140th Anniversary of the Birth of Jan Łukasiewicz)

Jan Woleński

University of Information Technology
and Management in Rzeszow,
Sucharskiego 2 Street,
35-225 Rzeszow, Poland

e-mail: jan.wolenski@uj.edu.pl

Andrew Schumann

University of Information Technology
and Management in Rzeszow,
Sucharskiego 2 Street,
35-225 Rzeszow, Poland

e-mail: andrew.schumann@gmail.com

Abstract:
It is a Preface to Volume 8:2 (2019) consisting of articles presented at the International Interdisciplinary Conference anniversary of the birth of Jan Łukasiewicz, Rzeszów, Poland.

Keywords: Łukasiewicz, many-valued logic, non-classical logic, Lviv-Warsaw School of Logic, Lvov-Warsaw School of Logic.

The International Interdisciplinary Conference *Philosophical Basis for Making Decisions and Non-Classical Logics* has been organized by dr. Andrew Schumann, dr. Włodzimierz Zięba, dr. Paweł Balcerak, dr. Konrad Szocik on the 140th anniversary of the birth of Professor Jan Łukasiewicz (born on December, 21st 1878 in Lviv (Polish: Lwów), a city in today's Ukraine, and died on February, 13th 1956), who was a famous representative of Lviv-Warsaw School of Logic with contributions to philosophical logic, mathematical logic, and history of logic. Using some philosophical ideas of Aristotle's *De Interpretatione* (ch. IX) (namely, his asserting that the application of the law of excluded middle to future propositions like, 'There will be a sea-battle tomorrow' should be categorically restricted), Jan Łukasiewicz proposed the first version of many-valued logic (1920). So, he showed that even some features of real world which are out of classical logic such as dynamics can be described and modeled logically still by non-classical systems. This finding that logic and rationality can be detected even in non-logical processes is quite typical for the Lviv-Warsaw School of Logic and distinguishes this school from the Vienna Circle (German:

Wiener Kreis) focusing only on classical logic and their natural extensions. Hence, the motto of this conference was that rationality can be observed everywhere. Over the past two decades our social world has changed a lot due to new media. One of the biggest changes is communications in social networks which became an important part of our everyday's life. But new forms of social communication are out of traditional forms of logical analysis of discourse. For instance, in these media the standard referential conception of truth is inapplicable – we cannot check uttered facts, but we can check contexts of uttering. In this way, we are interested to discuss non-classical logics in decision making and cognitions, new forms of communication and decision making, communication in new media.

Volume 8:2 (2019) of *Studia Humana* is a Postproceeding of the Conference described above. In this volume, the papers are devoted to different aspects of rationality. So, the paper *Logical Ideas of Jan Łukasiewicz* written by Jan Woleński discusses some logical ideas put forward by Jan Łukasiewicz within their historical context and development. The paper *Logical Determinacy versus Logical Contingency. The Case of Łukasiewicz's Three-valued Logic* submitted by Andrew Schumann is concentrated on explicating a logical intuition of Jan Łukasiewicz provided him to the idea of many-valued logic. The paper *Dispute over Logistic between Jan Łukasiewicz and Augustyn Jakubisiak. Why was it important?* written by Bartłomiej K. Krzych is devoted to the polemics with Łukasiewicz initiated by Augustyn Jakubisiak who criticized Łukasiewicz's logistics for its anti-metaphysical and anti-theological role. In the paper *The Analogy in Decision-Making and the Implicit Association Bias Effect* its author, Nataliia Reva, considers the thinking by analogy as a natural instrument human have because of the mirror neurons in our brain. The contribution *About Possible Benefits From Irrational Thinking in Everyday Life* written by Magdalena Michalik-Jeżowska is focused on indicating some benefits that may become a result of irrational thinking in the everyday human practice. The given examples of irrational thinking come from research in the field of cognitive and social psychology and behavioural economics. Magdalena Hoły-Łuczaj in her paper *Moral Considerability and Decision-Making* analyses "affectability" as a capacity of an agent to affect a considered entity. Such an approach results in significant changes in the scope of moral considerability and is relevant for discussing the human position in the Anthropocene. The paper *Practical Rationality – its Nature and Operation* prepared by Andrzej Niemczuk presents a proposal of explanation what practical rationality is, how it works and what are its criteria. Paweł Balcerak in his contribution *Can the Sense of Agency Be a Marker of Free Will?* analyses the relation between agency and responsibility. Finally, the paper *On Computers and Men* written by Tomasz Goban-Klas is addressed to the question how information technologies have transformed our thought on two levels: self-conception and relation to nature.

Studia Humana
Volume 8:2 (2019), pp. 3—7
DOI: 10.2478/sh-2019-0011

Logical Ideas of Jan Łukasiewicz

Jan Woleński

University of Information Technology
and Management in Rzeszow,
Sucharskiego 2 Street,
35-225 Rzeszow, Poland

e-mail: jan.wolenski@uj.edu.pl

Abstract:
This paper discusses the main logical ideas put forward by Jan Łukasiewicz within their historical context and further development.
Keywords: Łukasiewicz, many-valued logic, three-valued logic, non-classical logic, Lviv-Warsaw School of Logic, Lvov-Warsaw School of Logic.

Jan Łukasiewicz was born in Lviv (Lvov) in 1878 and died in Dublin in 1956. He studied philosophy in Lviv under Kazimierz Twardowski, obtained his Ph.D. in 1903 and Habilitation in 1906. In 1906, he became a *Privatdozent* at the University of Lviv, and in 1911, he was promoted to the position of extraordinary professor. Łukasiewicz moved to the University of Warsaw in 1915 and was appointed as the professor of philosophy at the Faculty of Mathematics and Natural Sciences. He formed, together with Stanisław Leśniewski, a powerful group of mathematical logicians (Warsaw School of Logic), including (there are mentioned only persons who began scientific career before 1939) Alfred Tarski, Adolf Lindenbaum, Mordechaj Wajsberg, Moses Presburger, Bolesław Sobociński, Jerzy Słupecki, Stanisław Jaskowski, and Andrzej Mostowski. Łukasiewicz organized the Polish Logical Society and essentially contributed in preparations to publishing *Collectanea Logica*, a specialized logical journal (unfortunately two first volumes printed in 1939 were destroyed). During World War II, Łukasiewicz taught at the Clandestine University in Warsaw. In 1944, Łukasiewicz obtained a permission (from the German authorities) to leave Poland. Finally, he settled in Dublin as the professor of mathematical logic at the Royal Irish Academy.

Scientific activity of Łukasiewicz can be divided into two periods. The first covers the years 1902-1915, and the second the years 1915-1956. Roughly speaking, he was occupied with various logico-philosophical problems in the first period. His Ph.D. thesis was devoted to the problem of induction. He considered induction as the inversion of deduction. Łukasiewicz’s Habilitation concerned an analysis of causality. He treated the causal relation as necessary. Perhaps [1] is the most important early work written by Łukasiewicz. This book offers a very detailed analysis of the principle of contradiction in Aristotle. This book has two tasks: firstly, an interpretation of the principle of contradiction (PCon, for brevity) and, secondly, an evaluation of arguments for and against PCon. Three interpretations of this principle can and should be distinguished: logical (concerning sentences), ontological (concerning things), and psychological (concerning judgments in the psychological sense). Łukasiewicz argues that the last understanding is irrelevant for logic, because it is an empirical fact that people assert contradictory assertions. However, Łukasiewicz denies that the logical (as well as ontological) PCon has a logical justification. We cannot deduce it from more basic principles. Finally, according to Łukasiewicz, one might say that PCon in its logical (ontological) meaning is accepted for ethical reasons, that is, as an indispensable device to distinguish between truths and falsehoods. The reported book has an Appendix presenting rudiments of mathematical logic in the version of algebra of logic as developed by Boole, Schröder and Couturat – it was the first account of the subject in Polish. Łukasiewicz shows that PCon is not an axiom (so he denies that it is the so-called highest principle of thinking) and can be proved as a logical theorem.

Works on induction led Łukasiewicz to the foundations of probability theory. Firstly, he hoped to solve the problem of induction via probability theory, but he abandoned this idea in his later works. In particular, Łukasiewicz became sceptical about a logical value of induction. His general approach (see [2]) consisted in ascribing probability to open formulas (formulas with free variable, indefinite propositions), not to full sentences which are true or false. He defined probability in the following way. Let *Fx* be a formula with free variables and *D* a finite domain. Assume that *n* is the cardinality (the number of objects) of *D* and *m* is the number of those which satisfy *Fx*. Thus, the ratio *m/n* can be defined as the logical probability of *Fx*. Łukasiewicz argued that the mathematical theory of probability allows an extension of the mentioned definition to infinite domains. Łukasiewicz introduced a classification of reasoning, very popular in Poland. He distinguished two main kinds, namely deduction (premises are the logical reason, conclusions are the logical consequents) and reduction in which the conclusion acts as logical reason and the premises as consequent. Induction is a kind of reduction, but is has no great scientific value, particularly in justifications. According to Łukasiewicz, deductive procedures are at the heart of science. His views on induction can be considered as an anticipation of Karl Popper's anti-inductivism.

Many-valued logic became the most remarkable Łukasiewicz's achievement. His above-mentioned doubts concerning PCon (and the law of the excluded middle expressed in one of his lectures before the Polish Philosophical Society) resulted in rejection of the principle of bivalence (PBiv) saying that every sentence is either true or false. Łukasiewicz announced his discovery of a non-Aristotelian logic in 1918 and elaborated its various details in two lectures in Lviv in 1920 (I skip bibliographical references – all relevant paper are included in [4]; see also [6], [7], [8] for further information). The Łukasiewicz's first motivation for introducing many-valued (more precisely, three-valued) logic was more philosophical than formal. Firstly, he believed in human freedom, creativity, and responsibility. Secondly, he was convinced that these facts and values are not coherent with determinism as an ontological theory. Consequently, he came to the conclusion that we need a non-deterministic ontology and three-valued logic as a proper background for creativity, freedom, and responsibility. Łukasiewicz considered determinism as closely connected with PBiv. He immediately observed that the issue in question has affinities with the old question, already discussed by Aristotle, concerning future contingents. If *A* is a sentence about a future contingent event, for example, the sea battle tomorrow, is it true or false at the moment of issuing it, for instance, today. The Stagirite himself argued that although the sentence $A \vee \neg A$, expressing the law of excluded middle, is universally true, its constituents, that is, *A* and $\neg A$ are not, if concern future contingents. It can be also expressed in terms referring to properties of the concept of truth. Define that *A* is occasionally true provided that if *A* is true at *t*, then *A* is true at every moment *t*' earlier than *t*. Furthermore, *A* is eternally true provided that if *A* is true at *t*, it is also true at every moment *t'* later than *t*. Łukasiewicz rejected occasionality of truth, but agreed that truth is eternal. According to him, this position suffices for considering truth as absolute. Incidentally, the absoluteness of truth as defined by Twardowski and Leśniewski consisted in its occasionality and eternality.

Łukasiewicz realized very soon that his new logic should not be called "non-Aristotelian". Since it was based on rejection of bivalence, he began to use the label "three-valued logic". The status of PBiv became the crucial issue. According to Łukasiewicz, this principle is not a theorem of logic, but a metalogical rule, which can be accounted as the conjunction of the metalogical non-contradiction and the metalogical excluded middle. Its acceptance or not cannot be reduced to purely logical circumstances, but requires assuming of some extralogical decisions, for instance, ontological. Anyway, there is no logical force to accept PBiv. If we reject this principle, we can introduce more than two logical values. Łukasiewicz introduced the third logical value, usually denoted by the fraction ½. Its meaning is explained by rules related to traditional truth-tables. In particular, we have the logical value v: if $v(A) = ½$, then $v(\neg A) = ½$, if $v(A) = ½$, $v(B) = ½$, then $v(A \vee B) = ½$ and $v(A \wedge B) = ½$. According to these equalities, if $v(A) = ½$, then $v(A \vee \neg A) = ½$ and $v(A \wedge \neg A) = ½$. This means that the (logical) law of the excluded middle and the (logical) law of non-contradiction are not theorems (tautologies) of three-valued logic. In the inter-war period, Łukasiewicz generalized the three-valued logic (usually denoted by the symbol $\mathbf{Ł_3}$) to logics with finite and infinite number of values as well as formulated various axiomatizations of theses systems. Several results concerning many-valued logic were obtained by Łukasiewicz's students, namely, Lindenbaum, Słupecki, Sobociński, Tarski and Wajsberg.

The problem of interpretation of many-valued logic was essential. In his first works on three-valued logic, Łukasiewicz understood the third value as possibility. Later he abandoned this intuition and decided to speak about ½ as a logical value, which has the same status as other. Yet Łukasiewicz believed that one of the systems, two-valued or many valued, is satisfied in the reality – he conjectured that the logic with infinitely many values is "true" on the world. However, he gradually became more and more formalistic in his thinking about logic. According to him, logical systems are formal constructions, independent of their relations to the reality or applicability to concrete scientific or technical problems. Historically speaking, Łukasiewicz's work on many-valued logic was pioneering. Nicolai Vasiliev, a Russian logician had some ideas about many-valueness, but he did not elaborated them in a formal way. Emil Post, an American logician, constructed a many-valued logic, but it was rather a purely formal system without an intuitive interpretation. Today, study of many-valued logics (plural as justifying for a considerable plurality of such logics) is a branch of mathematical logic. Many-valued logic has also several technical and philosophical applications, for instance, offers a basis for studies on paraconsistency. In fact, $\mathbf{Ł_3}$ is sometimes considered as the first formalization of paraconsistency.

The first intuitive interpretation of the third value as possibility immediately led to the problem of the relation between $\mathbf{Ł_3}$ and modal logic. Łukasiewicz accepted the following principles: (a) if it is not possible that A, then not-A; (b) if not-A, then it is not possible, that A; (c) for some A, it is possible that A and it is possible that not-A. Łukasiewicz demonstrated that (a)–(c) cannot be proved in two-valued logic. Hence, implementing modalities into three-valued logic appeared as a possible solution. Tarski proposed to define "it is possible that A' as $\neg A \Rightarrow A$. This definition functions in $\mathbf{Ł_3}$. However, Łukasiewicz did not construct a system of modal logic before 1950, partially due to various critical remarks about his modal ideas. In particular, Ferdinand Gonseth observed that Łukasiewicz's assumptions entail that the formula $A \wedge \neg A$ is possible just in the case if $v(A) = ½$, contrary to the common claim that contradictions are impossible. Łukasiewicz tried to solve this problem and other difficulties by the modal system based on four-valued logic, but this proposal did not gain an acceptance. One of the main features of all Łukasiewicz's logical systems is that they are strictly extensional. It means that if $v(A) = v(B)$, then the formulas A and B are substitutable per *salva veritate*. On the other hand, modal operators, possibility and necessity, are not extensional in Lewis' systems. Consequently, if, for instance, if A is possible, A is true, B is true, this set of premises does not implies that A is possible. On the other hand, possibility and necessity as understood by Lewis are not definable in two-valued logic. Consequently, Lewis' modal logic is extension of two-valued logic and this circumstance generates intensionality. Defining modality by Tarski's proposal, admits embedding modalities into $\mathbf{Ł_3}$ (similar constructions are possible in systems with more than three values) and keeps extensionality. Incidentally, the principle of extensionality functioned as a fundamental dogma of Warsaw School of Logic (it was particularly stressed by Leśniewski) – this circumstance blocked formalizing intensional context by the Polish logicians.

Łukasiewicz extensively worked on propositional calculus (or rather calculi; see [3] for a summary). He invented a special logical notation (called Polish notation or the Łukasiewicz notation). This symbolism avoids punctuation signs (brackets, points) – the structure of a formula is determined by the succession of signs. Functors are represented by the capital letters: *N* (negation), *C* (implication), *K* (conjunction), *A* (disjunction), *D* (bi-negation) and *E* (equivalence). For instance, the formula (I employ small letters as propositional variables) $(p \Rightarrow q) \Leftrightarrow (\neg q \Rightarrow \neg p)$ become *ECpqCNpNp*, the formula $p \vee \neg p$ is *ApNp*, the formula $\neg(p \wedge \neg p)$ is *NKpNp* and so on. Łukasiewicz built various axiomatizations of propositional calculi. He preferred the simplest constructions, for instance, with minimal number of possibly shortest axioms. Consequently, he was looking for single shortest axioms as the best. The most popular is his following axiomatization: *CCpqCCqrCpr* (in traditional setting: $(p \Rightarrow q) \Rightarrow ((q \Rightarrow r) \Rightarrow (p \Rightarrow r))$; the transitivity rule for implication), *CCNppp* (traditionally: $(\neg p \Rightarrow p) \Rightarrow p$; characterization of implication via falsity of the antecedent, *ex falso quolibdet*), *CpCNpq* (traditionally: $p \Rightarrow (\neg p \Rightarrow q)$; *ex falso quolibdet*). Of course, it is not the simplest one, because it consists of three axioms. Since *Dp* can be defined as *Dpp,* bi-negations suffices as the sole primitive concept of propositional logic (*C*, *K* and *A* must be supplemented by *N*). Consequently, the entire propositional calculus can be axiomatized by a formula consisting from *D*'s and propositional variables. Łukasiewicz also investigated partial propositional calculi, for instance, based on *E* as the sole functor and intuitionistic logic. After 1945, he introduced propositional calculus with the so-called variable functors. The idea is that this systems contains variables for functors (in standard version, propositional functors are constants). The resulting system is very powerful and allows proving that intuitionistic propositional calcuslus is more expressive than classical one.

Łukasiewicz had a deep interest in the history of logic. He proposed to look at historical logical doctrines as anticipations of modern formal logic. This research project required reading of older logic through glasses of modern tools. Łukasiewicz, guided by this methodology, achieved revolutionary discoveries. In particular, he showed that Stoic logic was another system than Aristotelian syllogistic. More specifically, the Stoics developed elements of propositional logic, but the Stagirite elaborated a logic of names. Aristotle was a favourite logician of Łukasiewicz. In fact, two books published by the latter during his lifetime concerned the ideas of the former (see [1] and [4]). Although Łukasiewicz did not agree with Aristotle in many important points, he was convinced that the Aristotelian logic requires a modern interpretation. It was offered in [4], where syllogistic was reconstructed as an axiomatic system assuming propositional calculus. The second edition of this book contains a detailed analysis of Aristotle's logic of modalities. Łukasiewicz's idea that old logic should be investigated as an earlier stage of contemporary logic became fairly revolutionary and essentially changed understanding of the history of logic.

Łukasiewicz was a philosopher by education. Although he maintained in the second period of his scientific activities that logic should be entirely purified from philosophical assumptions, he was continuously interested in philosophical problems of logic. He entirely rejected psychologism and protested against the use of the term "philosophical logic" as leading to conflating logic with psychology and epistemology. Mathematical logic is the only logic and must be separated from philosophy as well as mathematics. On the other hand, logic is a fundamental instrument of reasoning and rational thinking, the morality of speech and thought (Łukasiewicz's saying). In particular, philosophy should be axiomatized in order to be a science. In general philosophy, Łukasiewicz preferred ontology over epistemology. He argued that post-Cartesian philosophy with its epistemological orientation, culminating in Kant, poisoned logic by psychologism – Leibniz was the only exception. This assessment of the history explains Łukasiewicz's sympathies to Aristotle and the Schoolmen. Łukasiewicz defended logic against objections pointing out that it recommends the empty formalism, entirely inconsistent with needs of philosophizing. According to Łukasiewicz, logic as such does not privilege any concrete philosophy and can be reconciled with many philosophical positions. On the other hand, every philosopher should obey general logical principles as indispensable for rationality. Clearly, his philosophical views were more explicit in the years 1902-1903, but he did not lose philosophical interests until the end of his life.

References

1. Łukasiewicz, J. *O zasadzie sprzeczności u Arystotelesa (On the Principle of Contradiction in Aristotle)*, Kraków: Polska Akademia Umiejętności, 1910; Germ. tr. *Über den Satz des Widerspruch bei Aristoteles,* Olms, Hildesheim 1993.
2. Łukasiewicz, J. *Die logische Grundlagen der Wahrscheinlichkeitsrechnung,* Kraków: Polska Akademia Umiejetności, 1913; partial Eng. tr in [5].
3. Łukasiewicz, J. *Elementy logiki matematycznej (Elements of Mathematical Logic)* [Lecture notes], 1929; Eng. tr., Oxford: Pergamon Press, 1963.
4. Łukasiewicz, J. *Aristotle's Syllogistic from the Standpoint of Modern Formal Logic*, Oxford: Clarendon Press, 1951; 2nd ed. 1957.
5. Łukasiewicz, J. *Selected Works*, Amsterdam: North-Holland, 1970.
6. Malinowski, G. *Many-Valued Logics,* Oxford: Oxford University Press, 1993.
7. Rescher, N. *Many-Valued Logic,* New York: McGraw Hill.
8. Woleński, J. *Logic and Philosophy in Lviv-Warsaw School,* Dordrecht: Kluwer, 1989.

Studia Humana
Volume 8:2 (2019), pp. 8—15
DOI: 10.2478/sh-2019-0012

Logical Determinacy versus Logical Contingency. The Case of Łukasiewicz's Three-valued Logic

Andrew Schumann

University of Information Technology
and Management in Rzeszow,
Sucharskiego 2 Street,
35-225 Rzeszow, Poland

e-mail: andrew.schumann@gmail.com

Abstract:
In constructing the three-valued logic, Jan Łukasiewicz was highly inspirited by the Aristotelian idea of logical contingency. Nevertheless, we can construct a four-valued logic for explicating the Stoic idea of logical determinacy. In this system, we have the following truth values: 0 ('possibly false), 1 ('necessarily false'), 2 ('possibly true'), 3 ('necessarily true'), where the designated truth value is represented by the two values: 2 and 3.
Keywords: Łukasiewicz, many-valued logic, three-valued logic, modal logic, four-value logic, logical contingency, logical determinacy.

1. Introduction

There are two extreme approaches to modalities: logical contingency and logical determinacy. According to the first approach, there exist contingent events *A* which are possible to be and possible not to be simultaneously: '*A* is possible and non-*A* is possible'. This approach was formulated by Aristotle for the first time. The second approach is a negation of the first one: 'Either *A* is necessary or non-*A* is necessary'. According to this claim, each event is either necessarily to be or necessarily not to be. At first, it was formulated by the Stoics.

In this paper, I show that the modal logics **D** and **T** help to formalize the Aristotelian approach (Section 2) and the modal logics **CD** and **K=** formalize the Stoic approach (Section 3). Łukasiewicz proposed his system of three-valued logic as his own attempt to justify the Aristotelean idea of logical contingency (Section 4). Nevertheless, we can propose a four-valued logic to justify the Stoic idea of logical determinacy (Section 5). This logic is proposed for the first time.

2. Modal Logic for Logical Contingency

The vocabulary of modal logic is as follows:

- $p_0, p_1, \ldots$ – propositional atoms, *Prop*;

- $\neg$, $\vee$, $\wedge$, $\Rightarrow$, $\Leftrightarrow$ – propositional connectives: negation (“not…”), disjunction (“… or …”), conjunction (“… and …”), implication (if… , then, …”), equivalence (“… if and only if…”), respectively;
- $\Box$, $\Diamond$ – modal operators: the symbol $\Box$ is used for ‘necessity’ (“… is necessarily”) and the symbol $\Diamond$ is for ‘possibility’ (“… is possibly”).

On the basis of this vocabulary, we can define well-formed formulas:

- Each propositional atom from *Prop* is a formula;
- If A and B are formulas, then $\neg A$, $\neg B$, $A \vee B$, $A \wedge B$, $A \Rightarrow B$, $A \Leftrightarrow B$, $\Box A$, $\Box B$, $\Diamond A$, $\Diamond B$ are formulas, as well.

The meanings of well-formed formulas without modal operators are defined in a standard way within the two-valued logic, about the meanings of modal formulas please see [2].

The basic modal logic, denoted by **K** in honor of Soul Kripke, has the following axioms:

- All propositional axioms such as $A \vee \neg A$;
- All instances of the Kripke schema: $\Box(A \Rightarrow B) \Rightarrow (\Box A \Rightarrow \Box B)$.

The set of these axioms is closed under the following two inference rules:

- *modus ponens*: from $A \Rightarrow B$ and A it follows that B;
- *Necessitation rule*: if A is an axiom, then $\Box A$ is an axiom, too.

On the basis of **K**, we can obtain two additional systems of modal logic for logical contingency by adding to **K** the following two schemas [2]:

(D) $\Box A \Rightarrow \Diamond A$

(T) $\Box A \Rightarrow A$

If we add to **K** all the instances of (D), then the new modal logic is denoted by **D**. In the meanwhile, if we add to **K** all the instances of (T), then the new modal logic is denoted by **T**. Let us notice that all the axioms of **D** are contained in the class of axioms of **T**. For the first time, the intuition of this logic came to mind of Aristotle and some axioms of **T** were considered in his book Περί ερμηνείας (*De Interpretatione*). In this logic we cannot infer $\Box A \vee \Box B$ from $A \vee B$:

Δὲ οἷον ἀνάγκη μὲν ἔσεσθαι ναυμαχίαν αὔριον ἢ μὴ ἔσεσθαι, οὐ μέντοι γενέσθαι αὔριον ναυμαχίαν ἀναγκαῖον οὐδὲ μὴ γενέσθαι• γενέσθαι μέντοι ἢ μὴ γενέσθαι ἀναγκαῖον (*De Interpretatione* 9, 19a).

Necesse est quidem futurum esse bellum navale cras vel non esse futurum sed non futurum esse cras bellum navale necesse est vel non futurum esse, futurum autem esse vel non esse necesse est (*De Interpretatione* 9, 19a).

A sea-fight must either take place tomorrow or not, but it is not necessary that it should take place tomorrow, neither is it necessary that it should not take place, yet it is necessary that it either should or should not take place tomorrow.

Otherwise we should accept that $\Diamond A \Rightarrow A$ is ever false and $A \Rightarrow \Box A$ is ever true. But it is impossible: Οὐδὲν ἄρα οὔτε ἔστιν οὔτε γίγνεται οὔτε ἀπὸ **τύχης** οὔθ’ ὁπότερ’ ἔτυχεν, οὐδ’ ἔσται ἢ οὐκ ἔσται, **ἀλλ’ ἐξ ἀνάγκης ἅπαντα** καὶ οὐχ ὁπότερ’ ἔτυχεν (De Interpretatione 9, 18b). […] ἅπαντα οὖν τὰ ἐσόμενα ἀναγκαῖον γενέσθαι (De Interpretatione 9, 18b).

Nihil igitur neque est neque fit nec a casu nec **utrumlibet**, nec erit nec non erit sed **ex necessitate omnia** et non utrumlibet (*De Interpretatione* 9, 18b). […] Omnia ergo quae futura sunt necesse est fiery (*De Interpretatione* 9, 18b).

Then nothing is or takes place **occasionally**, either in the present or in the future, and there are no real alternatives; **everything takes place of necessity** and not occasionally […]. […] Then all that is to be must necessarily take place in the future.

The point is that Aristotle assumes the existence of logical contingency (ἐνδεχόμενον). For example, propositions such as ‘A sea-fight will be tomorrow’ (A) are logically contingent: $\Diamond A \wedge \Diamond\neg A$ (ἐν οἷς ἄμφω ἐνδέχεται καὶ τὸ εἶναι καὶ τὸ μὴ εἶναι; *in quibus utrumque contingit et esse et non esse*). These statements can be true by some valuations in **T**.

3. Modal Logic for Logical Determinacy

So, systems **D** and **T** are used for explicating logical contingency within a modal logic. Nevertheless, we can explicate logical determinacy within a modal logic, too. For this purpose, we should involve other schemas added to **K** [2]:

(CD) $\Diamond A \Rightarrow \Box A$

(=) $A \Rightarrow \Box A$

If we add to **K** all the instances of (CD), then the new modal logic is denoted by **CD**. At the same time, if we add to **K** all the instances of (=), then the new modal logic is denoted by **K**=. From (=) we can infer (CD). It means that all the axioms of **CD** occur among axioms of **K=**. The intuition for modal logics **CD** and **K=** was expressed by the Stoics (first of all, by Chrysippus):

Nihil enim fieri sine causa potest (Cicero: *De Divinatione* 2, 61).

Nothing happens without a cause.

Motum nullum esse sine causa (Cicero: *De Fato* 23).

No motion is without a cause.

It means that each proposition is necessarily true or necessarily false because of causes existed for all events described by propositions. So, the proposition ‘A sea-fight will take place tomorrow’ is either necessarily true or necessarily false right now, since there are or there are not causes for the event to be a sea-fight tomorrow right now. Some Stoic synonyms for the word ‘necessity’ (ἀνάγκη): ‘inexorable’ (ἀπαράβατος), ‘inflexible’ (ἄτρεπτός), ‘invincible’ (ἀνίκητος), ‘unconquerable’ (ἀνεκβίαστος), ‘unpreventable’ (ἀκώλυτος), ‘immutable’ (ἀμετάβλητος), and ‘unchangeable’ (ἀμετάθετος) [1].

In the logic of **K=**, the statement of contingency $\Diamond A \wedge \Diamond\neg A$ is always false, because the statement of determinacy $\Box A \vee \Box\neg A$ (the negation of $\Diamond A \wedge \Diamond\neg A$) is directly delivered from (CD) as an axiom.Hence, for logical contingency we deal with modal logic **T** and for logical determinacy we deal with logic **K=**.

4. Three-valued Logic for Logical Contingency

In his famous paper *On Three-Valued Logic* [3], Jan Łukasiewicz was mainly inspirited by the Aristotelian modal reasoning from the book Περί ερμηνείας (*De Interpretatione*), especially about sea-fights tomorrow. In order to describe logical contingency, Łukasiewicz decided to introduce the third truth value ½ with the meaning ‘possible’. So, in his logic there are the following truth values: 0 (‘false), 1 (‘true’), ½ (‘possible’), where 1 is the designated truth value. The intuition for this introducing was as follows. The value 0 was understood as ‘necessary false’, the value 1 as ‘necessary true’, and the new value ½ as ‘possible’. In the meanwhile, 0 < ½ < 1 so that we have the true implication ½ ⇒ 1 which corresponds to axiom (CD).

In this logic the meanings of propositional connectives are defined as follows:

Negation

A	$\neg A$
1	0
½	½
0	1

Conjunction

A	B	$A \wedge B$
1	1	1
1	½	½
1	0	0
½	1	½
½	½	½
½	0	0
0	1	0
0	½	0
0	0	0

Disjunction

A	B	$A \wedge B$
1	1	1
1	½	1
1	0	1
½	1	1
½	½	½
½	0	½
0	1	1
0	½	½
0	0	0

Implication

A	B	$A \Rightarrow B$
1	1	1
1	½	½
1	0	0
½	1	1
½	½	1
½	0	½
0	1	1
0	½	1
0	0	1

According to these truth valuations, the law of excluded middle $A \vee \neg A$ is not an axiom. Indeed, its truth valuation does not give only truths:

A	$\neg A$	$A \vee \neg A$
1	0	1
½	½	½
0	1	1

The law of contradiction $A \wedge \neg A$ has not only falsehood in this logic:

A	$\neg A$	$A \wedge \neg A$
1	0	0
½	½	½
0	1	0

In this logic we can define two modal operators: □ (“… is necessarily”) and ◇ (“… is possibly”), as follows:

A	$\Box A$
1	1
½	0
0	0

A	$\Diamond A$
1	1
½	1
0	0

From both truth tables, it follows that (D) and (T) are axioms of Łukasiewisz’s three-valued logic. Hence, Łukasiewicz supports the Aristotelian approach to logical modalities and, therefore, shares the Aristotelian ideas of logical contingency. Nevertheless, we can construct many-valued systems for the Stoic approach focused on logical determinacy.

5. Four-valued Logic for Logical Determinacy

Let us introduce the following four truth values: 0 (‘possibly false), 1 (‘necessarily false’), 2 (‘possibly true’), 3 (‘necessarily true’), where the designated truth value is represented by two values: 2 and 3. The intuition for these values is based on the following inequalities: $0 < 1 < 2 < 3$ so that we have the true implications $0 \Rightarrow 1$ and $2 \Rightarrow 3$ which correspond to axiom (CD). Now let us define propositional connectives on these values:

Negation

A	$\neg A$
3	0
2	1
1	2
0	3

Conjunction

A	B	$A \wedge B$
3	3	3
3	2	2
3	1	1
3	0	0
2	3	2
2	2	2

2	1	1
2	0	0
1	3	1
1	2	1
1	1	1
1	0	0
0	3	0
0	2	0
0	1	0
0	0	0

Disjunction

A	B	$A \vee B$
3	3	3
3	2	3
3	1	3
3	0	3
2	3	3
2	2	2
2	1	2
2	0	2
1	3	3
1	2	2
1	1	1
1	0	1
0	3	3
0	2	2
0	1	1
0	0	0

Implication

A	B	$A \vee B$
3	3	3
3	2	2
3	1	1
3	0	0
2	3	3
2	2	3
2	1	2
2	0	1
1	3	3
1	2	3
1	1	3
1	0	2
0	3	3
0	2	3
0	1	3
0	0	3

In the four-valued logic with the two designated truth values: 2 and 3, the law of excluded middle $A \vee \neg A$ is an axiom. We can check it:

A	$\neg A$	$A \vee \neg A$
3	0	3
2	1	2
1	2	2
0	3	3

The law of contradiction $A \wedge \neg A$ cannot take the two designated truth values:

A	$\neg A$	$A \wedge \neg A$
3	0	0
2	1	1
1	2	1
0	3	0

The two modal operators: $\Box$ ("… is necessarily") and $\Diamond$ ("… is possibly") are understood as follows:

A	$\Box A$
3	3
2	3
1	1
0	1

A	$\Diamond A$
3	2
2	2
1	0
0	0

So, the necessity operator preserves the value 3 ('necessarily true') and 1 ('necessarily false') and makes 3 ('necessarily true') from 2 ('possibly true') and makes 1 ('necessarily false') from 0 ('possibly false'). The possibility operator preserves the value 2 ('possibly true') and 0 ('possibly false') and makes 2 ('possibly true') from 3 ('necessarily true') and makes 0 ('possibly false') from 1 ('necessarily false').

Thus, (CD) and (=) are axioms of the new logic:

A	$\Diamond A$	$\Box A$	$\Diamond A \Rightarrow \Box A$
3	2	3	3
2	2	3	3
1	0	1	3
0	0	1	3

A	$\Box A$	$A \Rightarrow \Box A$
3	3	3
2	3	3
1	1	3
0	1	3

As we see, this logic is one of the possible formalizations of the Stoic idea of logical determinacy.

6. Conclusion

In developing many-valued logics, Łukasiewicz was highly inspirited by the Aristotelian modal approach towards logical contingency, although there is possible an alternative approach put forward by the Stoics towards logical determinacy. Within the Stoic approach we can appeal to the many-valuedness, too. So, we can propose a four-valued logic with the following truth values: 0 (‘possibly false), 1 (‘necessarily false’), 2 (‘possibly true’), 3 (‘necessarily true’), where the designated truth value is represented by the two values: 2 and 3.

References

1. Bobzien, S. *Determinism and Freedom in Stoic Philosophy*, Oxford: Clarendon Press, 1998.
2. Garson, J. W. *Modal Logic for Philosophers*, Cambridge: Cambridge University Press, 2006.
3. Łukasiewicz, J. O logice trójwartościowej, *Ruch filozoficzny* 5, 1920, pp. 170–171б; English translation: On three-valued logic, In L. Borkowski (ed.), *Selected works by Jan Łukasiewicz*, Amsterdam: North–Holland, 1970, pp. 87-88.

Studia Humana
Volume 8:2 (2019), pp. 16—24
DOI: 10.2478/sh-2019-0014

Dispute Over Logistic Between Jan Łukasiewicz and Augustyn Jakubisiak. Why Was it Important?

Bartłomiej K. Krzych

University of Rzeszów,
Rejtana 16c Av.,
35-959 Rzeszów, Poland;
Pontifical University of John Paul II
in Kraków, Poland

e-mail: bartlomiejkk@gmail.com

Abstract:
Augustyn Jakubisiak (1884-1945), Polish priest, philosopher and theologian, undertook polemics with Jan Łukasiewicz, whom he knew personally. A dispute concerning the so-called logistics (mathematical logic) and its relationship with philosophy developed between the two. The most important arguments were laid out, primarily in the following works: in the case of Jakubisiak, in the book *From Scope to Content* and in the case of Łukasiewicz, in the texts *Logistics and Philosophy* and *In the Defense of Logistics*. Jakubisiak criticized logistics for its anti-metaphysical, anti-theological and anti-religious attitude, which was based on neo-positivist philosophy, and led, in consequence to atheism. He also claimed that one should focus on what is concrete, avoiding idealization and abstraction (meaning the content of concepts, not their scope). Łukasiewicz defended logistics claiming that it possesses its own methods based on intellect, and is also an area of independent knowledge (but not completely detached) from philosophy, due to the fact it can consider the most important philosophical problems such as finiteness and infinity. This dispute, as the researchers identified, basically concerned the reduction of philosophy to the study of language (analytic philosophy) and initiated one of the most important discussions concerning the relationship between philosophy and logic. This debate was crucial because it also concerned questions related to fundamental metaphysical issues (naturalism – supranaturalism, rationalism – irrationalism) and epistemological issues (realism – idealism, boundaries and structure of cognition).
Keywords: Lviv-Warsaw School, philosophy of logic, Polish logic and philosophy.

1. Introduction

The importance of the Lviv-Warsaw School (LWS) for Polish philosophy and philosophy in general is undeniable and universally recognized [36]. The best-known achievements of the LWS are related to research and development of logic,[1] to mention only Józef M. Bocheński, Alfred Tarski and Jan Łukasiewicz. However, the polemics of the members and representatives of the LWS are less widely known, especially the disputes conducted in the circle of (international) Polish philosophical thought during the interwar period [37]. One of the most important discussions on the relationship and mutual relationship between philosophy and logic took place in 1936-1937, between Jan Łukasiewicz and Augustyn Jakubiskiak.[2] Their polemics also provoked reactions from other thinkers (e.g. Adam Żółtowski, Zygmunt Zawirski). Moreover, the question of the role of logic in philosophy was raised at the III Polish Philosophical Congress in Kraków in 1936 [22]. This dispute, however, being relatively unknown[3], results firstly from: firstly, the hermetic character of the environment in which it was conducted and later commented on [37, pp. 134nn], [35, pp. 24-49] and secondly, from the fact that Jakubisiak was forgotten and omitted in current philosophical, historical and theological research. This situation is cleverly described by Bohdan Chudoba:

> Jakubisiak, the author of three learned and penetrating books on the subject of creative freedom, was one of the most lucid and also most truly Christian thinkers of our century. His struggle against the pseudo-myths was only equaled by his defense of the Christian tradition against the spread of the Aristotelian, Thomistic and Cartesian obscurantism. In his faithfulness to the Christian message he evidently incurred the wrath of the servants of the false myths as well as of those Christians who are ready to bend over backward in catering to these servants. As result, his very name has been omitted form even most comprehensive encyclopedias as well as from textbooks of the history of philosophy [8, p. 113].

While one can agree with the final statement, the dispute between Jakubisiak and Łukasiewicz is a good example of the fact that it is impossible to be both simultaneously a specialist anda visionary in every field of philosophical and scientific research. Nevertheless, as Jan Woleński indicates [37, p. 134], polemics *en question* initiated one of the most important discussions on the relationship between philosophy and logic. Łukasiewicz himself wrote the following in one of his letters to Bocheński:

> I would not like much to be written about my pre-logistic philosophical works; both the dissertation about causation and my book about the principle of contradiction in Aristotle I consider old and unsuccessful. I attach some importance to the work *On science and probability*, and to the polemics with Fr. Jakubisiak and the article *In the defense of logistics*, and besides, another two philosophical articles [21][4].

In short: if today's analytical philosophers and logicians can be grateful to Jakubisiak for anything, then certainly it must be for his contribution to the development and precision of the thoughts of Jan Łukasiewicz [24, p. 117], [35, p. 24].[5]

2. *Virus Logisticus*?

The basic accusation formulated by opponents of the LWS, including – probably most significantly – Jakubisiak himself, consisted firstly of absolutizing logic and its tools (contrary to the intentions of the representatives of the LWS themselves), and then of categorically rejecting it as another attempted attack on the truth and metaphysics. *Virus logisticus* is a term used by Jakubisiak and other Catholic thinkers who opposed[6] this accusation – and, as it emerged, did not fully understand the ideas and methods cultivated within the LWS [33]. This term has become so prominent that it has entered both the general circulation, as well as current literature [e.g. 39, p. 162].

Jacek Jadacki in the article *Semiotics of the LWS: Main Concepts* [12] gives two statements, the authors of which are Jakubisiak and Bolesław Gawecki. The first wrote: "The virus logisticus brought from abroad was bred perfectly on the broth of the LWS school of philosophy, and from there it spread through the universities of Poland" [12, p. 131]. The second one: "Their favorite weapon is what is commonly called 'grabbing for words.' Their exactness is their deity, a scientific cry of a battle; they crush, annihilate their opponents, moreover, the few and speaking shyly" [12, p. 131]. Jakubisiak, as we already know, belonged to a group of these opponents. He was one of the most important participants in discussions, because his polemics – however we know today, that it was not justified [28], [38, pp. 150-153], and in a slightly broader context [40, pp. 150-153], [10, p. 231] – allowed Łukasiewicz to clarify his views and give, not only to him, a proper understanding of contemporary logic and its relation to philosophy and science [3, p. 337], [32, p. 341].

Jakubisiak's merit was mainly that he began a discussion on logistics (as mathematical logic was called at that time) relations with philosophy, which allowed Łukasiewicz not only to overcome erroneous ideas about the proposals and postulates of the LWS, but also change his own style of speech to a much less emotional tone. This fact is emphasized by Wolak [35, pp. 42-43] and Woleński [37, p. 164]. The sources of Jakubisiak's accusations can be found in the opinions of Twardowski himself about the loss of contact with reality by the school and "vain formalism" [27, pp. 148-150], as well as in the Christian-theological background of Jakubisiak's thoughts and his own philosophical conception, which for the sake of simplicity let us call autodeterminism.[7] For Jakubisiak, the discontinuity existing in the world (including the cases of human choices and actions) is recognized by the intellect in an intuitive way. It is a manifestation of the existence of closed and autonomous entities, which, moreover, can be self-determinant, thus escaping determinism (overthrown by modern science, especially quantum mechanics) and indeterminism (which denies the stability and rationality of free will). By what beings (or people), thanks to their salvation (free will, intellect, self-awareness), are able to form the first principles of reality, which are the first three scholastic principles of reason (non-contradiction, identity, *tertium non datur*).[8]

As for Łukasiewicz, let us give a short overview of his views in the context of his dispute with Jakubisiak, through the synthetic elaboration of Stanisław Borzym [5, pp. 517-518]. Thus, for Łukasiewicz, any philosophy without a scientific method that operates with ambiguous terms can have at most aesthetic or ethical value. Neglect of logic was the main factor in this state of affairs. The "new" logic gives a new criterion of accuracy and allows to formulate an outgoing program – according to the words of Lukasiewicz – above the emptiness of the current philosophical speculations. This program can be summarized as follows: one should deal with comprehensible issues, i.e. those that can be formulated based on contemporary knowledge and scientific methods. The method itself is based on mathematical logic, i.e. be deductive and axiomatic. The axioms should be intuitively clear and simple sentences, and the original concepts should contain such expressions whose meaning can be easily grasped and given their understandable examples. The results of such research should be controlled by combining them with experience data and the results of sciences, especially natural sciences.[9]

3. Dispute

The polemics itself was played out in three basic stages: 1936 – Jakubisiak's introduction to the book *From the scope to the content* which is a collection of his lectures and speeches [17]; 1936 – Lukasiewicz's answer in the pages of the "Philosophical Review" (article *Logistics and Philosophy*) [24]; 1937 – Jakubisiak's answer in the pages of the "Philosophical Review" (article *On the book "From the scope to the content"*) [15]. An article by Łukasiewicz entitled *In defense of logistics* [23] can be regarded as a kind of epilogue, it was originally published in the book *Catholic Thought Towards Contemporary Logic* in 1937 being the fruit of the Third Polish Philosophical Congress in Krakow (September 1936) [22]. At this point, Łukasiewicz does not quote Jakubisiak anywhere, but he refers to his article from the "Philosophical Review" and clarifies some of his thoughts.[10]

Before proceeding to delineate and describe the arguments of Jakubisiak and Łukasiewicz, it is necessary to emphasize once again the somewhat confessional and prejudiced nature of the attacks on the LWS, resulting largely from the misunderstanding of modern logical ideas, which in turn is rooted in the classic approach not only to logic, but primary to the basic problems of philosophy.[11] This is why Jakubisiak (and his supporters) could accuse Łukasiewicz and his disciples of the anti-theological, anti-religious and anti-metaphysical attitude (however, it is well-known, for example, that the LWS had different views, and Łukasiewicz considered himself theist). According to Jakubisiak, logistics is just another name for neo-positivism, the direct consequence of which is atheism. Łukasiewicz showed that there is a difference between logistics and philosophy, and furthermore that there are priests who recognize logistics and apply it in theological provinces. Jakubisiak, on the other hand, answered that it is a natural affliction of Poles to sanctify everything, to seek sanctity where it does not exist, even in logistics [cf. 35, pp. 24-49].[12]

Symptomatic of this way of thinking is the discussion of Jakubisiak's *book From the scope to the content* [17], which in the "Homiletical Review" in 1936 was published by Fr. Aleksander Syski [33]. He wrote:

> The slogan of struggle against mysticism, or religion, raised by the head of logistic school Bertrand Russell, may have been the most painful, and the highest scientific criterion was taken especially by bred Polish logists of Lviv school of philosophy with Łukasiewicz, Kotarbiński and other "strong heads" at the head, and therefore if, where, in Poland, in the face of the command of university chairs by this foreign pseudo-scientific logistic poison brought to us, it would be time for the reaction to be great. This reaction abroad, especially in France, is a great triumph – and its symptom is the book of Fr. Jakubisiak. He beats logisticians, and in general all pseudoscientists, or actually, philosophical determinists who refer to science, he beats them with science [33, pp. 376-377].

How was it really?[13] In the introduction to his book, Jakubisiak, at the very beginning, makes a program for his philosophy: "The individual is the end to which human cognition should go and *de facto* it do so. It is to make it to this end, because it has the source of everything that man knows about reality" [17, pp. 7-8]. In addition, according to the Polish philosopher, it is known from logic that the scope of the concept means the elements that make up its composition (e.g. the scope of the concept of "human" are all people), and its content are common features of elements falling under a given concept (e.g. common features of all people). The larger the scope, the more general the concept, the smaller the scope the more the concept is richer in content. The richest content has an individual – each time it belongs only to a given unit. For Jakubisiak, in the face of the crisis and the decline of determinism, the most important goal of science is to know the individual. This must be overcome by the thoughts of the ancient Greeks, as well as by Kant, who separated the being and thought and thus established the guiding principle of modern philosophical schools: being is unknowable. Jakubisiak calls this philosophical attitude criticism and also assigns it to logistics, which he calls logical empiricism and mathematical logic. He counts Russell, Whitehead, Kreis, Wittgenestein, Schlick and Carnap as one of these philosophical currents, besides of course the LWS. These thinkers "not only break radically with all metaphysics, but also speak inexorable to the philosophical struggle of the doctrines of the past" [17, p. 11]. Their main objection to the current philosophy is the lack of a method – says Jakubisiak citing the text of Łukasiewicz on the method in philosophy. According to Jakubisiak, although logistics also wants to break with Kant's criticism and the concept of the theory of cognition (according to Łukasiewicz, mathematics logic is a salutary solution to philosophy), yet its postulates coincide with the philosophy of thinker form Königsberg. According to Jakubisiak, these are: 1) the negation of metaphysics resulting from the negation of the relationship between the subject and the object of cognition, this time not in the creation of *a priori* categories, but in the closing of philosophy in the narrow frames of abstract formulas that impose on cognition *a priori* structure of assumptions that stop the spontaneity of the human mind (this what is not general is not scientific). This leads Jakubisiak to call logicians "new encyclopedists"; 2) a postulate of determinism, in essence opposed to indeterminism in quantum physics, which in turn manifests itself in the desire to "unify all sections of knowledge with the most general and all-binding binding law of causality" [17, p. 16]. They replace the necessary causal relationship of the former determinists with a functional relationship – and in this, according to Jakubisiak, they are wrong, because their efforts overthrow the Heisenberg uncertainty principle (functions cannot be one-determinant). In the next part of the introduction Jakubisiak discusses Tadeusz Kotarbiński, trying to show that the goal of modern logistics is the negation of faith and religion. "This is where the scientific philosophy ultimately leads. It begins with the negation of metaphysics and ends with the negation of God" [17, p. 23]. As Jakubisiak goes on: "The results of this philosophy do not bring anything fundamentally new to philosophical thought, they only restore old errors" [17, p. 23]. In the final analysis, Jakubisiak regrets that *virus logisticus* has been spreading in Poland and is calling for a reaction against it, and calling this virus "pseudophilosophy." Only at the very end, making a recapitulation of his argument, Jakubisiak writes: "Criticism has survived to this day in its most important postulates, namely: negation of metaphysics, denial of all transcendence and bringing a richer content of scientifically significant cognition to the *a priori* forms of the human mind" [17, p. 25]. According to him, criticism has reached its extreme form in logistics, the formalism of which is in turn "the extreme stage of the current of thought, going in the opposite direction to the progress of human knowledge, instead of going from scope to content, it goes backwards through the movement of cancer from content to scope" [17, p. 25].

It did not take long for Lukasiewicz to answer [24]. In fact, the Polish logician said that Jakubisiak's attacks might be silent, if only due to the lack of knowledge of the subject he so vehemently criticizes. Łukasiewicz's reply can be summarized as follows: 1) Logistics cares for contact with reality – here Lukasiewicz refers to the text about the method which Jakubisiak criticized, but which he did not read honestly: according to logic it is necessary to verify and control the results obtained in logistics through intuition, experience and natural sciences; 2) logistics does not defend the postulates of Kant's philosophy; 3) logistics is not logical empiricism; these two points result from the fact that it is not a philosophical or logical direction, but only science, such as psychology, and this is a science closer to mathematics than to philosophy; it can be considered "at most" as a branch of philosophy – as Wolak points out, although Łukasiewicz's view of logic and philosophy and their relations has changed, it can certainly be said that he did not want to replace philosophy with logic; 4) logic is non-philosophical and does not pretend to be a philosophy – in my opinion one of the most important arguments, because logistics has its own methods that may, but do not necessarily imply philosophical theorems. Łukasiewicz notes that Jakubisiak confuses concepts by identifying mathematical logic with philosophical logic, calling it a philosophical current (the second logician considers it the pre-scientific stage of the first); for Łukasiewicz, it is clear that the current task is to create a philosophy of logic that grows out of scientific logic; 5) Jakubisiak does not touch the main point of the problem, he does not speak a word about logistics, and his reflections on the relation of scope and content cannot be called strictly logical considerations; 6) logistics is not nominalism or the analysis of language (Carnap) – according to Wolak, such an argument could be followed by Jakubisiak's whole argument; it follows from this that the charge of neopositivism is therefore erroneously put forth by Jakubisiak; 7) logistics does not negate metaphysics – according to Łukasiewicz, Jakubisiak wrongly attributes the radical views of the Vienna Circle to the LWS, confusing, in addition, Kant with Hume; once again it is clear to the Polish logician that Jakubisiak has no idea what he is writing about; 8) Łukasiewicz does not want to limit philosophical issues, but wants to improve methods of practicing philosophy, like natural science (development of logic and mathematics clearly shows that their methods are effective and fruitful in researching philosophical problems, to mention only Salamucha's book about ontological evidence); here, too one sees Łukasiewicz's remark about the priests applying logistics in their research; 9) the LWS clearly and programically distinguishes philosophy from the outlook – it means that there are such matters that cannot be examined by methods of scientific philosophy (areas that are outside the boundaries of reason are a place of beliefs and religious feelings and can pervade, according to Łukasiewicz, activity of reason). Wolak observes that the polemic between Jakubisiak and Łukasiewicz should be regarded as a worldview rather than a philosophical clash [35, p. 24], [24, p. 117].

Jakubisiak's answer was rapidly forthcoming [15]. According to Wolak, it was not a discussion – contrary to what Jakubisiak himself wrote – but a declaration, because the Polish priest omitted some of the issues raised by Łukasiewicz, while in others he made further mistakes [35, p. 44]. Jakubisiak maintains his allegation of nominalism, not accepting that logistics as a formal science is philosophically neutral. Yes, it can give reflective methodological patterns, and even give premises for philosophical reasoning, but it is not a philosophy in the strictest sense. The formalization of issues, as practice demonstrates, can be very useful in considering philosophical problems. According to Wolak, Jakubisiak commits a serious mistake by calling logistics what is only his own interpretation of logistics and, furthermore, attributing it to the entire LWS [35, p. 45]. In the latter portion of his text, Jakubisiak again equates Łukasiewicz with Carnap, not recognizing that the neo-positivists cannot be considered faithful followers of Kant's thoughts, and also that in the LWS there was epistemological pluralism. In the end, Jakubisiak strongly disagrees with Lukasiewicz's statement, who spoke and wrote that philosophical speculation should be removed. The Polish priest did not see, however, that Łukasiewicz spoke about speculations in the meaning of inaccurate and ambiguous reasoning, not about speculations conducted on the basis of the best methods of reasoning. According to Wolak [35, p. 47], Jakubisiak was in error here, but on the other hand, not knowing the details of the functioning of the different meanings of the concepts used by him, and not seeing the broader perspectives of the status of metaphysics, which the simpliciter was often refused. At the end of his answer, Jakubisiak states that modern philosophy speaks of what is on the basis of what should be, while equating Łukasiewicz and neo-positivist logicians who, however, have only seen sources of knowledge in experience. The final objection against Łukasiewicz is that he does not answer the question about the relation of logistics to philosophy at all, which results from ignorance not only of Łukasiewicz's other writings and significant omissions of fragments of his counterargument, but also the equating of multiple and diverse, and sometimes even alien views [35, pp. 48-49].

At the end of this paragraph is a brief mention, in a somewhat broader context, of the article by Łukasiewicz *In defense of logistics* [23], where he defends logistics against allegations of nominalism, positivism, pragmatism and relativism. He states that the publication of multi-valued logics in 1930 does not change the fact of the validity and ruthlessness of the principle of exclusive non-contradiction, as well as the validity of the rules of inference. It does not exclude the existence of other similar principles that may be discovered while continuing logistical and philosophical research. At the end, Łukasiewicz states that whenever he faces an issue, he has the impression of communing with a compact and resistant structure that acts on him as concrete and tangible objects:

> I can not change anything in this construction, I do not create anything myself, but in the hard work I DISCOVER in it only ever new details, gaining truths that are not touched and eternal. Where is and what is the ideal design? A believing philosopher would say that he is in God and is his thought [23, p. 26].

4. Closing Remarks

This laconic examination of the polemics between Łukasiewicz and Jakubisiak, especially in the context of today's knowledge and development of both philosophy and logic, allows us to obtain a broader understanding of the inaccuracies and shortcomings of Jakubisiak's arguments.[14] However, as Woleński notes, Jakubisiak's criticism and attacks occured in the 1930s, when the achievements of LWS were something new and not yet solidified, and many issues were not fully clarified or were only beginning to be understood [37, p. 164]. Jakubisiak's warnings about the logistic virus turned out to be unwarrented, as evidenced, among others, by the development of broadly understood analytical philosophy and logical tools in other models of practicing philosophy, but his attitude and attacks on the LWS significantly (*implicite*) contributed to the development of ideas cultivated within the sphere.

The example of this polemic demonstrates that in philosophy there is a need for theoretical clashes and discussions which, if they do not change views and positions, can significantly contribute to their clarification.

References

1. Ambrożewicz, Z. Jednostka podstawą racjonalności świata. Próba indywidualistycznej metafizyki ks. Augustyna Jakubisiaka, In Z. Ambrożewicz (ed.), *Oblicza racjonalności*, Opole: Wydawnictwo UO, 2005, pp. 79-95.
2. Andrzejuk, A. Philosophers among the lecturers of PUA (1939–2009). Profiles of Scholars, In M. Płotka, J. Pyłat, A. Andrzejuk (eds.), *Philosophy at the Polish University Abroad*, Warszawa-Londyn: Wydawnictwo von borowiecky, 2014, pp. 23-54.
3. Barbaszyński, D. Inspiracje światopoglądowe w myśli Jana Łukasiewicza, In S. Janeczek, R. Charzyński, M. Maciołek (eds.), *Światopoglądowe odniesienia filozofii polskiej*, Lublin: Wydawnictwo KUL, 2011, pp. 329-337.
4. Bochenek, K., Gawor, L., Jedynak, A., Kojkoł, J. *Filozofia polska okresu międzywojennego. Zarys problematyki*, Gdynia: Wydawnictwo Akademii Marynarki Wojennej, 2013.
5. Borzym, S. Filozofia międzywojenna (1918-1939). Przegląd stanowisk, In S. Borzym, H. Floryńska, B. Skarga, A. Walicki (eds.), *Zarys dziejów filozofii polskiej 1815-1918,* Warszawa: PWN, 1986, pp. 511-547.
6. Borzym, S. O filozofii międzywojennej w Polsce, *Znak* 329, 1982, pp. 210-241.
7. Bouchet, J.-M. L'homme le plus extraordinaire que j'ai rencontré: Augustin Jakubisak, *Cahiers de Chiré* 8, 1993, pp. 31–39.
8. Chudoba, B. *Of Time, Light And Hell: Essays In Interpretation Of The Christian Message*, Mouton: The Hauge, 1974.
9. Coniglione, F. *Nel segno della scienza: la filosofia polacca del Novecento*, Milano: FrancoAngeli, 1996.
10. Dunning, D. E. The logic of the nation: Nationalism, formal logic, and interwar Poland, *Studia Historiae Scientiarum* 17, 2018, pp. 207-251.
11. Gałęzowska, I. *Ksiądz Augustyn Jakubisiak*, Paryż: Biblioteka Polska i Towarzystwo Historyczno-Literackie, 1947.
12. Jadacki, J. Semiotyka szkoły lwowsko-warszawskiej: główne pojęcia, In M. Hempoliński (ed.), *Polska filozofia analityczna. Analiza logiczna i semiotyczna w szkole lwowsko-warszawskiej*, Wrocław-Warszawa-Kraków-Gdańsk-Łódź: Zakład Narodowy Imienia Ossolińskich, Wydawnictwo PAN, 1987, pp. 131-173.
13. Jakubisiak, A. *Nowe Przymierze. Z zagadnień etyki*, Paryż: Komitet Wydawniczy Dzieł Księdza Augustyna Jakubisiaka, 1948.
14. Jakubisiak, A. *Vers la causalité individuelle*, Paris: Société Historique et Littéraire Polonaise, 1947.
15. Jakubisiak, A. W sprawie książki „Od zakresu do treści", *Przegląd Filozoficzny* 40, 1937, pp. 90-96.
16. Jakubisiak, A. *La pensée et le libre arbitre*, Paris: Libraire J. Vrin, 1936.
17. Jakubisiak, A. *Od zakresu do treści*, Warszawa: Wydawnictwo Droga, 1936
18. Jakubisiak, A. *Essai sur les limites de l'espace et du temps*, Presses Universitaires de France, Paris: Libraire F. Alcan, 1927.
19. Kiczuk, S. *Przedmiot logiki formalnej oraz jej stosowalność*, Lublin: Redakcja Wydawnictw KUL, 2001.
20. Łukasiewicz, J. *Pamiętnik*, J. Jadacki, P. Surma (eds.), Warszawa: Wydawnictwo Naukowe Semper, 2013.
21. Łukasiewicz, J. Letter to Father Józef M. Bocheński of October 7, 1947, *Zagadnienia Filozoficzne w Nauce* 2 (18), 1997.

22. Łukasiewicz, J. (ed.). *Myśl katolicka wobec logiki współczesnej*, Poznań: Księgarnia Św. Wojciecha, 1937.

23. Łukasiewicz, J. W obronie logistyki, In J. Łukasiewicz (ed.), *Myśl katolicka wobec logiki współczesnej*, Poznań: Księgarnia Św. Wojciecha, 1937, pp. 12-26.

24. Łukasiewicz, J. Logistyka a filozofia, *Przegląd Filozoficzny* 39, 1936, pp. 115-131.

25. Murawski, R. *Filozofia matematyki i logiki w Polsce międzywojennej*, Toruń: Wydawnictwo Naukowe UMK, 2011.

26. Polak, P. Zmagania polskich filozofów z ogólną teorią względności: przypadek neoscholastycznej recepcji teorii Einsteina przed II wojną światową, In P. Polak, J. Mączka (eds.), *Ogólna teoria względności a filozofia. Sto lat interakcji*, Kraków: Copernicus Center Press, 2016, pp. 29-64.

27. Rechlewicz, W. *Nauka wobec metafizyki. Poglądy filozoficzne Kazimierza Twardowskiego*, Kielce, 2015.

28. Rolleri, J. L. Review of J. Łukasiewicz, *Estudios de Lógica y Filosofía* (1975), *Crítica. Revista Hispanoamericana de Filosofía* 58, 1988, pp. 100-109.

29. Simons, P. Jan Łukasiewicz, In *Stanford Encyclopedia of Philosophy*, retrieved from: https://plato.stanford.edu/entries/lukasiewicz/ (April 23, 2019).

30. Sosnowski, L., Jakubisiak Augustyn, In *Encyklopedia filozofii polskiej*, vol. 1, A. Maryniarczyk et al. [eds.], Lublin: Polskie Towarzystwo Tomasza z Akwinu, 2011.

31. Suchoń, W. Logika czasu kryzysu, In W. Suchoń, I. Trzcieniecka-Schneider, D. Kowalski (eds.), *Parerga z logiki praktycznej*, Dialogikon vol. 16, Kraków: Wydawnictwo UJ, 2013, pp. 41-47.

32. Surma, P. Jan Łukasiewiz o determinizmie i logice, In S. Janeczek, R. Charzyński, M. Maciołek (eds.), *Światopoglądowe odniesienia filozofii polskiej*, Lublin: Wydawnictwo KUL, 2011, pp. 339-349.

33. Syski, A. Review of A. Jakubisiak, *Od zakresu do treści* (1936), *Przegląd Homiletyczny* 4, 1936, pp. 376-377.

34. Wolak Z., Sztuka prowadzenia sporów, *Zagadnienia Filozoficzne w Nauce* 17, 1995, pp. 117-118.

35. Wolak, Z. *Neotomizm a Szkoła Lwowsko-Warszawska*, Kraków: Ośrodek Badań Interdyscyplinarnych, 1993.

36. Woleński, J. Lvov-Warsaw School, In *Stanford Encyclopedia of Philosophy*, retrieved from: https://plato.stanford.edu/entries/lvov-warsaw/ (April 23, 2019).

37. Woleński, J. *Szkoła Lwowsko-Warszawska w polemikach*, Warszawa: Wydawnictwo Naukowe Scholar, 1997.

38. Wolsza, K. Józef M. Bocheński OP (1902-1995) i środowisko „Verbum”, *Studia z Filozofii Polskiej* 2, 2007, pp. 133-156.

39. Woźniczka, M. Alternatywność konwencji kształcenia filozoficznego jako wzorzec edukacyjny, *Analiza i Egzystencja* 10, 2009, pp. 151-172.

40. Zegzuła-Nowak, J. *Polemiki filozoficzne Henryka Elzenberga ze szkołą lwowsko-warszawską*, Kraków: Wydawnictwo Scriptum, 2017.

Acknowledgements

Scientific work funded from the 2018-2021 budget for science, as a research project under the Diamond Grant Programme (Decision No 0002/DIA/2018/47 of 18 July 2018 and Contract No 0002/DIA/2018/47 of 7 September 2018). The official website of the project: www.augustyn-jakubisiak.pl.

The author would like to thank Paweł Balcerak and Tomasz Grzebyk for their valuable comments on the manuscript of this text.

Notes

[1] "As far as the matter concerns international importance, one thing is clear. The logical achievements of the LWS became the most famous. Doubtless, the Warsaw school of logic contributed very much to the development of logic in the 20th century. Other contributions are known but rather marginally. This is partially due to the fact that most philosophical writings of the LWS appeared in Polish. However, this factor does not explain everything. Many writings of the LWS were originally published in English, French or German. However, their influence was very moderate, considerably lesser than that of similar writings of philosophers from the leading countries. This is a pity, because radical conventionalism, reism or semantic epistemology are the real philosophical pearls. But perhaps this is the fate of results achieved in cultural provinces" [36].

[2] Polish Catholic priest, theologian and philosopher associated with the Historical and Literary Society and the Polish Library in Paris. He lectured and published in French and in Polish, and served as a chaplain among the Polish community and soldiers. In his intellectual work he dealt with Polish philosophy, criticism of totalitarianism, philosophy of man and freedom, ethics, as well as issues in the field of philosophy of nature and philosophy of science. He was born in Warsaw in 1884, after graduating from high school he entered the catholic seminary, which he completed in 1906 and was ordained a priest. In 1910 he travels to Paris, where he takes up philosophical studies (Catholic Institute in Paris). Two years later, he defended his doctorate in morality with Count August Cieszkowski (1912). He also wrote a dissertation on the philosophy of the absolute in the thought of Józef Hoene-Wroński, which he presented at the Sorbonne in 1914. He also continued his studies in specific sciences: mathematics, physics and chemistry (Sorbonne). This allows him to complete his work, which he wrote for many years, on time and space limits (*Essai sur les limites de l'espace et du temps*), for which he received a distinction from the French Academy of Moral and Political Sciences (1927). In the meantime, he returned to Poland as an army chaplain to General Józef Haller (1919-1920). Then he returns to France. 1936, he published a collection From scope to content and a second important work in French - *La pensée et le libre arbitre*. In the years 1939-1940, he was the first professor of philosophy at the Polish University Abroad. He also performed various pastoral, social and political functions. He dies on November 23, 1945. For further information see e.g. [11], [30, pp. 542-545]

[3] Short mentions about Łukasiewicz's polemics with Jakubisiak can be found e.g. in [31, p. 41], [9, pp. 95, 230]. Wider discussion with a broader historical-theoretical context: [37, pp. 134nn], [35, pp. 24-49]. Recently, the polemics have been mentioned in [10].

[4] It is interesting how Łukasiewicz talks about Jakubisiak in his private journal. In May 1936 Łukasiewicz was invited to lecture at the Sorbonne. Jakubisiak also came to his lectures. "I had a problem with this priest who was considered a great philosopher in the Polish circles of Paris, because he attacked me and my school in a way that was both stupid and ugly. He became frustrated when we invited him to dinner at *Lutecja*, watered with wine, but when he later read my article *Logistyka a filozofia* (*Logistics and philosophy)* in *Przegląd Filozoficzny* (*Philosophical Review*) after a few weeks, he became mortally offended" [20, p. 58].

[5] For Łukasiewicz's thought and writings see e.g. [29], [25, pp. 69-89].

[6] It should be noted that even today, thinkers of Christian (Catholic) provenance formulate skeptical judgments about logic as a tool for solving philosophical (metaphysical) problems: logic cannot be a fully adequate method of justification in metaphysics, nor can it justify all the statements made in metaphysics [19, pp. 67-70].

[7] Jakubisiak develops and finally formulates his concept in subsequent works: [13], [14], [16], [18]. Zbigniew Ambrożewicz attempts to discuss his concept synthetically [1].

[8] Synthetic development of the outlined ideas can be found in [14] or in a more popular form in [7], [4, pp. 75-77, 120-121] and [2, pp. 30-31].

[9] See also [6, pp. 215-217].

[10] It should be added that Łukasiewicz's article Logistics and philosophy has also become the subject of Henryk Elzenberg's remarks and reservations, as Joanna Zegzuła-Nowak writes in detail in his recent book, precisely in the context of Łukasiewicz's dispute with Jakubisiak [40, pp. 150-153].

[11] That is why Wolak [35] includes Jakubisiak among the neo-Thomists, although he does not do it without any reservations, which is also emphasized by Paweł Polak who characterizes Jakubisiak's philosophical silhouette in the following way: "Jakubisiak was perceived by his contemporaries as an original philosopher, building his philosophy in the spirit of the Ockham's nominalism and criticizing most philosophical positions, including scholastics and Thomism. In newer studies, accents are differently distributed in relation to his views: according to Sosnowski, Jakubisiak's interests were directed towards the sciences, and Wolak considers him (with some cautions) as a representative of the Polish neo-Thomist movement. Apart from attempts at a comprehensive assessment of Jakubisiak's position, let's keep in mind that he tried to integrate his original concept into Christian philosophy in Poland, and that his reflections on the theory of relativity were for him one of the elements of the analysis of contemporary science and philosophy, which was to serve him in the construction of fundamental concepts of new philosophy and in the criticism of Kantian apriorism" [26, p. 56].

[12] The order of the argumentation was given in favour of another text by Wolak, in which he presents Łukasiewicz's polemics with Jakubisiak as an example of a dispute conducted within the framework of Schopenhauer's eristic [34].

[13] I'm basing the following reconstruction especially on [35, pp. 24-49] which is more hermetic for the subject than [37].

[14] For more detailed discussions see [37, pp. 134nn], [35, pp. 24-49].

Studia Humana
Volume 8:2 (2019), pp. 25—31
DOI: 10.2478/sh-2019-0015

The Analogy in Decision-Making and the Implicit Association Bias Effect

Nataliia Reva

Taras Shevchenko National
University of Kyiv,
Volodymyrska 64/13 Street,
01601 Kyiv, Ukraine

e-mail: natalie.reva@gmail.com

Abstract:
The author stands that thinking by analogy is a natural instrument human have because of the mirror neurons in our brain. However, is it that infallible to rely on? How can we be sure that our hidden biases will not harm our reflections? Implicit Association Bias (IAB), for instance, is a powerful intruder that affects our understanding, actions, and decisions on the unconscious level by cherishing the stereotypes based on specific characteristics such as ethnicity, sex, race, and so on. To check if there is a correlation between the IAB effect and the people's capacity to reason logically, the author had created an online-survey. The focus was on analogical reasoning and IAB tests concerning the question of gender equality in science and everyday life and age prejudices.
Keywords: analogy, analogical thinking, Implicit Association Bias, IAT.

1. Introduction

Human makes decisions a thousand times a day using different apparatus to help themselves. Often we consider a new situation for the one they know. It is in our nature. From our childhood, babies try to copy their parent's habits, their facial expressions, and gestures, their manner of talking, their walk and posture – the first thing they get from their moms and dads. Cognitive scientists say that this type of behavior is possible thanks to the mirror neurons in our brain, which are focused on finding similar patterns.

Nevertheless, this copying is unconscious. Small child, as well as apes or parrots, imitate what they see and hear without re-thinking. Growing up, they find the new role models and start to analyze and compare their parents with the new idols choosing the elements they like more. At this point, the child begins reason by analogy consciously.

The ability to spot existing or emerging patterns is one of the most (if not the most) critical skills in intelligent decision making, though we are mostly unaware that we do it all the time [14]. Human brain works by patterns and associations – if a perception fits roughly into an existing pattern, then it may be taken as definitive. We see a half-hidden person dressed or coiffured as someone we may know and "recognize" this person by his/her type of clothes or hairstyle. We distinguish a fake smile from a genuine smile, predict from person's body language if he/she is telling the truth or read from the facial expression what people are thinking about at this particular moment.

This ability of our brain makes our life easier, but at the same time, it leaves some loops. The not only analogy works in this manner. Not so far, American scientists, Kahneman and Tversky, have discovered the whole series of biases that imperceptibly intrude into our decision-making process. For instance, Implicit Association Bias (IAB) arises from the quick automatic association by noticing patterns between two or more similar things, i.e., creates rapid mental connections between the objects, actions, and ideas that share the same patterns as well as the analogical reasoning. In this article, we are going to take a better look at both of them to find their similarities and differences.

2. Analogy

In early 80s Dedre Gentner developed the structure-mapping theory according which an analogy is a mapping of knowledge from one domain (the base or source) into another (the target) such that a system of relations that holds among the base objects also holds among the target objects. Meaning that analogy works by establishing the correspondences between two objects, so the new inferences derive by importing connected information from one object to another. To do so, these objects should have some common patterns [4].

Analogical mapping is often used because of its simplicity and familiarity for our brain. However, sometimes, it requires a good level of creative thinking. There is a small number of analogies that can be taken from our memory. Basically, because our memory is not limitless. Besides, it is in human nature to forget things. Thus, the other analogies suggest that we use some "unexpected" sources, like the creativity of our mind. The last one is claimed to be the origin of novelty, which analogical reasoning is glad to benefit from.

Gentner and Loewenstein insisted that analogical reasoning as a reflection process from particular to particular can be divided into simple stages. They distinguished 4 steps that people "pass" reasoning by analogy: (1) retrieval of potentially useful related case given another (2) mapping between two cases in working memory (finding the correspondence/ likeness) (3) evaluating the analogy and its inferences (using a source analog to form a new conjecture) (4) abstracting the common structure (good analogy is structurally consistent). [3]

Holyoak and Thagard supporting Gentner's idea stressed that good analogical reasoning follows three kinds of constraints: similarity, structure, and purpose. They do not operate like rigid rules but assist the internal coherence of the analogical reasoning. Returning to the Little Aaron, his example adheres to all restrictions. Aaron's mom hit her hand as her son done for hundreds of time (similarity). The boy kissed his mom's hand she had done before (structure). Aaron wanted to make his mom feel better by easing her pain by a kiss (purpose) [6].

As straightforward and widespread analogical thinking is, it also proved to be useful. Gick & Holyoak conducted an experiment whose results reported some curious data. Only 10% of people who simply read and tried to solve the problem succeeded. While 30% of those who were given a story with an analogous solution, yet, with different specific content, before receiving the insight problem, solved it. Three times as many as without the analogy! Seems incredible! [9].

Using analogical reasoning in decision-making process simplifies the last one, proposing to use some already established model instead of creating a new laborious resolution. Thus, the more previous experiences one have, the more connections this person easily can make. Analogies may be applied at various levels: in the same case or a case with similar structure; in social interaction with the same individual or with individuals who are considered analogous (e.g., are in similar relations to me, like family or team members), etc. [13].

However, if analogies are so quick and easy-made, moreover based on our elusive memory and previous doubtable experience, how can we be sure that analogical reasoning is infallible to rely on? How can we be sure that our hidden biases will not harm our reflections? Does our level of critical and logical thinking preserve us from errors? Marijke Breuning insists that analogies can fail if and only if they are constructed based on superficial similarity, not deep causal traits [1].

3. Implicit Association Bias

Implicit stereotyping is systematically studied through well-established methods based upon principles of cognitive psychology that have been developed in nearly a century's worth of work. The IAB is demonstrated in two paradigms: (1) says that the cognitive salience of a familiar stereotype can implicitly bias social judgment in stereotype-consistent ways (Devine's critical experiment); (2) states that social attitudes – including prejudice and stereotypes – are empirically captured by the degree to which they are linked through speed and efficiency to semantically related concept [11].

IAB has the next characteristics: (1) It belongs to the I System of human "Machinery of Thought" that is represented by the quick automatic mode of decision-making. We can identify it with human intuition and instinct [12]; (2) It is unconscious, so humans are unable to catch its presence at once. Just as Freud suggested that we push our sexual troubles and traumas out of consciousness, yet they continue to follow us and have an influence on us in the form of our dreams, linguistic errors or even some kind of depression. Cognitive implicit biases hide in the dark corners of our mind waiting for the right time to show their effect on us; (3) It works on the rapid mental associations attached to people behavior and attitude; (4) It can contradict human conscious beliefs and positions.

Where can we see the manifestation of the IAB? Literally everywhere! It could happen in any domain: recruitment, healthcare, outcomes in criminal justice, etc. For example, if meeting a person for the first time, you, rather than being neutral, have a preference for (or aversion to) he/she based on such characteristics as race, gender, ethnicity, age, or even appearance; this is the manifestation of the implicit association bias. I think anyone of us had in our experience a person who was more loved by our teacher/boss or whoever else only because he/she had some characteristics this person likes or on the contrary you have some flaws this person hates.

However, you should understand that it does not make you or anyone else a racist, sexist, etc., anytime such a stereotype popped into your mind. It just means that your brain is working properly, noticing patterns, and generalizing! Racism or sexism is a decision made by a sharp mind. The implicit associations are 'rogue' processes, which are not properly seen as part of the agent's character - not indicative of 'who she is.' Merely being influenced by implicit bias does not mean that one has the nature of a racist or sexist person; it takes something other than the operation of implicit racial biases to be properly ascribed the character trait racist [7].

IAB have both positive and negative results. From the bright side, IAB facilitates fast-made judgments and decisions. Although it does it by undermining the true intentions, changes the behavior, and sets people up to overgeneralize. For instance, imagine a police officer that believes in his superior role to protect and serve people. He is deeply committed to these principles. Yet, most of the time, he stops only men of color. If you ask him why he is doing that he will not be able to give you a rational answer. The truth is – he is biased! Unconsciously he associates a black or a brown face to the criminal one (without being aware of it). This police officer suffers from the implicit race bias that could have appeared in his childhood or forced by his everyday environment.

At the same time, we can have some harsh prejudices against a government, for instance, associating it with the word corruption, or bankers associating them with greed, or militaries associating them with aggression. Thereby when you meet a man, who claims to be a banker, but dreams of going to politics, you may unconsciously associate this desire with his greed and corruption inclinations. That is how the implicit bias can penetrate the human decision-making process and affect or even modify the decision. The good news is that any bias could be corrected by simple re-thinking! When you get to know better this man, you can change your mind, finding him kind and fair, or confirm your first assumptions.

4. Analogical Reasoning and Implicit Association Bias Experiment

Our brain is a powerful machine that knows how to simplify the word for us. Just imagine if we would need to think of everything in a logical, meticulous way comparing all the propositions and desires. It would take a lot of time and mental energy. Therefore, our brain constructs a vast amount of models for quick understanding and processing the information in our memory, so we would not have to deal with it later in the future one more time. The same principle works in so-called "Holistic learning," where you learn things by connecting them to other ideas and creating mental constructs of concepts. [17]

Nevertheless, nothing is perfect! Even our brain. That is why the loops open for the biases in the System I become possible. As it was mentioned before, IAB works in the same manner as the analogical reasoning by gathering commonalities together. That means that IAB like a virus or a tree fungus clings to our brain and unnoticed functions with it. Let us sum up in the table below the main characteristics of the analogical reasoning and IAB.

Reasoning by analogy	**Implicit association bias**
The reasoning is typically considered with a high-level awareness and rationality that belongs to System II.	Manifest themselves using the loops of the System I, creating rapid mental connections between the objects, actions, and ideas.
Identifies a common relation between two situations and generates further inferences driven by these commonalities.	Arises from the quick automatic association by noticing patterns between two or more similar things.
Consciously makes generalizations to come to a specific decision.	**Unconsciously** makes generalizations to come to a certain decision.
Rational good analogies are structuralized.	Unconscious associations are driven by indefinable emotional impulses.

As these two phenomena work similarly, the next questions arise: (1) could the Implicit Association Bias intrude our Analogical reasoning? (2) Could the stereotypes or prejudices take the place of the rationally made analogies in the name of fast thinking? My hypothesis is that IAB not only can but also do so quite often. Therefore, there should be a correlation between them. We can assume that if there will be found a robust and significant correlation (r = more than .05) between the level of Analogical reasoning and the IAB a person shows, we may say that human analogical reasoning (sufficiently) suffers from the unconscious impulses.

5. Method

To check the hypothesis, I have created an online survey on the Lime Survey platform. The study was performed in Ukrainian language, so here I am giving you the translations. It was composed of three parts: two on the analogical reasoning test and one on the implicit association bias. The analogical test part was run twice (before and after the IAB) to see if the implicit association bias mutually with the pressure of time affects human decisions. The study run in the next sequence: (1) First, the participants have as many time as they need to reflect on the ten questions on analogy. (2) After they do the IAB test (that is limited in time) to rate the level of gender and age stereotypes they unconsciously have. (3) Then they have a new analogical test, and they need to answer these questions as quickly as they can accordingly to the timer (10 seconds per question). Besides, in the last task, five questions out of ten concerned the same topics as the IAB test, i.e., have gender and age implication.

The original idea to take the IAT was rejected, firstly, because of its complexity and long time-consuming; secondly, because the new studies showed its shiftiness. [2, 15] Thus, I decided to create my test that will not take much time and will be simpler to do, yet, that will be still based on the same principle as the IAT – the association test on millisecond reaction time. For example, to see the gender preference unconscious, I named some professions, like an astronaut, first-grade teacher, nurse or mathematician, to the students and asked them to choose to whom it is more suitable – for a girl named Olia and a boy called Tolia. To check the race prejudice, I gave the students some examples of the presents, like a book, laptop or a bike, and asked to choose which of them are good for a son and which for his grandpa. Subjects had only 30 seconds to make their decision.

Additionally, I decided to check Holyoak's assumption that analogical reasoning is a congenital ability. Holyoak & Thagard gave an example of little Aaron who at his second year of life was able to derive an analogy from to similar situations, that shows that analogical reasoning does not require any tutoring in logic or critical thinking [8]. Thus, I invite not only people who studied logic, but also those who never had a deal with critical thinking. The total number of the participants is 50 (25 from each side). All of them are students from my alma mater – Taras Shevchenko National University of Kyiv. Half of them – the third grade – had attended a course on classical logic. The first grade did not have any logical classes at that time or before.

After the revision, only half questionnaires turned out to be competently and correctly full-filled. Seven of the participants were male subjects, while eighteen were female subjects. Only two representatives out of 25 said that they prefer men to women in work, while four gave their preference to women. At the same time, 12 represents indicated that they prefer to work with young people and none favored elders. An interesting fact is that all four who suggested their preference to women showed prejudice against them in the question of career and profession.

6. Conclusion

Out of 100 answers on modified IAT test, 28 did not show any biased, while 72 were the biased answers. Overall, 50% of the representatives showed a gender-bias and 70% – the age-bias! For instance, *all the respondents* (100%) recognized the primary school teacher as a female profession, while 86% of participants chose the astronaut as a job for men. At the same manner, the subjects decided that the exact sciences, like mathematics and astronomy, are more suitable for men (72% and 80%) while the Humanities and Art fit the women (89% and 95%).

Talking about analogical thinking, all subjects made more mistakes in the second part after they pass the IAB test. If we analyze the two analogical tests, out of 250 answers, only 69 were incorrect in the first test compared to the 135 in the second one. Looking at these results, we can previously agree with the hypothesis that in time-limited circumstance the implicit association bias easily intrudes the decision-making process and to save time replace the analogical reasoning simultaneously pushing people to make the wrong choices.

Moreover, in the three of five questions that had gender or age stereotypes, the results showed in average 60% of biased answers. For instance, to the question "apple tree: apple: father:?" only five people gave the right answer "a child" when all the other chose the wrong variant – "a son." At the same way, to the question "fast: slow: immature:?" only three participants selected the correct answer "mature/developed," while seventeen chose "aged," the rest also gave wrong answers "young" or "green," probably not understanding the analogy of this task.

Statistics were done in the SPSS. It reveals a significant correlation (Pearson's r = .06) between the level of implicit association bias people have and their use of analogical thinking. As a result, the subject's answers showed that the author's hypothesis was right. Besides, preliminary study or non-study of logic did not affect the test results. Thus, we can conclude that Holyoak's theory of analogical thinking as natural human ability may have sense. For future researches, the author plans to check it in correlation with the other social biases and heuristics, like, for example, the anthropomorphism.

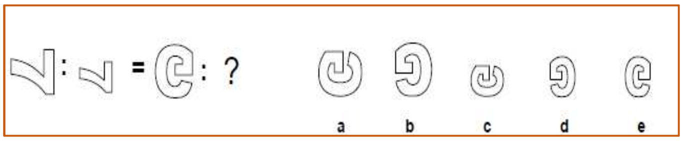
References

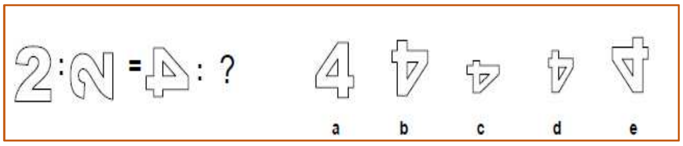
1. Breuning, M. The Role of Analogies and Abstract Reasoning in Decision-Making: Evidence from the Debate over Truman's Proposal for Development Assistance, *International Studies Quarterly* 47 (2), 2003, pp. 229-245.
2. Brunel, F., Tietje, B., Greenwald, A. *Is the Implicit Association Test a Valid and Valuable Measure of Implicit Consumer Social Cognition?* 2003.
3. Gentner, D., & Loewenstein, J. Learning: Analogical Reasoning. *Encyclopedia of Education, Second Edition*, New York: Macmillia, 2003.
4. Gentner, D. & Maravilla, F. Analogical reasoning, In L. J. Ball, V. A. Thompson (eds.), *International Handbook of Thinking & Reasoning*, New York: Psychology Press, 2018, pp. 186-203.
5. Gentner, D. & Smith, L. Analogical reasoning, In V. S. Ramachandran (ed.), *Encyclopedia of Human Behavior* (2nd Ed.), Oxford: Elsevier, 2012, pp. 130-136.
6. Gick, M., Holyoak, K. J. Analogical Problem Solving, *Cognitive Psychology* 12, 1980, pp. 306-355.
7. Holroyd, J., Kelly, D. Implicit Bias, Character, and Control, *Personality to Virtue*, Oxford University Press, 2014.
8. Holyoak, K. J. Analogy and relational reasoning, In K. J. Holyoak, R. G. Morrison (eds.), *The Oxford handbook of thinking and reasoning*, New York: Oxford University Press, 2012, pp. 234-259.
9. Holyoak, K. J. & Thagard, P. The Analogical Mind, *American Psychologist* 52 (1), 1997, pp. 35-44.
10. Hristova, P. Unconscious Analogical Mapping? *New Bulgarian University* 2009, pp. 655-660.
11. Jost, J. T., Rudman, L. A., Blair, I. V., et al. The existence of implicit bias is beyond reasonable doubt. Review, *Research in Organizational Behavior* 29, 2009, pp. 39-69.
12. Kahneman, D. *Thinking Fast and Slow*, New York: Farrar, Straus, and Giroux, 2011.
13. Kokinov, B. Analogy in decision-making, social interaction, and emergent rationality, *Behavioral and Brain Sciences* 2003.
14. Miemis, V. Essential Skills for 21st Century Survival: Part I: Pattern Recognition, Retrieved from: https://emergentbydesign.com/2010/04/05/essential-skills-for-21st-century-survival-part-i-pattern-recognition/.
15. Nagai, A. The Implicit Association Test: Flawed Science Tricks Americans into Believing They Are Unconscious Racists, Special Report no 196, The Heritage Foundation, 2017.
16. Reva, N. Logic, Reasoning, Decision-Making, *Future Human Image* 10, 2018, pp. 76-84.

17. Young, S. A Brief Guide to Learning Faster (and Better), 2011, Retrieved from: https://www.scotthyoung.com/blog/2011/01/11/learn-faster-and-better/.

APPENDIX

Visual analogy tests





Verbal analogy tests

? : tradition : hedonist : pleasure	**outlaw : ? : offend : affront**
a. purist (term) b. Eden c. displeasure d. Agnostic	a. chase b. police c. crime d. forbid (synonym)

IAB analogy test

#1 fast: slow: immature: ?	**#2 newborn: diaper: ?: coffin**	**#3 apple tree : apple: father: ?**	**#4 unclear: clarity: ?: flexibility**	**#5 dentist: teeth: ?: money**
1. developed 2. aged 3. old man 4. green	1. undertaker 2. old man 3. thief 4. dead	1. wife 2. son 3. child 4. boy	1. flexible 2. hard 3. young 4. straight	1.businessman 2. bank 3. accountant 4. lawyer

Modified IAT test (30 seconds per question)

Who would you recommend for the astronaut's position in NASA?	**Whom would you recommend to give a new IPad for a New Year?**
Olia Tolia	Grandpa Son

Studia Humana
Volume 8:2 (2019), pp. 32—46
DOI: 10.2478/sh-2019-0016

About Possible Benefits from Irrational Thinking in Everyday Life

Magdalena Michalik-Jeżowska

University of Rzeszów,
Rejtana 16c Av.,
35-959 Rzeszów, Poland

e-mail: madelaine.mmj@gmail.com

Abstract:
In this work, no denying the role, or even more so, the value of rational thinking, it is assumed that it is not the only effective tool for man to achieve his valuable goals. It is conjectured here that sometimes irrational thinking is an equally good (and sometimes even better than rational thinking) means of achieving them. In the light of these assumptions, the goal of my work is to indicate the benefits that may be the result of irrational thinking in the colloquial (i.e. unscientific) domain of everyday human practice. The given examples of irrational thinking come from research in the field of cognitive and social psychology and behavioural economics. Their results prove that irrational behaviours (including thinking) are neither accidental nor senseless, and on the contrary systematic and easy to predict, they constitute important arguments for considering the phenomenon of irrational thinking. I also discuss this issue although only to a limited extent.
Keywords: rational and irrational thinking, cognitive psychology, behavioural economics, morality.

1. Introduction

It has been assumed that thinking, including its subtype – reasoning,[1] is crucial for the effective, everyday functioning of people. As such, it is supposed to increase the probability of undertaking an optimal action to achieve valuable goals set by man. Although what has been written applies to thinking in general, it refers in particular to rational thinking, because in our culture rationality is considered a desirable value and at the same time a norm for most (if not all) types of human activities. In the context of the cultural depreciation of irrationality,[2] the idea to consider the potential benefits of irrational thinking can therefore look like an intellectual provocation. The commonness of irrational thinking indicated by many researchers[3] as the basis for drawing conclusions and making decisions in everyday life proves that this is not the case. Of course, the commonness of a phenomenon cannot be an argument in its favour. In particular, it cannot prove its merits. Nevertheless, considering the case from the evolutionary point of view, persistently repeated behaviour (here: the commonly occurring tendency to irrational thinking) can be seen as a useful (and therefore evolutionarily conserved) adaptation to environmental requirements.[4] Accepting this point of view in this work, I focus on looking for possible benefits that irrational thinking can bring as a means of achieving specific individual goals in the sphere of everyday life.

2. The Issue of Distinguishing Acts of Rational and Irrational Thinking

Indication of possible benefits from irrational thinking is a challenge, because it is very difficult and sometimes impossible to accurately distinguish acts of rational thinking from acts of irrational thinking. This difficulty is connected with the fact that "rationality" as well as "irrationality" in accordance with the new paradigm currently in development are not completely disjunctive or mutually contradictory concepts. On the contrary, the new approach to the issue of rationality and irrationality emphasises the relative character of the opposition of both concepts and their mutual, complex relationships. This corresponds to the currently advanced model of human nature and psyche, according to which "man is not – and cannot be – a being that is fully and consistently rational (…) [He is – MMJ] a complex mental: cognitive-emotional and emotional-impulsive structure (…) [which is – MMJ] internally omnifariously intertwined, that is, there is a mutual interaction of components, rational ones affect emotional ones, emotional components affect rational ones (this direction of impact is generally stronger), impulsive ones impact on emotional ones" [17, p. 90]. The distinguished, qualitatively different, integral components of the human psychic structure appear in it with varying intensity and dimension. Their mutual relationship depends, among others, on the phylogenetic and ontogenetic level of human development, culture, personality and needs, and the specific situational circumstances of thinking and acting. More importantly, although different types of human activity entail mutual relations, varying in scope and proportions, between the above-identified structural elements of the human psyche, no sphere of human activity can be said to be the exclusive domain of rationality or irrationality.[5]

In the face of the syncretic character of human thinking,[6] signalled above, the issue of adequate rationality criteria seems to be a serious problem. Traditionally, rational thinking is perceived as a methodical activity, focused on cognition, that meets clearly defined criteria. In the context of the irrationality of colloquial thinking considered here, the criteria of rationality, assigned to its three types – logical rationality, pragmatic[7] and practical rationality – seem to be important. The test of logical rationality is the consistency of adopted premises and the deductive character of reasoning. In the domain of pragmatic rationality, efficiency is important and sufficient. In turn, the criteria of practical rationality are: goal orientation, preparation through prior reflection and efficiency.[8] All the above characteristics and at the same time requirements of rational thinking are imperfect criteria, therefore they should not be always treated as reliable indicators of rationality of a given piece of thinking.[9]

3. Suggested Additional Definition of Irrational Thinking

In the face of the indicated difficulties with distinguishing rational and irrational thinking is it possible to consider the benefits of one or the other? I think so. Clearly conclusive criteria seem to be possible only in a world with little complexity. Social reality, as the domain of everyday life, and thus the conceptual apparatus describing it, lacks this feature. Therefore, should we depart from its conceptual categorisation? Of course not. Even if our judgments about the rationality or irrationality of a particular act of thinking have to be, to some (even a minimum) degree an estimation, the conceptual order obtained in this way is conductive to better human orientation in the world and functioning in it. On the other hand, perhaps a more fruitful strategy would be the qualification of thinking in terms of its rationality or irrationality, focused on its outcome and/or its procedure. Both possibilities are indicated by the definition of irrational thinking proposed by Cezary Mordka. Under this definition, irrational thinking is "any kind of thinking that does not solve the motivational problem or solves it inconsistently with the accepted criterion (such criteria as:

economy, simplicity, fruitfulness in the sense of predictability, etc.), i.e. in a non-optimal way" [10].[10]

In this work attention is focused on these cases of irrational thinking which, although they lead to solving the motivational problem, they do it in a non-optimal way. In other words: for the purpose of this work, irrational thinking is understood as thinking containing (first of all) systematic errors in reasoning. These may be very different errors, e.g. they may consist in: not taking into account all relevant premises,[11] overestimation (not necessarily conscious) of a selected piece of information,[12] or in concluding on the basis of clearly insufficient premises[13] or even inadequate information, such as "red-haired people are deceitful". I assume that these and other errors in thinking that will be considered in this work may benefit the person who commits them, although often this potential profit can be seen only in a slightly wider perspective than the context of a particular motivational problem.

4. General Reasons for Irrational Thinking

Before I move on to consider some errors in reasoning, resulting in irrationally made decisions, five general reasons for irrational thinking indicated by Stuart Sutherland will be discussed.[14] These general motives inform us about the benefits that are associated with this kind of thinking. As such, the reasons highlighted below can be seen as arguments suggesting that irrational thinking is not an accidental freak of nature or an incomprehensible deviation from rational thinking treated as a model and norm, but rather a kind of sensible mechanism that optimises, together with rational thinking, the human decision-making process and the resulting action.[15]

According to the first explanation derived from evolutionary psychology, the animal ancestors of man due to living in a very unfriendly environment usually had to act hurriedly – fight or flee. In this situation, reflection was an ill-advised strategy, reducing the possibility of survival. From the point of view of this most important goal (survival) it was better to quickly make the wrong choice (here: escape when there was no danger) than none (due to too long reflection). This explanation would explicate why people act according to set patterns[16] in stress or rush, instead of considering all the circumstances of the case. Why has this irrational mechanism survived? Because in our society, survival (and at the optimum level) does not require only rational decision making.[17]

The second general reason for the irrationality of thinking is related to the structure and functioning of the human brain, in particular to the nerve cell networks. Initially these cells are connected together at random. In the process of learning, some of the connections are strengthened while others are weakened. Mastering a given concept, e.g. "house" or "bird" means that it is represented by activating many cells scattered over a vast area of the brain that form a certain system. "The cells that are activated fire simultaneously (…) so that processing is very fast (…) moreover, such systems of cells generalise readily. If presented with a number of different birds, they will classify as a bird a member of a species not previously shown" [15, p. 307]. Just like every mechanism, this one also turns out to be unreliable sometimes. Because the same cells participate in learning of different things, as a result of acquiring new material, sometimes the previous connections change, and (generally) small errors can happen. The existence of such systems would explain errors caused by the availability and the halo effect because in both of them man pays too much attention to the most striking feature – i.e. the one that at the cellular level corresponds to the activation of these cells between which there are the strongest connections. Despite possible errors the functioning of this data processing system is beneficial for us, because it is quick, effective and effortless due to its unconscious character.

The third reason for irrational thinking is directly connected with mental laziness. An effective way to avoid strenuous and prolonged mental effort are heuristics – "ways of thinking that will usually produce a passable but not perfect result quickly" [15, p. 308].[18]

The fourth reason for irrational thinking is the inability to use elementary probability theory, statistics and derived concepts, which is largely the result of the current education system. Sutherland believes that this inability is responsible for the error of not knowing the principle of regression to the mean, according to which "if an event is extreme (either way), the next event of the same kind is likely to be less extreme. It affects all events in which chance plays a role" [15, p. 252].[19]

Sutherland's last reason for irrational thinking and behaviour is self-serving bias expressed, among others, in the desire to show that one is right or to raise one's self-esteem. This bias combined with other factors would explain the unwillingness to reject a hypothesis one has accepted, as well as aversion to changing one's own wrong decision, and even persistent failure to notice the disadvantage of a purchase.

An interesting supplement to the presented general reasons of human irrationality is the concept of "haphazard brain" proposed by Gary Markus. Starting from an evolutionary perspective, Marcus writes about two main systems of thought coexisting in man – the ancestral system, also called the reflex system and the deliberative system. The ancestral system as evolutionarily older than the deliberative one is found in virtually all multicellular organisms. It performs its tasks quickly and automatically, consciously or unconsciously. It administers many of everyday behaviours such as the automatic adjustment of the step to an uneven surface or sudden recognition of an old friend. In its operation this system depends on evolutionarily old brain structures – the cerebellum and basal ganglia responsible for motor control and the amygdala responsible for emotions.[20] Marcus emphasises that we should not assume that the ancestral system is inherently irrational. In his opinion this system "likely wouldn't exist at all if it were completely irrational. Most of the time, it does what it does well, even if (by definition) its decisions are not the product of careful thought" [9, p. 64].

The other system of thinking assumed by Marcus – the deliberative one "deliberates, considers, chews over the facts – and tries (sometimes successfully, sometimes not) to reason with them" [9, pp. 63-64]. This system "consciously considering the logic of our goals and choices" is evolutionarily young, and hence if found in other species, it is only in few. Perhaps it is characteristic only for humans. According to Marcus's presumption, this system has its cerebral location mainly in the forebrain, in the prefrontal cortex.[21] The aim of calling it "deliberative" and not "rational" is to emphasise the lack of guarantee as to the quality of the results of its work, i.e. the real rationality of its considerations. Despite its intelligence, this system often settles for reasoning that is less than ideal. Moreover, although it is more evolutionarily advanced, it has not taken complete control of the cognitive process, because it almost always relies on indirect information, which, coming from the not really objective ancestral system, may not constitute a balanced set of data from which the deliberative system could carefully draw rational conclusions. Worse, in a situation of stress, fatigue or distraction (and therefore when a reliable analysis is most needed), the individual deliberative system usually switches off, giving way to the primitive reflex system.

Gary Marcus maintains that, from the point of view of the rationality of human thinking and functioning, a serious problem is the way in which the systems he identifies interact with each other. In theory, the deliberative system worthy of this name should be

> above the fray and unbiased by the considerations of the emotional. (…) [As such – MMJ] it would systematically search its memory for relevant data, pro and con, so that it could make systematic decisions. [It would be also – MMJ] attuned as much to disconfirmation as confirmation and utterly immune to patently irrelevant information (…) Such a system [would be also able to – MMJ] (…) stifle violations of its master plan. (<<I'm on a diet. No chocolate cake. Period>>) [9, p. 103].

Unfortunately, the above description of the deliberative system is a catalogue of wishful thinking, for which three circumstances are responsible: the relative "youth" of the system, its "building materials" which are inadequate old parts (e.g. contextual memory) and the lack of true independence from the ancestral system, which only partly takes into account the general goals of the organism.[22] In the light of the outlined concept, the irrationality of human thinking is the result of far from perfect cooperation between the brain structures (the ancestral and deliberative systems) that manage the cognitive functioning of man.

5. Benefits Resulting of Irrational Thinking in Everyday Life

Recent studies in the field of behavioural economics provide a body of evidence for the irrational nature of human decisions, and thus indirectly the irrationality of thinking.[23] This new field of research, based on psychology and economy, rejects the assumption about the fundamental rationality of human decisions, attributed to neoclassical economics.[24] According to this criticised assumption, people make decisions on the basis of the information available to them, they can calculate the value of various options that they choose from (e.g. using the probability theory), they are able to understand the consequences of any potential choice. Thus characterised actors are presumed to be making logical and sensible decisions and even if they happen to make a wrong decision from time to time, they quickly learn from their mistakes either on their own or with the help of market forces.[25]

The presented assumptions of neoclassical economics correspond to the old understanding of human nature, as a structure essentially (or even exclusively – Plato, Descartes) rational and therefore predestined for "functioning as a logical machine."[26] Researchers from the field of behavioural economics note that the assumption about the rationality of human choices is contradicted by the observed anomalies occurring when market participants make decisions. Neoclassical economics could only "explain"[27] them if they were rare and/or accidental. The problem is that, as research shows, irrational behaviours are neither accidental nor senseless. On the contrary, they are systematic and easy to predict. It is claimed here that "people are susceptible to the influence of the immediate environment[28] [the so-called context effect – MMJ], emotions, short-sightedness,[29] and other forms of irrationality" [1, p. 287] resulting in systematic errors in the decision-making process. Emphasising the regularity and predictability of cognitive errors (*nota bene* the basic concept of behavioural economics) creates space for the development of countermeasures, a kind of "glasses", correcting the picture "distorted" by someone's vision defect. Behavioural economists believe that the procedures[30] developed thanks to analysing the results of their research will help prevent people from making irrational decisions that result in such behaviours. To this end, they design experimental research to determine how to achieve the correction of the cognitive error as systematic as the error itself. For the purposes of this article, only the experimental examples of human irrationality are relevant, so I will limit myself to them here. Out of many constantly committed errors in thinking, resulting in an irrational decision, those that can be seen as beneficial have been chosen (according to the subject of this text).[31]

Several studies designed by behavioural economists aimed to establish real relationships between wages, motivation and results at work. These studies tested the common sense, proper to neoclassical economics, thesis that higher motivation (here corresponding to a higher bonus) causes increased effort, resulting in achieving the established goal (here better results at work). The research results showed that the above reasoning was wrong. It turned out that small and medium bonuses result in improvement of the performance, while very high ones, on the contrary, mean "overmotivation," i.e. a state of increased motivation pressure causing distraction, and as a result, a sudden deterioration of the achieved results. Further research showed that the negative impact of a very high performance bonus is related to the type of rewarded activity. It turned out that the more cognitive skills a given job required, the more likely the fiasco of the expected results was. On the other hand, when the rewarded activity was purely mechanical a very high bonus resulted in increased efficiency.[32] Commenting the above research, Dan Ariely stresses that the negative impact of a high bonus is related to the increased stress experienced by the employee. This observation suggests that "our tendency to behave irrationally and in ways that are undesirable might increase when the decisions are more important" [2, p. 63].[33] In other words: a fully rational action is more likely when decisions are made about abstract or less important matters. In such matters, the mind has the best conditions for cool, detached and objective concentration on the circumstances of the case.[34]

In the inference scheme which would correspond to the preliminary assumptions adopted in the above studies (i.e. the tested hypothesis), a high bonus for improving performance at work as a premise for action leading to this improvement would result in the actual enhancement of the results in question. For this to happen, it would have to really motivate people to put in more effort and this multiplied effort would have to be effective. The motivational role of a high bonus seems to be unquestionable. The situation is different with the efficiency of the effort. To say the obvious: the potential capability for increased and adequate effort (e.g. guaranteed by the education and/or experience) is a necessary condition, but not sufficient to achieve the established goal. Situational factors are always important, but some of them are difficult to predict and therefore they cannot be taken into account earlier in a rationally carried out analysis aimed at estimating the probability of achieving a given goal. The conducted research confirmed the significant influence of situational factors, ignored by the neoclassical economics, (here: increased stress corresponding to a high stake and the type of the task). Thus, while neoclassical economists actually thought that for more productive work, apart from individual disposition, a high bonus would suffice, their reasoning was irrational as a result of not taking into account all the relevant premises. On the other hand, if, as it has been suggested, it is impossible to take into account in reasoning all premises relevant to a given issue, a rational strategy may be to stop at what is undisputed (here: the motivational nature of a high bonus). Sometimes the belief that "where there's a will there's a way", usually overly optimistic, is confirmed in practice, proving that a person really determined in his actions is in some way independent of what in "normal circumstances" would surely limit him. It may happen, therefore, that for someone who really needs money[35] a high bonus, triggering extraordinary determination, will become a sufficient condition for effective action.

We can use the presented studies proving the difficulty of the rational estimation of the probability of achieving the desired result (here: a significant improvement in the performance of a given task), citing Rafał Krzysztof Ohme, to emphasise benefits we all derive from irrationality. He notes that

> thanks to the fact that we are irrational, it is impossible to totally predict our behaviour (...) and thus we cannot be controlled. (...) Secondly, due to irrationality in our naivety we do not know that something is impossible. Knowledge about lurking difficulties does not encourage us to change the *status quo*. However, discoveries, inventions and innovations are born thanks to the questioning of the existing state of affairs. It drives the development of mankind. Irrationality is adaptively desirable because it offers security and develops civilization. Although it works against reason, it is undoubtedly the work of our mind [13 p. 13].

What has been written so far indicates that rational thinking does not have to be the best or the only tool for making important decisions. This conclusion has strong empirical confirmation in experiments devoted to supportive behaviour.

It has been known for a long time[36] that lending support to another person or a group of people is greatly affected by the potential "donor's" emotions, especially his ability to feel compassion for those in need and/or his disposition to feel empathy. Experimental studies in the field of social psychology have established that whether help will be provided or not is greatly connected with the whole situational context in which supportive behaviour is desirable. This context consists of, among others, the features: of the potential donor (e.g. his current mood, whether he is in a hurry or not, etc.), the situation itself (the place of the incident – a city or a village, the number of witnesses, etc.) and the recipient (his affinity or friendship or only similarity to the potential "donor,"[37] his appearance, which may suggest the need for help or lack of it[38]). Simply put, it can be said that all these circumstances of the case result in creating or not a compassionate attitude towards the victim and, consequently, in giving or not giving help. Other studies,[39] in turn, confirmed the influence of what sociologists call "identifiable victim effect" on the occurrence of assistance activities, in the form of payment of a donation to the needy. These studies showed that people are more than twice as eager to help (here: generous) when they know the face (even from a photo) and the data of a specific person in need, (the identifiable victim effect), rather than when the information about the needy is not individualised. It means that only an identifiable, and not a statistical, victim of a natural disaster, war or poverty arouses sympathy and or empathy, while a statistical victim (in the sense: anonymous) does not. Why does it happen? In this context, the so-called "drop-in-the-bucket effect" is more important than the other two factors distinguished by psychologists (i.e. closeness and vividness). It is connected with someone's faith in his ability to single-handedly and completely help the victims of a tragedy. It occurs when our own possibilities of providing help are assessed as irredeemably and wholly insufficient to change a dramatic situation for the better (for example, to prevent starvation of millions of people suffering from drought in a remote African country[40]); in this situation, the futility and senselessness of potential help efforts leads to emotional detachment from the needy, resulting in failure to give them any help.[41]

Rationally speaking, if saving one person is good, then saving a few is more so. Similarly, the misery of a nation seems to be more evil than the tragedy of one person. Therefore, the above-mentioned studies, contradicting these common-sense conclusions, seem to prove the irrational character of reasoning of the examined persons. On the other hand, the indicated reasons explaining the results of these studies, elucidating the mechanism of the creation of compassion and consequent help, paradoxically show that rational thinking, like a Hobbesian calculation (weighing reasons) plays a significant role in the occurrence of help. The preference for an action aiming to aid someone specific instead of helping many abstract persons or such an objective is rational, because it is easier to rectify the situation of one person rather than of many anonymous people. In addition, in the case of helping an individual, it is easier to control how the help will be used. In short: helping a specific victim is more rational, because it is at least potentially more effective and effectiveness proves instrumental rationality.

The third of the above indicated reasons for the lack of help for statistical victims – the "drop-in-the-bucket effect" – which *nota bene* has a rational nature (a futile effort is irrational), leads us to an experiment designed to check whether greater rationality in thinking promotes aid. Before the test one group of respondents was asked to solve a simple mathematical equation. The goal was to prime (i.e. to put people in a particular, temporary state of mind) the participants so that they would be in a special disposition to think logically during the experiment. The respondents from the other group were asked a question aimed at evoking emotions in them – "When you hear the name George W. Bush, what do you feel? Please use one word to describe your predominant feeling." After that the respondents were given the information either about Rokia or about the general problem of food shortage in Africa. In the next step, the experiment participants were asked about the sum of money they would allocate for a given cause. The results showed that people who were primed to experience emotions, and therefore those whose reasoning was irrational (because it was under the influence of additional emotional premises, irrelevant from the point of view of a rational procedure of drawing conclusions), allocated much more money to Rokia than to the fight with the general the problem of hunger. Their results were, therefore, similar to the results of previous studies, the participants of which were not primed in any way. This means that without the priming, the respondents were guided by compassion when individualised information was involved. On the other hand, people who were primed to think logically (in the sense dispassionate) turned out to be misers – they allocated equally small amounts to both causes. It suggest that:

"A cold calculation does not increase our concern for large problems; instead, it suppresses our compassion. So, while more rational thinking sounds like good advice for improving our decisions, [purely rational – MMJ] thinking can make us less altruistic and caring" [2, p. 296].

From the point of view of the goals of this work, other experimentally confirmed irrational phenomena are also interesting. By this I mean the overestimation of what is the product of our own labour, unwarranted by its objective value (the so-called Ikea effect),[42] and the equally irrational favouring of own ideas (“not-invented-here”). Both phenomena have their negative and positive sides. The most obvious benefit associated with these cognitive errors is the motivation to act. The tendency to perceive the effects of one’s own work or creativity as better and more useful than similar works of other people seems to be an extremely effective incentive to undertake a new task, as well as one that requires a long and/or strenuous effort (scientific work).

One of the most interesting, in my opinion, research carried out by behavioural economists concerned the issue of honesty. In the set of crimes consisting in theft, two subgroups can be distinguished: 1. “evident” thefts committed by “professional” criminals; 2. thefts and frauds committed by people who consider themselves to be honest. Every year in the United States, the value of theft and fraud perpetrated by people in the latter of the “categories” exceeds the material losses caused by “professional” criminals.[43] This circumstance provoked researchers to experimentally determine whether and to what extent people, who deem themselves “honest,” will succumb the temptation of fraud when exposed to it.

The respondents were the students of the Harvard Business School. The first group was asked to take a test consisting of 50 multiple-choice, general-knowledge questions. The questions should be answered within 15 minutes and then the answers should be transferred to a scoring sheet. At the end both sheets should be submitted to the examiner. It was possible to obtain 10 cents for each correct answer. The second group of students took the same test and just like the first one had to mark the answers on the scoring sheet, but in this case this sheet already contained the correct answers, hence the participant were tempted to “correct” their mistakes. After transferring their answers, they were to calculate those that were correct, write that number at the top of their scoring sheet and hand both sheets to the examiner who paid the respondents the due amount. The third group was asked to do the same as the second group, the only difference was that they were told to destroy their worksheet and submit only the scoring sheet to the examiner. The best conditions for cheating were created for people from the fourth group. After completing the task, they were supposed to destroy both cards and instead of informing the examiner about the obtained result, they were to collect the prize from a jar with coins on the table.

As it could be expected, the most honest were the students from the first control group as they did not know the correct answers, unlike the other three groups, and therefore could not benefit from this knowledge. The average of correct answers was 32.6 out of 50 questions. The results of the respondents from the subsequent groups were higher: in the second group the average number of correct answers was 36.2; in the third - 35.9; in the fourth group - 36.1. What is important, the researchers found that it was not just a few individual students that significantly overstated the number of their correct answers – the majority of participants cheated. Similar results were also obtained in studies conducted at MIT, Princeton, UCLA and Yale. The similarity of the obtained results allowed the researchers to come to the following conclusion: when given the opportunity, people [often – MMJ] cheat. The banality of this conclusion contrasts with another regularity observed in the above-mentioned experiments – the lack of relationship between the scale of fraud and the amount of risk of being caught red handed.[44] According to the authors of the experiment, the lack of such a connection proves that: “even when we have no chance of getting caught, we still don’t become wildly dishonest” [1, p. 243].[45]

There is still a question about the reason for this limitation. Perhaps one[46] of the researchers, Don Ariely, administering these test is right. In his opinion, people generally care about honesty and want to be honest. However,

> their internal honesty monitor is active only when they contemplate big transgressions, like grabbing an entire box of pens from the conference hall. For the little transgressions, they don’t even consider how these actions would reflect on their honesty and so their superego stays asleep [1, p. 246].

What is more, these minor offences are not prevented by the rational cost-benefit analysis, accented by neoclassical economics, or the probability of being caught. It is suggested here that even if such considerations take place, they do not affect the integrity of the one who contemplates.

According to Dan Ariely's suggestion, cheating at tests in order to obtain a small financial gain is not (just like cheating an insurer or tax office) treated as denying someone's general integrity. This conclusion is supported by the results of the above experiment, showing that generally everyone who had such an opportunity cheated, even though the respondents representing the elite of society (students of one of the best universities in the United States) probably believed in socially supported moral values, forbidding, among others, committing crimes. The results of this experiment testify to the irrationality of their thinking for two reasons: 1. Reasoning that uses a double standard: one for "criminal" theft and the other for "minor" fraud (dishonesty shown in the above experiment) is irrational; 2. Assuming that cheating at a test is a rational phenomenon, because it is a sensible grasp of an opportunity and not taking it would be a "sin of omission", the lack of correlation between the benefits from the fraud and the risk of being caught, established in these studies, proves the irrationality of the respondents' actions, and indirectly the irrationality of their reasoning.

Does the tendency to small scams found in the studies have any advantages? The indicated predominance of material losses resulting from theft committed by "honest" people over the amount of bandit spoils seems to deny this possibility. On the other hand, perhaps, above-average honesty, as an actual, not desirable, characteristic of the general public might not be as socially useful as it seems. Every day, each of us makes many decisions. Their number and limited resources at our disposal (time, attention, information available) result in the necessity of sorting them into important, less important and irrelevant. It is probable that assigning the same weight to all decisions will result in the failure of the entire system. Perhaps just as it is impossible to simultaneously receive all external stimuli that come to us from the outside, so it is equally unrealistic to analyse all decisions we make on a day-to-day basis in terms of their compliance with our moral values and standards.

6. Conclusion

Perhaps, as I have suggested, a functioning, relatively moral society (in the sense of: "roughly" and officially adhering to the most important laws[47]) is better than an inefficient community of morally scrupulous people. However, can we always and/or in every area of life afford this kind of nonchalance as a society?

It is clear that not all decisions made every day are equally important. Similarly, not every one of them has a moral aspect. Nevertheless, in our times many, once morally neutral, private matters have gained moral significance. We can mention here the question of nutrition, consumer choices, holiday arrangements, lifestyle, the standard of living, etc. As moral problems, they all demand resolution in the form of a specific individual decision. What is more, individual solutions to these issues, having a direct impact on the natural environment, have ceased to be private matters of specific people. It is connected with the threat of ecological disaster on the scale of our entire planet pointed by many nature researchers (including philosophers[48]). In this situation, someone's rational thinking oriented towards achieving individual happiness, within the constraints of the current law and available possibilities, considered in a wider context of what is good of future generations, or even the current generation, but in the perspective of the next 30 years, maybe turn out to be irrational, because it leads to a significant deterioration of the living conditions of all inhabitants of the Earth. This possibility is emphasised by Andrzej Szahaj who notes that "the sum of micro-rationality may add up to macro-irrationality, which changes this micro-rationality into micro-irrationality" [19, p. 94].

Let us return to the indicated oversupply of problems demanding rational consideration, enforcing in some way their selection in terms of importance. Does this surplus inevitably and irrevocably imply that we must choose between them, i.e. give up the rational consideration of many of them? It is difficult to disagree with Sartre, Jonas or environmental ethicists, who in unison opt for a very broad range of individual moral responsibility, and therefore for moral scrupulousness (i.e. hyper-rationality), as a condition for the survival of our planet and species. However, as a result of the overabundance of issues perceived as requiring rational consideration, our existential situation seems to resemble Dworkin's dilemma associated with the design of a just social system, i.e. at the same time sensitive to ambition and indifferent to natural endowment.[49] If this comparison is legitimate, what can we do? Perhaps a compromise solution would be the inurement from an early age to practising modesty/humility understood as an attitude always taking into account the possibility of one's error.[50] In this sense, a humble person would always be willing to consider a given question in detail, should the need arise, and in its absence would rely on standard, previously worked out solutions.[51] What would attest to such a need? Own doubts about how to proceed, and in their absence – reservations or criticism from third parties, not necessarily close or significant. We are left with the problem of coexistence, openness to criticism and trust in one's own judgment and possibilities[52] It is difficult to be self-assured with a constant, or even abstract, conviction that you can always be wrong. On the other hand, perhaps it is exactly the point that the belief in the possibility of error should remain abstract. As such, it would not cause decision-making stalemate, nor moral pedantry consisting in an equally meticulous analysis of all circumstances of the case before any decision is made.[53]

References

1. Ariely, D. *Potęga irracjonalności*, transl. by T. Grzegorzewska, Wrocław: Wydaw. Dolnośląskie, 2009.
2. Ariely, D. *Zalety irracjonalności*, transl. by T. Grzegorzewska, Wrocław: Wydaw. Dolnośląskie, 2010.
3. Bińczyk, E. *Epoka człowieka. Retoryka i marazm antropogenu*, Warszawa: PWN, 2018.
4. Bombik, M. Typy racjonalności (I), *Studia Philosophiae Christianae* 37, 1, 2001, pp. 5-42.
5. Dobrowolski, J. *Filozofia głupoty. Historia i aktualność sensu tego, co irracjonalne*, Warszawa: PWN, 2007.
6. Gajewski, M. *Funkcje i dysfunkcje myślenia irracjonalnego*, *Annales Universitatis Mariae Curie-Skłodowska* XLII., 2, Lublin: Wydaw. UMCS, 2017, pp. 9-26.
7. Kleszcz, R. Kryteria racjonalności, *Filozofia Nauki* IV, 2 (14), 1996, pp. 121-133.
8. Kleszcz, R. *O racjonalności. Studium epistemologiczno-logiczne*, Łódź, 1998.
9. Marcus, G. *Prowizorka w mózgu*, transl. by A. Nowak, Sopot: Smak słowa, 2009.
10. Mordka, C. *Filozofia jako doksologia* (unpublished).
11. Nęcka, E., Orzechowski, J., Szymura, B. *Psychologia poznawcza,* Warszawa: „Academica" Wydaw. SWPS, PWN, 2006.
12. Niemczuk, A. *Racjonalność praktyczna – jej natura i działanie* (unpublished).
13. Ohme, R. K. *Preface to the Polish edition of G. Marcus's book, Prowizorka w mózgu,* In G. Marcus, *Prowizorka w mózgu*, transl. by A. Nowak, Sopot: Smak słowa, 2009.
14. Solek, A. Ekonomia behawioralna a ekonomia neoklasyczna, *Zeszyty Naukowe Polskiego Towarzystwa Ekonomicznego* 8, 2010, pp. 21-34.
15. Sutherland, S. *Rozum na manowcach. Dlaczego postępujemy irracjonalnie*, transl. by H. Jankowska, Warszawa: Książka i Wiedza, 1996.
16. Szczepański, J. Sfera irracjonalności, In *Sprawy ludzkie*, Warszawa: Czytelnik, 1978.
17. Szmyd, J. *Myślenie i zachowanie nieracjonalne*, Katowice: „Śląsk" Wydaw. Nauk., 2012.
18. Tałasiewicz, M. O pojęciu «racjonalności», *Filozofia nauki* 3 (1-2) 1995, pp. 79-100.
19. *Co jest racjonalne w życiu? Rozmowa z profesorem Andrzejem Szahajem przeprowadzona przez Jacka Żakowskiego,* Niezbędnik współczesny inteligenta (suplement Polityki), nr 5, 2019, pp. 90-98.

Notes

[1] I define reasoning here as “the process of formulating conclusions on the basis of premises, i.e. using previously acquired or commonly available knowledge” [11, p. 420].

[2] “To acknowledge the rationality of some view, act or man means to define it in a positive way. To deny them rationality means to show disregard, to exclude from the range of acceptable controversy” [18, p. 79].

[3] Cf., among others, [1], [15], [9], [17], [6], [5].

[4] This is not the only way we can use the evolutionary paradigm. Following Gary Marcus, we may as well argue that the universality of irrational thinking, and even the fact that it outnumbers the acts of rational thinking, results from the fact that rational thinking is one of our youngest capabilities, shaped in the process of the evolution of our species. The “young age” of rational thinking capability, explained by the “provisional”, i.e. “unfinished” character of brain structures responsible for rational thinking, accounts for the frequency and regularity of errors made by people in thinking, resulting in the irrationality of thinking [9].

[5] Although “scientific thinking mainly releases rational factors and makes them dominant, it does not disencumber itself – and cannot disengage – from other factors, e.g. feelings, intuition, faith, etc. And the ludic action of man releases mainly emotional and sentimental elements, but even they are completely devoid of rational elements” [17, pp. 90-91].

[6] In the sense of the co-occurrence of rational (e.g. criticism) and irrational elements (e.g. excitement or bias) in one act of thinking.

[7] J. Życiński distinguished pragmatic rationality as a type of rationality [4]. Although this type of rationality could be reduced to practical rationality, such a “procedure” would be unfortunate, because as a result the term “practical rationality” would gain a (“permanently”) instrumental sense.

[8] In the strict sense, these are the criteria of the rationality of action indicated by R. Kleszcz [8, pp. 44-85] cited in [4, pp. 38-39]. Since thinking is a kind of activity and is often a direct incentive to act, the criteria distinguished by Kleszcz can be also applied to it (thinking). Other criteria of the rationality of practical reasoning are indicated by [12] A. Niemczuk in his unpublished text. He distinguishes 5 criteria of practical rationality: 1. Affirmation of being, 2. Criticism and self-knowledge, 3. Non-contradiction, 4. Realism and effectiveness, 5. Respect for the hierarchy of values. These criteria correspond to the meta-principles of rationality highlighted by R. Kleszcz, described in the next footnote.

[9] Pundits are usually aware of the shortcomings of various rationality criteria, therefore, in accordance with the postulate advanced by Jan Szmyd [17, p. 93] they try to modify the existing criteria of rationality in such a way that they conform to modern knowledge about the complexity of the human cognitive apparatus and the specifics of cognitive activity of people. Such an attempt was made by R. Kleszcz who criticizes “standard conditions of rationality” (such as 3 conditions of rationality indicated by K. Szaniawski: 1. Proper (strict) articulation, 2. Respect for logic requirements, 3. Proper justification). He replaces them with a two-storey model of rationality, i.e. two levels of principles (criteria) of rationality. Level I – the level of meta-principles – would contain general and universal principles adequate for all areas of cognition and activity that would not be “rigid” rules. This means that their every use would entail the necessity to specify them, taking into account given circumstances. Kleszcz distinguished 4 meta-principles: 1. Language precision, 2. Observance of logic requirements (minimum rationality), 3. Criticism, 4. Ability to solve problems. All the rules are important and necessary, but the author assigns particular importance to the requirement of observing the rules of logic. On the other hand, the criteria of rationality of level II, as adequate for certain specific spheres (types) of cognition would correspond to the models of rationality proper for these different domains [7, pp. 122-131].

[10] Cited in [6, p. 17].

[11] What can be expressed in constant and tendentious disregard for information contrary to the decision made earlier or to one’s own view on some matter or even to one’s own worldview (dogmatism).

[12] Concretisation/examples of this error are: 1. “The halo effect” as a result of which one very positive trait of the object affects its overall assessment; 2. “The devil effect” – object assessment based on one negative feature; 3. stereotypical perception of the object – it can be positive (“All Richards are nice chaps”) or negative (“All blacks are lazy”) [15, pp. 34-36].

[13] Stuart Sutherland calls the tendency of coming to unjustified conclusions on the basis of clearly inadequate information the most common manifestation of irrationality [15, p. 10].

[14] Cf. [15, pp. 305-309].

[15] *Nota bene* emotions, similarly to irrational thinking, are complementary to rational thinking. This is not surprising, because emotions are traditionally included in the sphere of irrationality. Researchers like Damasio (cf. idem *Descartes’ Error*) emphasise that emotions, as the basis of a reaction to a stimulus that is faster than reflection (thanks to not engaging neocortex), improve the decision-making process. For this reason, many contemporary emotion researchers (among others Damasio, philosophers: R. Solomon and M. C. Nussbaum or evolutionary psychologists) regard emotions as a kind of “mechanisms” complementary to slower reflective thinking. On a more general level Gary Marcus writes about the insufficiency of rational thinking as the basis of an effective decision-making process. Starting

from an evolutionary point of view, he maintains that evolution has provided people with two complementary systems – the ancestral, unconscious reflex system and the evolutionarily posterior (and thus badly underdeveloped) deliberative system. These systems have different skills and a different scope of activity. The domain of the ancestral system are routine tasks, and of the deliberative one – new situations that require going beyond the usual patterns. However, their competences are not completely disjunctive. The reflex system not only works better when there is not enough time for a thorough analysis of the circumstances of the case. It also works well (if we give it enough time!) when it is necessary to take into account many variables. Similarly, because the ancestral mind is focused on estimating statistical data (it originally served to estimate the likelihood of finding food and predators in a specific area), it may be a better tool than the deliberative system in a situation where solving a problem requires compiling a spreadsheet. In short: the ancestral reflex system can sometimes have an advantage over the deliberative system in synthesising extensive data (*vide:* a "blink" described by Malcolm Gladwell or "intuition" understood as, following Ap Dijksterhuis – a Dutch psychologist – a premonition that is the result of insightful, unconscious thought processes, brought to perfection by years of experience). What is more, "it is not completely irrational, but only less deliberative" [9, pp. 104-105].

16. It should be emphasised that "acting according to set patterns" is neither a thoughtless act, nor is it "automatic" or "reflexive" in the strict sense of these words. Some insight in the "circumstances of the case" is always necessary, as in stereotypical thinking which although brief (a stereotype as a kind of cognitive pattern allows us to improve, i.e. shorten the time of reasoning) is still thinking though not as precise and reliable as reflective rational thinking. On the other hand, the same mechanism – acting according to set patterns – seems to occur in the case of emotional priming. A single situation resulting in a particular emotion in a given person may generalise to a situation of a similar or even different type in the future, resulting in an automatic interpretation of the new situation in a previously "primed" way cf. [2, pp. 312-316].

17. Sutherland claims that negative effects of irrational thinking in the private sphere are rather small, because most matters in this sphere are trivial. Only four are truly important in this domain: "which neighbourhood to live in and which house to buy; which career to follow and which options to choose within that career; whom, if anyone, to live with and when to stop doing so; whether to have children (an outcome that is in any case often involuntary). In all these choices, there are usually many unknowns, which means that rational thinking may only marginally increase one's chance of a successful outcome" [15, p. 315].

18. "If you select a job applicant because you are greatly impressed by his fluency at interview (the halo effect), he is unlikely to be totally unsatisfactory even though he might not be the best of those applying" [15, p. 308] *Nota bene* the use of heuristics resulting from mental laziness instead of "full-blown" rational thinking can sometimes be pragmatically rational, as M. Bombik indicates writing that "actions in which «strong» measures to achieve a goal are used without objective need cannot be considered rational (…) [Similarly – MMJ] when with relatively little effort there is a non-zero probability of achieving a high value goal, the pursuit of this goal cannot be considered irrational, even if the probability coefficient is very low" [4, p. 13].

19. Ignorance of this principle was demonstrated by Israeli Air Force officers complaining about their trainees who when praised after a particularly good flight flew poorly next time. Since a reprimand given to those who flew extremely badly resulted in a better next flight, they concluded that reprimanding was the best method of training the champions cf. [15, pp. 251-252].

20. At the same time Marcus warns us against equating this system with emotions. He argues that although many emotions (e.g. fear) seem to be reflexive, not all can be characterised in this way. Moreover, a great deal of this system has little to do with emotions [9, p. 64].

21. Because this part of the brain is also found in other mammals, but in their case it is much less developed, this may be the premise of the thesis about the evolutionary kludge of this solution.

22. The influence of the ancestral system on the deliberative one is visible, e.g. in individual beliefs. "We feel as if our beliefs are based on cold, hard facts, but often they are shaped by our ancestral system in subtle ways that we are not even aware of" [9, p. 65].

23. Behavioural economics is a relatively new field of knowledge that is interested in how people actually act as economic agents. Among others, psychologists Amos Tversky and Daniel Kahneman are considered its precursors, who in their work *Prospect Theory: An Analysis of Decision under Risk* used cognitive psychological techniques to explain many documented discrepancies in making economic decisions in relation to the neoclassical theory. Figures important for the development of behavioural economics were also two Nobel Prize winners in economics: 1. Gary Becker – an economist and sociologist, Nobel Prize winner of 1992 and 2. Herbert Alexander Simon – an economist, computer scientist, sociologist and psychologist, who received this award in 1978. The former was the author of *Crime and Punishment: An Economic Approach* (1967), a work that drew attention to psychological factors as important for making economic decisions. The latter was the author of the theory of limited rationality, which explained how people irrationally tend to be contented, instead of trying to maximise usability.

24. In specialist literature, e.g. in the books of Dan Ariely (one of the leading behavioural economists), the term "classical economics" is used instead of the term "neoclassical economics" (cf. idem *Potęga irracjonalności* as well as *Zalety irracjonalności).* On the other hand, authors such as Adrian Solek (cf. Idem [14]) identify what Ariely calls classical economics with neoclassical economics. It is argued here that at the beginning of its development, classical economics contained numerous references to psychology, ethics and morality. For example, the author of *The Wealth of*

Nations, Adam Smith was also the author of the book *The Theory of Moral Sentiments* in which he showed that the criterion of moral principles is not the consideration of one's own benefit (Hobbes) or the compatibility of these principles with reason (Kant), but a feeling of sympathy. Similarly, Jeremy Bentham's utilitarianism, which is the ideological basis of classical economics, had many references to psychology. In contrast to this early period of development of economics as a science, "flirting" with the psychology and ethics, neoclassical economics has moved away from these sciences. As a result, neoclassical economists emphasised the rational nature of economic behaviour. In the light of this new approach, consumers as economic people (*homo oeconomicus*) are actors whose decisions and actions result from their will to make rational choices.

25. "On the basis of these assumptions, economists draw far-reaching conclusions about everything from shopping trends to law to public policy" [1, p. 285].

26. The author of this term is J. Szczepański [16, p. 127]. He emphasised that in every human being, apart from the sphere of rationality, there is a domain of irrationality, hence there is no man who functions as a logical machine.

27. In the strict sense, the "explanation" is in this case an exaggerated term, because the explication of the existence of something (here: a cognitive error) as an exception to the applicable rule seems to be rather an evasion.

28. An example of such an impact are, for example, consumer behaviours, which instead of serving to satisfy personal needs or tastes of individuals (one of the assumptions of the classical economics) sometimes serve other purposes, for example a public image. These include the experimentally determined tendency of Americans to emphasise their individuality by ordering beer of a different brand from the ones chosen by accompanying people (if it is ordered in public, i.e. orally, and not when the order is submitted in writing). The important thing here is that beer chosen by the person willing to emphasise their distinctness is often not what they would really like to order. This means that by their choice they sacrifice their own pleasure. Also in Hong Kong, the surveyed people were clearly under the influence of their surroundings in their choices. However, because Asians belong to a collectivist culture and therefore tend to emphasise (also through their consumer choices) belonging to their group, the people surveyed in the bar ordered (aloud) what their companions had previously ordered. However, when alcohol was ordered in writing (no influence of the environment on the decision), the orders of the respondents differed from the choices of their companions and thus reflected their true preferences [1, pp. 279-284].

29. A manifestation of short-sightedness is, e.g. not saving enough for future retirement. Neoclassical economics does not attach any importance to this phenomenon, because according to its rational vision of human nature, people (rational market participants) save as much as they want. Thus if the sums they save are really very small, it means that saving of this type is a rationally (though not necessarily fortunately) chosen option. However, in the light of behavioural economics, as it does not assume the rationality of human actions, the statement that people do not save enough is logical. Several reasons for this are indicated: procrastination, having a hard time understanding the real cost of not saving as well as the benefits of saving, a false belief that if someone owns a house, he is indeed rich, etc. [1, pp. 287-288].

30. One of the remediation strategies proposed by Dan Ariely is based on the earlier discovery of social psychologists who found that honesty of people is enhanced by activating their self-awareness, and strictly its part containing information about moral norms a given person identifies with. In their experiments, the activator was, for example, a mirror. In Ariely's experiments it was established that the same role can be played by the principles of the Ten Commandments written down by the subjects directly before solving a task, during which they were exposed to the temptation of cheating. Since recalling the Ten Commandments raised the honesty of the participants (they did not cheat during the test) regardless of whether they remembered all of the rules or just some of them, Ariely concluded that just thinking about a certain moral pattern encouraged honesty. This supposition was confirmed in further experiments, in which, before taking a test that was to check their honesty, the subjects had to sign the following statement: "I understand that this study falls under the MIT honour system." People who signed this pledge obtained the same results as those in the control group who did not have a chance to cheat. In the above statement, Ariely sees a kind of professional oath that obliges people of specific professions (doctors, lawyers, employees of science) to act in accordance with the ethos of a given profession. It is stressed here that occasional swearing of an oath or signing a statement on compliance with the rules is insufficient. These acts must be repeated and they always must precede making a decision in the situation of temptation because "When social and market norms collide, the social norms go away and the market norms stay" [1, p. 257], cf. [1, pp. 250-256].

31. Dan Ariely also shares this conviction, and in *Zalety irracjonalności* he stresses that irrationality has its advantages. "It allow us to adapt to new environments, trust other people, enjoy expending effort, and love our kids" [2, p. 19].

32. The stress caused by a high bonus looks like that which accompanies the presence of other people during performing a given task. In the latter situation, it was observed that if the required activity observed by onlookers is well-learned and quite easy (such as riding a bicycle) then the presence of spectators is conducive to better fulfilment of this task. However, when the activity is difficult, its performance in the presence of witnesses results in a worse level of its performance. This experimentally proven relation is called social facilitation.

33. Cf. also chapter 1 [2, pp. 25-65].

34. It is quite often the case that we provide the most rational advice to others, that is when we consider matters that do not pertain to us. This seems significant, given that people generally reluctantly act on it.

35. Just like for the hero of the film *Slumdog Millionaire*.

36. Largely due to experiments carried out by social psychologists.

37. In the strict sense, these features characterise the "donor" to the same degree.

38. It means not only that often help is not given to unconscious people because their appearance (the swollen and red face as a result of high blood pressure) suggests potential helpers that they are dealing with an intoxicated person, not someone who needs medical intervention. It seems that people often think that a person in need of help (e.g. a mother raising money for the treatment of her child in a foreign clinic) should look according to her situation, i.e. the situation of someone who comes asking for money. An unpretentious, and preferably poor look is appropriate for this situation.

39. I am talking about an experiment conducted by Deborah Small, George Loewenstein and Paul Slovick. Researchers gave each experiment participant $ 5 for completing several questionnaires. After receiving the money, the respondents received a piece of information about hunger in the world. Then they were asked how much of their just earned five dollars they would be willing to donate for the case they were reading about. The respondents were divided into 2 groups. In the first one, which was called statistical, the information concerned the need for immediate financial assistance for several millions of people threatened by hunger in four African countries. On the other hand, the respondents from the second group (called the identifiable victim group) were presented with information about Rokia – a very poor seven-year-old girl from Mali who faced starvation. Moreover, the respondents were shown a photo of this child and an additional piece of information: "thanks to your donation and support of others, her life can change for the better". This difference in the content of the information translated into the results obtained in both groups. In the first group, the average donation to famine victims in Africa was 23% of the five-dollar earnings, while in the second group – more than double that amount, i.e. 48% [2, pp. 285-287].

40. Help for a seriously aggrieved person, a person harmed in many ways, is similarly treated as not making sense. Paradoxically, the more help someone needs, the harder it is to find those willing to give it, although it should be easier, because a more injured person seems to need help more than someone less injured. And although it really is the case, helping someone who is very disadvantaged often seems senseless. This is explained by the fact that in the face of great harm each instance of help seems too small, because it has no power to completely eradicate the ill caused by this harm. Thus, cases of providing aid to those who are aggrieved to a smaller extent are more frequent. Acting on behalf of such people is considered sensible because it completely or significantly reduces their harm, and it "makes a difference."

41. Cf. [2, pp. 289-293]. The drop-in-the-bucket effect is also, according to D. Ariely [2, p. 300] one of the important reasons why many people do nothing to counter global warming in any way. They assume that their extraordinary efforts to save the Earth from the environmental disaster – for example by driving a hybrid car, changing all light bulbs to energy-saving ones, switching to veganism, etc. – would be too insignificant to solve this problem.

42. Research on this error showed that: 1. The effort put into something changes not only the object but also the subject and his evaluation of that object (product); 2. Harder labour leads to assigning greater value to the product; 3. Our overvaluation of the things we make is so deep that we assume that others share our biased perspective; 4. The impossibility to complete something that requires great effort results in the lack of attachment to it. All the above conclusions can be used to indicate the benefits from favouring your own products [2, p. 126].

43. Every year the value of American employees' theft and fraud at the workplace is estimated at $ 600 milliard. For comparison, the total value of robberies, burglaries, larceny-thefts, and automobile thefts committed in the USA in 2004 amounted to about $ 16 milliard. Every year American insurance companies deal with individual customers who overstate the value of lost property by $ 24 milliard. According to the estimates of the American tax office, it loses about $ 350 milliard every year – this is the difference between the value of taxes that the government expects to collect and the sum it actually receives. In turn, the retail industry loses $ 16 milliard every year, due to customers who buy clothes and wear them for some time, and then they get bored with them and return to the shop, which is possible because they have not removed price tags [1, pp. 237-238].

44. Cf. results obtained in groups II-IV.

45. Alternatively, it can be assumed that the conditions that were created for the subjects from the fourth group, as too openly conducive to fraud, could generate in the experiment participants the conviction of the existence of some "catch" that would enable the disclosure of their deception. If that was the case, then their non-cheating, and the reasoning that led to it would be rational.

46. The integrity tests described here are the result of work of three researchers – Nina Mazar (a professor at the University of Toronto), On Amir (a professor at the University of California in San Diego) and Dan Ariely (at that time a professor at MIT in Massachusetts).

47. It would be a community of students from the quoted experiments who although cheated "did not go beyond a certain (relatively low, not to say «decent») level of dishonesty."

48. For example, by H. Skolimowski, H. Jonas, W. Tyburski, Ewa Bińczyk [3] and many others.

49. According to Dworkin, a just social system must attain two conflicting goals: 1. equalising the chances of all citizens, 2. creating conditions for the development of talents. Any system that wants to be just must pursue both goals. The problem is that they are practically contradictory. Likewise, making a decision often requires some reconciliation of the values and/or goals involved. In addition, decisions made by a person must also be in some way compatible with one another. In this situation, quick, based on the learnt disposition to respond appropriately to the situation, decision making, if it occurs (and often must occur), is vulnerable to errors.

50. It seems that an advocate of this solution was, for example, G. Marcel, who in *Being and Having*, noted that a thinker, as oriented at being before having, ought at any time to criticise his own thought, not to attach to it, not to treat it as own property, let alone identify with it.

51. An example of such a standard solution is the development of virtues – constant dispositions to appropriately respond to situations of a certain type.

52. It seems that the coexistence (in a given person) of openness to criticism (internal humility) and self-confidence is particularly problematic in the case of very young people. The suggested difficulty would explain arrogance typical for many young people, or even disregard for the opinion of the older generation about their own person and/or how to live, what to cherish, etc. Disregarding older people can be seen as a defence mechanism that prevents young people from losing their confidence in their own competence to make the right decisions. On the other hand, the co-existence of humility and self-confidence can be similarly difficult for mature people. If maturity brings knowledge about the inevitable relativity of things, including the relative character of one's own judgments, it can not only counteract adamant attitudes but also foster doubt in the sense of making choices.

53. On the other hand, there is a danger that such a general belief about the possibility of error might turn into a mere hypothesis. In this case, this belief would lose the status of a real possibility, which should always be taken into account seriously.

Studia Humana
Volume 8:2 (2019), pp. 47—54
DOI: 10.2478/sh-2019-0017

Moral Considerability and Decision-Making

Magdalena Hoły-Łuczaj

University of Information Technology
and Management in Rzeszow,
Sucharskiego 2 Street,
35-225 Rzeszow, Poland

e-mail: mholy@wsiz.rzeszow.pl

Abstract:
The paper revisits metaphysical and deontological stances on moral considerability and offers a new criterion for it – “affectability”, that is a capacity of an agent to affect a considered entity. Such an approach results in significant changes in the scope of moral considerability and is relevant for discussing the human position in the Anthropocene. This concept, given especially the assumption of the directness of moral considerability, is also substantial for the decision making process on the ethical, as well as the political plane.

Keywords: moral considerability, non-human beings, decision making, science advice.

1. Introduction

The paper aims to contribute to the discussion on moral considerability by offering a new criterion for it, which can redefine its scope. The issue of moral considerability is related to the question *to whom* our actions should be morally considered. It was not asked for ages since ethics took as obvious that only relations with people (whose definition has changed over the centuries) need to be evaluated this way.

The first to oppose this claim was environmental ethics. It holds that our dealings with non-human natural beings should be morally assessed, or, to put it differently, non-human natural beings *deserve* to be considered morally and to be treated correspondingly. This state of ‘deserving’ is referred to as ‘moral considerability’ [1].

The set of non-human (natural) beings is, however, internally very diverse and thus the question which specifically should be within the scope of moral considerability and on what criterion is the core issue in the debate on that matter. Most commonly employed approach in defining it was a “metaphysical” one – oriented toward indicating a trait of a considered being which can qualify it to have moral considerability. It is opposed by a “deontological” stance, according to which we should think of moral considerability in terms of obligations of an agent (the one who performs the action) [2]. In the paper, I critically compare and revisit both stances. I shall argue, building upon this reexamination, that a valid criterion for granting moral considerability for an entity should be *a capacity of an agent to affect it*.

My argument will be structured as follows. First, I introduce the idea of moral considerability. Next, I reconstruct the debate on criteria of moral considerability and groups of being granted it accordingly. In the third section, I revisit deontological approach to moral considerability. In the final part, I lay out the concept of affectability as the criterion for moral considerability.

2. Moral Considerability

The term "moral considerability" was coined by Kenneth Goodpaster in 1978 [1]. It defines the status of a being which requires human beings to consider implications of their actions toward this being in normative terms.

This idea is complementary to the concept of 'moral patienthood' as distinct from 'moral agency.' The class of moral patients is the class of beings to whom we consider that we owe ethical obligations, whereas moral agents are defined as that class of moral patients, usually only human beings, to whom we owe obligations and who, in turn, are held to be morally responsible for their actions. This distinction assumes that all moral agents are moral patients, but not all moral patients are moral agents [3]. By the same token, beings which can be granted moral considerability do not need to morally consider their actions.

The tendency to shift the boundaries of moral considerability toward non-human beings is closely linked to the efforts to uproot anthropocentrism and related to it the perspective of subordination of one being to another. It is believed that only such a move will enable us to eradicate mechanisms of violence and control [4]. Yet, today there is also another approach, which argues that only when we admit that not only human beings are agents, human hubris will be restrained. The belief that solely human beings are agents is often thought to result in that there are no constraints to their activity which becomes destructive hyperactivity. This can be observed in our current situation to which we refer as to the Anthropocene.

This name was suggested by the Nobel prize-winning atmospheric chemist and climate scientist Paul Crutzen [5], [6]. He believed that this is the most apt expression to describe current instable and unpredictable state of the Earth [7]. Even though neither the International Commission on Stratigraphy nor the International Union of Geological Sciences has yet officially approved the term as an indication of a particular geological period, the idea of the Anthropocene became a great source of inspiration for scholars working in various fields, creating new frames for their investigations.

The concept of Anthropocene consists basically of two assumptions. First, that the human (*anthropos*) has gained geological agency and has become the most important geological factor on the planet, trumping all the natural (non-human) factors. That is to say, the *anthropos* becomes a geological layer, just like ice before, in the sense that human agency determines the face of the Earth. Second, that it severely changed the Earth's atmosphere and biosphere, resulting in global warming and the collapse of vital ecosystems [7], [8].

In this sense, the Anthropocene is the epoch of fulfilling anthropocentrism and all its threats. The remedy for that, according to some scholars, is granting agency – including its moral dimension – also to non-human beings. The representatives of such an approach are primarily posthumanists. They put a premium on the activity of non-human beings, arguing that taking away agency from non-human beings supported a human sense of superiority, which translated into an arrogant way of dealings with beings other than human [9]. To eliminate this practice, posthumanists emphasize the ability of things to act, which according to them is the most solid ground for ascribing things a moral status.

Decreasing human hubris by means of ascribing moral agency to non-humans can have however one unintended result: decreasing human responsibility. If human beings could say that this is not only their wrongdoing to Earth but also, for instance, of some machines, it would wash away human guilt. And this is undesirable. Thus, I argue we should rather get focus on the problem of moral considerability to make human beings realize that they have to think, in ethical terms, not only about their fellow humans, but also about many other beings in the world. The question remains which non-human beings this should concern.

3. Metaphysical Criteria of Moral Considerability

The debate about criteria of moral considerability is inseparable from the discussion on beings that should be granted it. In what follows, I briefly reconstruct the main positions in this discussion.

Before the crystalizing the idea of moral considerability, Aldo Leopold, the founder of environmental ethics, claimed that we should morally respect nature as a whole for how harmoniously it functions, among other reasons [10]. Following such a claim, "a thing is right when it tends to preserve the integrity, stability, and beauty of the biotic community. It is wrong when it tends otherwise" [10, p. 15]. By this token, the feature which qualifies some being to moral considerability is the ability to affect ecosystems (this approach was thus referred to as "land ethic").

With this perspective, however, we can lose sight of the good of individual beings, such as animals from non-endangered species [11]. Thus, environmental ethicists started to advocate ascribing moral considerability to particular groups of non-human beings. Peter Singer argued that such a capacity should be granted only to animals, as they are sentient beings that are able to suffer/feel pain [11]. So if the spectrum comprises only human and non-human animals, improper behavior would be inflicting pain, while proper behavior would entail helping it to avoid, reduce or cure its pain/suffering.

Other theorists found that rationality, the capacity to suffer and to enjoy, or having interests are all arbitrary stopping points and thus they argue the above limitation is too restricted and we need to include also animate beings (e.g., plants) into the domain of ethics as living beings – teleological centers of life [12], [1]. Granting moral considerability to all animate beings results in that its criterion is whether we hinder or enable the self-development/self-reliance of such being.

Another group of environmental philosophers took a step further and argued for the moral considerability of inanimate natural beings (rocks, mountains, rivers) insofar as they belong to the sphere of physis, or natural order [13], [14]. If we take into account inanimate beings as well, the criterion of moral considerability is whether we interfere or help maintain their existence and integral identity. Of course, the above criteria apply upwards. For instance, the criterion of avoiding pain concerns exclusively sentient animals, but the right to be undisturbed is valid for all beings.

Extending the limits of moral considerability shows also how important is the *directness* of moral consideration and responsibility is [15], [16]. Only direct attention, being focused on the particular type of things, guarantees sufficient respect for them, eliminating the threat that their well-being will be overlooked or ignored for the sake of some other, more human-like being or nature as a whole.

The last caveat is not accidental: the horizon of moral considerability for environmental ethics is nature (either as whole or as the set(s) of individual natural entities). What about other beings? For instance, artifacts? They seem to be excluded by default. A good illustration of this problem is Thomas Birch's theory of a "universal consideration."

Thomas Birch was very sympathetic to the idea of widening the scope of moral considerability to non-human beings discussed in the sphere of environmental ethics. Birch accused, however, environmental ethicists of marasmus – getting bogged down in pointing to a specific feature of beings which should be a final criterion for granting them moral considerability. He challenged this need, offering a remedy which would be a 'universal consideration'. Its basic principle is simply taking *everything* into moral consideration in a given situation [16, pp. 314, 331]. He defined "to consider X morally" as attending, looking at, thinking about, sympathizing with X, etc. with the goal of discovering what, if any, direct ethical obligations one has to X [16, p. 315].

Such an approach seems to be very promising and open new possibilities for ethics concerned with non-humans, significantly broadening its scope. Unfortunately, it is only seemingly so. Even though Birch writes about “everything” [16, pp. 313, 314, 318, 321, 327, 330.] or “all others of all sorts” [16, p. 313] throughout the entire text, at the end he makes clear that “all things” refer solely to the “whole biosphere” [16, p. 331]. He does so without further explanation. We can only suspect that this is due to perceiving artifacts as inferior in environmental ethics as they lack intrinsic worth understood as a non-instrumental value. We can challenge such a framing of the problem of intrinsic worth can by pointing to the functionality of nature and mutual dependency of all the natural beings, but analyzing this problem goes beyond the scope of the paper. What is important for us is that Birch’s position turns out to be also grounded in metaphysical claims, even though some scholars read his theory as a deontological one, on which we comment in the next section.

But before we discuss that let us note there are approaches which advocate taking artifacts into moral consideration. An example of that can be Luciano Floridi’s “information ethics.” Information ethics suggests that there is something even more elemental than life, namely *being* — that is, the existence and flourishing of all entities and their global environment — and something more fundamental than suffering, namely *entropy*. Entropy here refers to any kind of *destruction* or *corruption* of informational objects, that is, any form of the impoverishment of *being* including *nothingness*, to phrase it more metaphysically [17, p. 47], [18, p. 146]. According to this, “what makes someone or something a moral patient, deserving of some level of ethical consideration (no matter how minimal), is that it exists as a coherent body of information. Consequently, something can be said to be good, from an IE perspective, insofar as it respects and facilitates the informational welfare of a being and bad insofar as it causes diminishment, leading to an increase in information entropy” [17, p. 300], [18, p. 146].

The question is whether everything which has a body of information can be in fact destroyed, or, if we would like to nuance this objection, we can ask whether beings vary when it comes to their permanent damaging and destroying. Digital beings seem to be in a significantly different situation in this regard than physical ones. Likewise, it appears that ideal beings cannot be destroyed or altered to any extent. These remarks, however, are not to challenge Floridi’s stance, but to show that his sound and inspiring theory calls for further examination. In this paper, however, I suggest approaching the problem of moral considerability from yet another angle. Namely, *a capacity of an agent to affect a particular being*.

However, before elaborating on this, we shall investigate insightful criticism of the above metaphysical concepts of moral considerability offered by deontological stance.

4. Deontological Criticism

Aforementioned Thomas Birch’s theory of universal consideration received significant feedback [19], [20]. Among scholars who responded to that concept, more or less critically, was Benjamin Hale. He offered, being inspired by Birch’s stance, a “deontological approach” to moral considerability.

The core claim of it is that “moral considerability should be understood narrowly and centrally as an agent-relative deontological question” [2, p. 37]. Hale argues that moral considerability is better understood as a question about a moral agent’s duty than about a moral patient’s status. Hale holds that rather than focusing on the properties, attributes, or capacities of other beings that qualify them as moral patients, we instead should focus on the obligations of rational agents to consider others [2, p. 37].

Hale explains his position in the following way: if we ask whether something is comprehensible, for instance, it would be odd to say that it “has” comprehensibility. We ask *ourselves* the question “can we comprehend it?” and do not seek to locate this comprehensibility in any of its constituent parts. Thus, when we suggest that something is comprehensible, we ask about its comprehensibility, and we suggest that it is or is not comprehensible *for us* [2, p. 41]. This allows Hale to conclude that moral consideration is an obligation of the agent [2, p. 45].

According to him, such a perspective offers a fresh solution to a problem that has plagued environmental ethicists for years. Hale, following Birch's criticism, says that seeking to establish a ground for moral considerability in a specific attribute of a particular group of beings leads to a dead-end – environmental ethicists cannot reach the consensus which beings deserve moral consideration.

Hale's stance is, however, entangled in a metaphysical presupposition about the superiority of nature in a similar manner as Birch's is. We can see it if we take a look at the assumption of considering *everything*. While Hale believes we must consider all about implications of our behavior, he does *not* think that *all entities* in the world are morally considerable [2, p. 39]. That is to say, he suggests that "everything must be considered – all factors unique to a given situation must be considered – not that every object or entity in the world is morally considerable" [2, p. 40]. Entities which he excludes are artifacts. Unlike Birch, he does so explicitly and comments on that, referring to the fact that the production of artifacts requires the use of natural resources. He says that it would be "double counting" [2, p. 50]. Hale argues that creators have already considered their component parts when producing specific artifacts. In his view, only damaging a tree (by cutting it) to produce a chair deserves moral consideration, while damaging a wooden chair does not as it is made from an already cut tree [2, p. 50]. So for Hale only other humans, animals, plants, or mountains are worthy of our moral attention [2, pp. 40, 47].

I do not agree with such a justification, but I believe Hale's criticism is worth attention. In what follows I offer, drawing upon ideas from a traditional metaphysical stance on moral considerability, as well as a deontological approach to it, a new criterion: affectability.

5. Affectability

What is particularly interesting in the theory of universal consideration, as well as deontological stance to moral considerability is the emphasis on human responsibility. They make us aware this is our doing which we need to think about. We cannot forget that these are our actions that we assess in the act of moral consideration. Shifting the burden of proof on human beings in developing the idea of moral considerability does not have anthropocentric connotations here. It rather reminds us that non-human beings can be harmed or benefited *by human beings*. This is an undeniably significant advantage of Birch's and Hales positions.

Furthermore, Birch's theory nicely explains the nature of consideration in moral consideration. It shows it is a specific attitude, or a willingness to be reflective about one's own behavior, rather than making calculations concerning each separate action, which would be quite alien to our everyday practices. Unfortunately, this is the direction in which Hale's concept is heading. He argues that "we are obligated to consider as much as is practically feasible" [2, p. 45] and that "whatever the case, to reframe the question of moral considerability deontologically only demands that we are obligated to consider the full spectrum of features of our actions before we act" [2, p. 50]. Such deliberations, however, seem to be quite unrealistic. My skepticism about it should not be read as a call for being thoughtless. On the contrary, I believe moral considerations are a significant component of making decisions process, but in the way Birch proposes it, not Hale.

Another drawback, this time of both stances, is that they do not explicate hidden metaphysical assumptions, of which they were supposed to be free of. To paraphrase Hale, their approaches mask the metaphysical underpinnings of the question [2, p. 41]. First, these theories are limited to natural beings. Birch does not explicate why, but we can assume he think they are of inferior status, while Hale directly states that he believes artifacts are secondary to natural entities (as produced from them). He sees his position as related to consideration level, but in fact he holds strong metaphysical assumption on that artifacts are not fully-fledged beings.

But this "metaphysical" objection can go further. Birch and Hale speak of considering "everything" or "every being": what if we were to take the terms literally? Between the biosphere, to which both theories in question are limited, and "everything" stretches the ontological abyss, which is inhabited not only by artifacts but also by, depending on ontology one accepts, fictional beings, abstract object, universals, tropes, properties, etc. Should they be morally considered as well, along with, for instance, concrete individuals? We need to review the answer in terms of what morality and ethics are about.

The most straightforward description is that morality and ethics deal with actions, or more precisely, *actions affecting other beings* [21, p. 20]. Ethics is then the analysis of the specific relationship between two subjects (an agent and its patient) – how one affects the other, changes it, makes a difference to it, etc.

In accordance with this claim, *only those beings that we are able alter or impact can qualify in the domain of ethics*. By the same token, ethics cannot include beings that human beings (or any other beings) cannot affect. This is the case for abstract beings (if we accept their existence), such as the color white, the idea of a triangle, acidity, or fictional beings, such as literary characters [22, pp. 19-20]. The (ideal) triangle is by definition immune to any changes. We cannot make it better or worse, thus we cannot consider it *morally* – we can think about (consider) it, but only in a non-moral way.

The entry-level in moral considerations should be then not the measure of ontological "perfectness," but the possibility to affect something. It is not that abstract or fictional beings do not deserve to be considered morally; we are simply unable to relate with them in this way. That is to say, moral considerability is not about the respect and reverence for such beings, but only about the possibility to influence them.

We need to consider then the implications of our actions toward other entities. To what extent do we influence them, how permanent are changes introduced by us, if these effects are irreversible or not, etc. When thinking about the appropriate course of action, we should take into account what actually can be harmed, damaged, violated, trespassed upon, and so on. In this sense, the capacity to absorb moral consideration is grounded in both a *trait* of an entity in question *and* agent's *ability* to impact this entity.

These questions become of crucial significance when considering our activity towards those with whom we are asymmetrically related [see 2, p. 57]. In such a case, it often turns out that entities which were "mere things" start to matter morally. A good illustration of it can be the aforementioned artifacts. When we directly devote our attention to them, they become able to reveal their unique character as concrete individuals. This may support our willingness to care for them and not replace them so easily in a more effective way than threatening us with harming nature by making more waste (when we dispose our things) and using more natural resources (to produce a new artifact).

The above example also illustrates how the issue of moral considerability is substantial in decision-making. It concerns both the ethical plane, as well as the political. Moral considerability does not only shape our moral sensitivity as individuals, but is also an important factor in setting pro-environmental policies. For instance, including artifacts to the scope of moral considerability can translate into alternative strategies of reducing plastic waste. Instead of negative argumentation (plastic things as a threat to nature – garbage, use of resources, etc.), we can provide a positive one by appealing to that they are our belongings for which we are responsible.

Finally, referring to the idea of moral considerability in policy- and decision-making can be read as the case of science advice. Thanks to using philosophical arguments, we can challenge existing patterns of actions and construct new ones, or we can validate and strengthen ideas from outside of the mainstream to build better society and better future for the planet [23, p. 238].

6. Conclusions

The debate on moral considerability encourages us to widen its scope. We invite to this club, to cite Thomas Birch, more and more non-human beings, agreeing we cannot cut them out of the picture when thinking of the sphere of ethics [16]. The question remains, however, *how* inclusive this club can be. Difficulties in finding a trait of the entities which should be granted moral considerability prompted some scholars to claim that simply everything deserves such our attention. This assumption was accompanied by the claim that we should place the burden of proof on the agent ("deontological" stance), and not on the patient ("metaphysical" stance), when addressing the problem of moral considerability. This paper offers a third way. It argues that the criterion should be *affectability*: a possibility of an entity to be affected by an agent, or, looking form the other side, an ability of an agent to affect in any way a given entity. In doing so, such an approach attempts to achieve equilibrium between "deontological" and "metaphysical" stances by showing that moral considerability of a particular entity should be grounded in the ontological condition of it as well as agent's ability to affect it. This deontological aspect of moral considerability, which translates into highlighting human responsibility for the actions towards other, non-human beings, is targeted at showing that moral considerability is a significant component of the decision-making process.

References

1. Goodpaster, K. On Being Morally Considerable, *Journal of Philosophy* 75 (6), 1978, pp. 308-325.
2. Hale, B. Moral Considerability: Deontological, Not Metaphysical, *Ethics & the Environment* 16 (2), 2011, pp. 37-62.
3. Light, A., Holmes Rolston III. Introduction: Ethics and Environmental Ethics, In A. Light, Holmes Rolston III (eds.), *Environmental Ethics. An Anthology*, Malden (MA): Blackwell Publishing, 2006.
4. Sessions, G. Ecocentrism and the Anthropocentric Detour, In G. Sessions (ed.), *Deep Ecology for the twenty-first Century: Readings on the Philosophy and Practice of the New Environmentalism*, Boston: Shambhala Publications, 1995.
5. Crutzen, P. J. Geology of Mankind, *Nature* 415 (6867), 2002.
6. Crutzen, P. J., Stoermer, E. F. The 'Anthropocen,' *Global Change Newsletter* 41, 2000.
7. Lemmens, P, Blok, V., Zwier, J. Toward a Terrestrial Turn in Philosophy of Technology, *Techné: Research in Philosophy and Technology* 21 (2-3), 2017.
8. Blok, V. Earthing Technology: Towards an Eco-centric Concept of Biomimetic Technologies in the Anthropocene, *Techné: Research in Philosophy and Technology* 21 (2-3), 2017, pp. 127-49.
9. Bennett, J. *Vibrant Matter: A Political Ecology of Things*, Durham: Duke University Press, 2010.
10. Leopold, A. *A Sand County Almanac: And Sketches Here and There*, Oxford: Oxford University Press, 1986.
11. Singer, P. *Animal Liberation: A New Ethics for Our Treatment of Animals*, New York: Avon, 1975.
12. Taylor, P. *Respect for Nature*, Princeton: Princeton University Press, 1986.
13. Naess, A., Sessions, G. Platform Principles of the Deep Ecology Movement, In A. Drengson, Y. Inoue (eds.), *The Deep Ecology Movement. An Introductory Anthology*, Berkeley: North Atlantic Books, 1995.
14. Brennan, A. *Thinking about Nature. An Investigation of Nature, Value and Ecology*, Athens: The University of Georgia Press, 1988.
15. Gorke, M. *The Death of Our Planet's Species: A Challenge To Ecology And Ethics*, trans. By P. Nevers, Washington, DC: Washington Island Press, 2003.
16. Birch, T. Moral Considerability and Universal Consideration, *Environmental Ethics* 15 (4), 1993, pp. 313-332.
17. Floridi, L. Information Ethics, its Nature and Scope, In J. van den Hoven, J. Weckert (eds.), *Information Technology and Moral Philosophy*, Cambridge: Cambridge University Press, 2008.
18. Gunkel, D. *The Machine Question Critical Perspectives on AI, Robots, and Ethics*, Cambridge (MA): MIT Press, 2012.
19. Hayward, T. Universal Consideration as a Deontological Principle: A Critique of Birch, *Environmental Ethics* 18 (1), 1996, pp. 55-63.
20. Weston, A. Universal consideration as an originary practice, *Environmental Ethics* 20 (3), 1998, pp. 279-289.

21. Singer, P. *Practical Ethics*, Cambridge: Cambridge University Press, 2011.
22. Bernstein, M. *On moral considerability. An essay on who morally matters*, Oxford: Oxford University Press, 1998.
23. Gare, A. Ethics, Philosophy and the Environment, *Cosmos and History: The Journal of Natural and Social Philosophy* 14 (3), 2018. pp. 219-240.

Acknowledgements

This research was supported by the grant DIALOG 0023/2019 from the Ministry of Science and Higher Education (Poland).

Studia Humana
Volume 8:2 (2019), pp. 55—68
DOI: 10.2478/sh-2019-0018

Practical Rationality – its Nature and Operation

Andrzej Niemczuk

University of Rzeszów,
Rejtana 16c Av.,
35-959 Rzeszów, Poland

e-mail: aniemczuk@ur.edu.pl

Abstract:
The article presents a proposal of explanation what practical rationality is, how it works and what are its criteria. In order to define practical rationality, the author starts from the general characteristics of reason, and then in the realm or reason activity distinguishes practical rationality from theoretical rationality. The necessary conditions of practical rationality are presented, as well as its standing between freedom and values. Next, the sources and nature of practical reasons are characterized, as well as their relation to values and desires. The problem of practical syllogism is briefly commented on. In the final part of the article the author proposes five criteria of practical rationality.

Keywords: practical rationality, reason, practical reasons, criteria of rationality, practical syllogism, freedom, values.

1. General Remarks on the Concept of Rationality

The typologies of rationality proposed in the literature on the subject are too varied to comment on them in detail in the present article. The differences between them vividly come to light when relevant works by Ryszard Kleszcz, Mieszko Tałasiewicz, Józef Życiński, Ryszard Szarfenberg and other authors are compared [3], [16], [17], [18], [14], [7]. Differentiated types are usually created by adding a domain adjective to the notion of rationality (e.g. epistemic, methodological, ontological, pragmatic rationality etc.), or a noun defining a field of thinking (e.g. rationality of science, politics, common thinking, decision, economic choice, etc.).

Examining these typologies, one finds it difficult not to throw a fundamental doubt whether they meet the requirements of logical division – even in those (rare) cases when they seem to be complete, doubts remain as to their disjunction, their uniform criterion, and the precision of their scope (whether they include in the scope of X something that is not X). These doubts could be ruled out if the authors of these typologies, firstly, at the beginning presented a clear description of what is subsequently to be divided into types, i.e. rationality itself, and secondly, if the reasons for distinguished types were derived from individual characteristics of the differentiated general rationality. Meanwhile, it is almost a rule in the proposed typologies that although the types are analysed in great detail, it is not characterised almost at all what they are types of (*totum divisionis*), while the reason for specifying such and no other types is not substantively justified, but accepted only on the basis of random, customary divisions, applied to interlocking spheres of thought. These not quite methodological strategies of creating various typologies of rationality observed here can even be ironically described as insufficiently rational. It is hard to resist the conclusion that as a result of such a method of multiplying various types, there are far too many of them. To ask rhetorically: how can one justify the differences (and disjunction!) between such types of rationality as conceptual, logical, methodological, and scientific rationality?

An example of general methodological shortcomings in the construction of typologies of rationality can be the puzzling fact that the author of the broadest study on this subject only at the end of his work (i.e. after characterising various types of rationality) comes to the conclusion that there are meta-principles of rationality which satisfy – as the author writes – "the need for the existence of certain universal criteria of rationality" [3, p. 113]. The text indicates that these meta-principles identify rationality as such – and therefore, as must be inferred, they must appear in each of its differentiated types, because otherwise the type in which not all of them would be present would be differentiated in an erroneous way. I am not making any accusation here against establishing such meta-principles; I merely wish to point out that their articulation was indispensable as a factual basis for the types distinguished. If the types were distinguished in an *ad hoc* manner (without referring to the rules determining rationality), then after revealing these principles, these types should be re-tested to demonstrate that they are types of rationality, and not of something else. It is also not known why the established principles have the characteristics of meta-principles, since they only concern rationality itself. Probably only because the principles previously attributed to individual types were principles of something other than rationality itself – which in turn would imply that the types were not types of rationality.

What I am questioning here - namely, the assignment of a separate type of rationality to almost every way of thinking about the world (science, ontology, epistemology, economics, common thinking, etc.) – seems to be just as unjustified as distinguishing as many types of truth as there are disciplines in which it is present, or as many types of life as there are species of organisms. In all these cases, distinguishing between types is undoubtedly correct – in each of them, however, the identification of the "medium" (or subject) of types is incorrect. If, after all, rationality occurs in X as well as in Y, only its variables change (the kinds of particulars to which it applies), and it itself – as long as it is rationality – remains the same.

Confining myself to this general critical remark against the known typologies of rationality, I would like to add, as a positive proposition, that the question of typologies can, in my opinion, be arranged in the following way. Two types of rationality that should be distinguished – cooperating with each other, but irreducible to one another – are theoretical (or cognitive) rationality and practical rationality. Theoretical cognition of the world and practical functioning in it constitute, on the one hand, the two most general ways of activity of the same reason, and on the other – ways that are so distinct that there are strong reasons for treating them as truly different types of this activity. On the other hand, the numerous types of rationality distinguished in the literature on the subject, which I mentioned earlier, are, in my opinion, models of criteria of justification (or correct argumentation) used in various fragments and disciplines of theoretical knowledge and practical thinking – and in fact there can be as many such models as there are areas in which the ways of justification specific for them were formed. The approach proposed here, on the one hand, gives justice to a certain factual multiplicity (multiplicity of ways of justifying), and on the other, it does not break up rationality to such an extent that the existence of its permanent core becomes doubtful – and therefore there is no need, in order to defend this core, to create additional meta-principles of rationality with their problematic reference to lower (type-specific) principles.

The subject of my further discussion will be only practical rationality – or rather its very essence, described by typologists of rationality, probably interchangeably, as axiological rationality, rationality of values or rationality of goals. I call axiological rationality the core of practical rationality, because I think – contrary to the previously mentioned typologies – that the so-called instrumental (or pragmatic) rationality is in fact not rationality, but only a fragmentary sphere of practical justifications, which (or more precisely: their assumptions) must be consistent with axiological rationality in order to be rational. There are no reasons to consider the selection of effective acts and measures leading to irrational goals as rationality. After all, if someone stuck to a false conclusion and then developed (equally false) premises to go with it in order to obtain a logically correct consequence, then his or her reasoning would not become true.

Axiological rationality is usually neglected by theorists of rationality (mentioned, but not explained in more detail) because its essence is inseparably entangled in the fundamental problems of axiology, which the researchers of rationality do not want to be concerned with. For the same reason, axiologists and ethicists formulate many important, though usually not systematised, conclusions about axiological rationality as it were on the outskirts of their normative theories. On the problem map of philosophy, the topic discussed in the present article is located in the region where theory of rationality and theory of values overlap. The need to address this borderline topic is dictated by the aforementioned circumstance which consists in the fact that, briefly speaking, although both neighbouring fields, i.e. the theory of rationality and the theory of value, need a systematic theory of their common borderland, today's highly specialised competences of researchers leave this borderline highly undetermined.

As far as the concept of rationality itself is concerned, I think that there are no sufficient reasons to distinguish and separate it from the concept of reasonableness. Therefore, I understand rationality as a field of the functioning of reason. It can therefore only characterise the activity of such entities that have the capacity of reason. Therefore, strictly speaking, they can also think and act irrationally.[1] The natural world, on the other hand, can be neither rational nor irrational – contrary to the ideas of the so-called ontological rationality. On the other hand, reason itself is the highest thinking ability we know. It is people and people only that are entitled to it – as it is closely related to self-awareness, self-knowledge, and reflexivity. The term "the highest ability" means that reasoning is directed primarily toward a lower type of thinking, i.e. toward thinking focused on external reality (or on the empirical). This reflective and meta-level nature of reason determines the fact that it is the source of doubts, criticism, and the idea of justification – both the requirement of justification itself and the creation of justifications, as well as the construction of the theory of justification. Therefore, rationality and justification seem to be the same.

These several features of reason suggest that – as in German philosophy – perhaps reason should be distinguished from the lower ability to think, which would be intelligence or intellect. This lower type of thinking would be responsible for using conceptual contents regarding empirical phenomena and links between them, but without reflective questions about justification. The main difference between reason and intellect would be – if such a distinction is accurate – that intellect amounts to thinking which functions as the subject dictates it and in order to serve his or her vital interests (it performs instrumental-adaptive functions), and reason, as reflective meta-thinking, has at its disposal – as Thomas Nagel called it – "a view from nowhere," which means that: (a) on the part of the subject, it is not determined in terms of its content by the psycho-vital sphere, (b) and in turn its object are the limits of intellectual thinking, or the subjective conditions and assumptions of this thinking, as well as what is ultimate on the objective side of intellectual thinking. Therefore, rational thinking, in a natural way for itself, gravitates in the sphere of theory towards epistemological and metaphysical issues, and in practical thinking – towards philosophy of freedom and philosophy of values (or the theory of good).

In order to answer the question about the participation of reason in scientific knowledge (in empirical sciences), one should carefully examine the logical structure of scientific theories and the history of particular sciences, and only on the basis of these studies determine (a) to what extent sciences go beyond recognising correlations between facts (i.e. beyond the procedures of intellect), and (b) how the ways of justifying the claims of particular sciences, distinctive for these sciences, were historically shaped in them. Such an analysis of science is, of course, a separate problem, and here I merely want to point out that in the scientific thinking there is as much reason as there is reflection on justification. I suppose that both the origin of sciences and their prevailing part which constitutes empirical knowledge are rooted in the intellect, and only the historically increasing requirement of justifying scientific knowledge, the need to reconcile its various fragments, and, finally, philosophical reflection on its non-empirical assumptions, included in it the participation of strictly rational thinking and expanded it.

2. Identification of Practical Rationality

Since practical rationality is a rationality, it has some features in common with theoretical rationality, but, in turn, as a practical one it also has properties that distinguish it from the theoretical one.

I have already mentioned the common features. They are namely: (1) metal-level thinking that is autonomous in relation to psycho-vital reasons; (2) the ability to doubt (that is, to question empirical data and conclusions obtained as a result of intellectual thinking); (3) looking for justifications for lower levels of thinking and reflective revision of the assumptions of this thinking; (4) not yet mentioned here, the use of the basic laws of logic (the law of identity, of contradiction, and of excluded middle).

In order to determine the practicality of rationality, categories that identify practice as such (i.e. the specificity of practice) should be distinguished. Practice differs – it can be said: categorially – both from theory and from processes of nature. In contrast to theory, or even to the process or operation of theorising, it is a sequence of changes not only related to thought (and even less is it a system of claims), but it consists in real changes that modify the real course of events. However, these changes – this time unlike natural changes – are not just a consequence of causes and effects, but are initiated and led by consciousness (although in terms of energy they are performed by the body of the acting subject). Thus, practice is a series of changes in the real world, but initiated and directed by consciousness.[2]

I will disregard the question of the difference between practice and natural processes, since the identification of practical rationality (to which I am leading up to) requires a differentiating juxtaposition of practice and theory rather than of practice and nature. More specifically, it is about capturing the most general differences between how consciousness functions in the mode of theorising and how in the modus of practice.

In both these fields – theory and practice – consciousness seems to have three levels. In the field of theory, these are: *sensory perceptions* providing content for thinking, *intellect* (which thinks about the connections between the contents of perception using language and concepts), and *reason*, which in turn, as meta-level thinking, problematises and fills with the content it constructs both the objective and subjective limits of intellectual thinking, at the same time going beyond these limits. In this way, reason in theory produces – speaking in historical order – metaphysical and critical-epistemological issues. It constitutes meta-empirical thinking in the sense that it directs its questions and answers towards the non-empirical conditions of the empirical. Of course, philosophy is not the only manifestation of reason. In everyday life (as well as in non-dogmatic religious speculations) its activity is expressed in questions and hypotheses concerning what is ultimate (including those, of course, which concern reason itself).

In the practical function of consciousness, the first level consists of positive and negative *drive and emotional reactions*. Their contents are – one could say – empirical data for practical thinking. In turn, the *intellect* in the practical function is thinking which, basing on the recognised correlations between objects of emotions, selects the ways of instrumental actions. These actions are admittedly intelligent (and not spontaneously emotional), but the intellect that designs them remains instrumental thinking, because, firstly, it is itself guided by psycho-vital dynamics and, secondly, its calculations are limited to the ways of avoiding discomfort and satisfying desires imposed on it by the leading emotions. This is because the practical intellect does not keep a cognitive distance from the subject's own desires and emotions (they are not an object for it), so it also does not evaluate them, but functions as an instrument of their implementation. Therefore, the goals pursued by it are not critically selected and therefore they have no other justification than the very fact of their energy advantage over the tendencies that are weaker in a given situation. In other words, practical intellect only plays the power game of fears and desires that takes place behind its back. It is not familiar with freedom of thought and reflection on justification.

The meta-thinking of reason becomes practical when the subject seeks justifications for his or her future practical activities, but with the assumption that he himself or she herself is their free agent. If he or she understood his or her future actions as necessary cause and effect chains, he or she would not be able to ask for their justification. We can see, therefore, that when asking for the justification of its future practice, practical reason not only thereby assumes its own freedom, but at the same time challenges what is an uncritical assumption for practical intellect – namely, it questions the treatment of psycho-vital desires as sufficient reasons for the undertaken actions. It is also worth noting that if the subject did not assume freedom and understood his or her future deeds as the necessary effects of his or her psycho-biological desires, it would mean that he or she does not understand his or her future as a practice they pursue, but as a natural process that does not have an author. In this understanding of one's own future, practical reason would destroy itself – it would be pointless. Such self-destruction of practical reason takes place – I believe – in every determinist who, if logically consistent, must translate his or her life only in the perspective of theoretical reason.

On the basis of the above, it must be stated that for practical reason the assumption of the existence of one’s freedom is a *sine qua non* condition of its functioning – without this assumption, all its operations would be based on a fundamental internal contradiction (and thus they would lead to self-destruction).

Freedom, however, is not just independence from causes. For a rational being to be able to initiate and pursue practice as a free activity, he or she must have a positive justification for his or her actions, other than psycho-vital causes (if they were the only determinants of action, it would not be free - so they would not be practice). Practice therefore requires not only that its beginning is independent from causes, but also requires positive determinations that are different from these causes. Without them – so to speak – independence itself would not move off, and practical reason would still have no object. These practical determinations are *practical reasons*. Reasons, in turn, can neither have no object (they would not be reasons then), nor can their objects be facts, beings, or laws of nature – because then they would not be of practical but of theoretical nature and nothing would result from them for practice. It therefore remains to accept that *values* are objects of practical reasons.

If, therefore, we call practical rationality – as I suggest – the functioning of reason that directs practice, we can already mention its two necessary conditions. The subjective condition is freedom, while the objective one – values. On the other hand, practical rationality itself consists of practical reasons, and it enters real practice in the form of decisions.

Reasons are something other than causes, so – what should be emphasised – only acting on the basis of reasons makes this action free. It seems that this issue needs to be interpreted more precisely as follows: the very fact of consciousness makes a person negatively free and, thanks to that, enables positive freedom of decision and action – but only enables it! And only adding something positive to negative freedom – something that “launches” a decision independent of the causes – completes the negative aspect of freedom with a positive aspect (and only then does a person perform a free deed). This positive factor which triggers action and at the same time preserves its freedom is either a single reason or, generally speaking, practical rationality. This means, first of all, that rationality, and also positive freedom with it, are not, like consciousness, facts that occur in people, but that they are normative tasks (which are sometimes achieved and sometimes not) for the consciousness which has a much wider scope than them. Secondly, it also means that only practically rational action is free action – while irrational action, even if its beginning was negative freedom, slides into causal determination. On the other hand, since practical reasons are related to values, it follows from the above diagnosis that free action is impossible without values (i.e. practice is impossible).

To conclude the presented argument in the simplest way possible, it can be stated that in order to act in a free way, one must be rational in a practical way; rationality in turn depends on the proper respect for and selection of values. The problem of values cannot be overlooked here. Explaining what values themselves are, I will introduce, in addition to freedom and values, a third category necessary to identify practical rationality – the category of happiness.

Values as objects of practical reasons can neither be subjective creations of desires nor any kind of objectively existing entities. Both of the interpretations of values mentioned here – in brief – fall into the naturalistic fallacy and are irreconcilable with the necessary assumption of freedom of practice.[3] Any attempts to explain practice that do not take into account freedom as its source must become trapped within the limits of theoretical-descriptive thinking about reality (and about the practice itself, which is then only alleged), and there is no logical transition from such thinking to a normative discourse. The assumption about the existence of freedom (or the point of view of freedom) sets – as it seems to me – the only perspective in which one can sensibly talk about values. “Sensibly” means here neither falling into the naturalistic fallacy nor (by reducing values to the ontic causes of action) denying freedom. Only then is practical rationality not reduced to theoretical one.

In order for values and freedom not to exclude one another, and for the former not to reduce themselves to the realm of beings, they must be interpreted as something which in the face of freedom is its necessary complement (and not an antagonistic factor). Values are indispensable for freedom in the sense that freedom cannot survive real changes without them (i.e. it cannot retain its self-continuation). For if the change taking place in a person is not directed by free affirmation of some value, then this change is determined by causes, degrading the human being to the position of a multiple-conditioned automaton. It means that a person cannot preserve their freedom in any other way than through the affirmative recognition of values. Only such affirmation makes it possible to set oneself and to achieve a goal alternative to what would emerge from an inert cause-effect sequence. Values are therefore objective conditions of the persistence of freedom – after all, freedom is not a substance that could last independently of the changeability of the world. But they can be non-threatening to freedom only when they are dependent on it themselves – when it establishes their validity. What are values, then?

My construction of the philosophy of values – in brief – is as follows [11]. The sphere of values and the sphere of beings are objectively one and the same reality. And what requires their differentiation are different points of view of the same reality. Different points of view also force the use of different concepts in relation to the aspects of one reality determined by these points of view. From a theoretical point of view, reality is a sphere of beings without values. On the other hand, from a practical point of view – in other words, from the point of view of freedom – reality, in turn, is not the domain of beings but of valuable objects.

My answer to the question where the valence (or the axiological significance) of existing objects comes from is as follows: its source is the absolutely first free decision of the existing subject – not the first in chronological terms, but logically, i.e. the one whose content is logically first in relation to the content of all other decisions. Its primary nature consists in the fact that the subject chooses in it not these or other existing objects, but one's own existence in the world. “Own” means here: the existence of oneself as a free subject. In simple terms, reality is valuable only because free subjects want to exist in it, continuing their freedom. This is the basis of my philosophy of values, while the rest of its claims are subtleties and conceptual and logical details; there is no space here, however, to elaborate on them [11], [12]. Let me just mention that in the development of this conceptual and logical instrumentarium, the most important thing is to consistently carry out conceptual analogies between ontological and axiological categories (which I try to show meticulously in my other works). What I consider to be the advantages of this concept is that it does not fall into the naturalistic fallacy (because the source of value is a decision, and not some beings), that it does not reproduce the dualism of being and values, and that it seems to avoid the typical difficulties of axiological subjectivism. However, the burden of explanation shifts to the clarification of differences (and mutual references) between the two above-mentioned subjective points of view. In the area of required clarifications, there is also the question of practical rationality.

Concluding the thread of values, it must also be added that the question of why free beings want to exist (or why their logically original decision is positive) does not seem to have a more accurate answer than the view that philosophy has always held: that the ultimate profit which the subject draws from his or her participation in reality is smaller or greater *happiness*. And in fact I think that without a concept of happiness philosophy cannot satisfactorily explain values, practice, or even practical rationality. Without referring to this concept (or a synonymous one), there are no good answers to the questions about the first reasons of subjective action. What I have in mind here, for example, are such questions as: why is it better to live than not to live? Why act at all and not remain a drifting object? Why should I be moral? And finally, why is it better to be rational than irrational? If we do not take into account and do not name a specific “profit” which the subject receives as a result of the undertaken positive involvements, the fact of these involvements will remain inexplicable and incomprehensible. The category of happiness, being the third – apart from freedom and value – category of practice, is a category that also identifies practical rationality.

As we already have the most important categories of practice at our disposal, we can conclusively distinguish practical rationality from theoretical one – and they differ in almost everything except the presence of attributes of reason in them. While the subject of theoretical rationality is the sphere of beings (or facts), the subject of practical rationality is the sphere of values. While the subject of the former is impersonal reason (impartial, individualised, leaving aside values), called theoretical reason, the subject of the latter is individual freedom existing in the real world, which uses reason to ensure its own persistence in an environment of permanent change. And finally, the third difference concerns objectives: for theoretical rationality, the goal is the universal and impersonal truth of theory (objective knowledge), while the objective of practical rationality is the individual happiness of the subject, obtained as a result of free action directed by values.

3. Practical Reasons Between Freedom and Values

I have already mentioned the relationship between practical rationality and freedom: namely, it is two-sided. On the one hand, negative freedom, or independence from causes, is a necessary condition for practical rationality – after all, without this independence, it would be pointless to construct justifications for practical actions (it would only be Marx's false consciousness). On the other hand, only practical rationality complements negative freedom with a positive aspect, that is, it makes it able to perform real actions – for without positive reasons provided by rationality conscious action would either be unable to "move off" or would fall to the level of determined natural processes. The question remains, however, where the contents of practical reasons originate from, since they must refer to values, whereas reality – as I have stated earlier – consists only of beings and not of values.

The contents of practical reasons come from rational affirmation, which the subject's freedom directs towards individual beings as the conditions of his or her existence and happiness. To put it more precisely, practical reasons have three aspects: *content, practicality, and rationality.* (a) The *content* itself is a mental correlate of the purely ontic definiteness of something that exists (it is axiologically neutral in itself, as is its ontic counterpart). In turn, the *practicality* of these contents (and this is what gives them axiological character) comes from the attitude towards the reality of two subjective factors: first, the one that the Greeks called *thymos*, or the drive and emotional sphere, but to a decisive extent only from the affirmative relation to freedom. Their *rationality*, on the other hand, consists in the fact that they are interrelated in a logically correct manner and that they are justified by reason (which is why they are reasons, not just motives or desires).

To explain briefly the constitution of practical reasons, one can refer to the known figure of the so-called practical syllogism (formulated by Aristotle and developed by John Stuart Mill, as well as Donald Davidson).[4] It is known that the scheme of this syllogism shows that an individual imperative conclusion results deductively from the general greater imperative (or value) premise and from at least one lesser descriptive premise accompanying it. Practical reasons are what I call these larger value premises and the above-mentioned imperative conclusions. The deepest problem (which was not solved satisfactorily by Aristotle or the subsequent researchers of practical syllogism) concerns the sources of the greater premise which would be the absolutely first and not one of intermediate greater premises. It is obvious that this absolutely first axiological premise must concern what distinguishes the entire domain of values from the domain of being that is not concerned with values. It is easy to guess that in the light of what I have stated earlier, this absolutely first premise originates from the logically first decision of the subject, in which his or her freedom affirms his or her existence. Because of the primary nature of this decision, I call it the *primaeval decision*. Because the subject cannot exist without the world, the practical choice of one's own existence is identical with the affirmation of existence in general. Therefore, the primaeval decision establishes the first practical reason, which is the first axiological premise for all other practical reasons, narrower in their scope. The simplified content of this first premise – its very essence – is as follows: “existence is valence.” The former theory of transcendentals expressed it in another version: “existence is goodness.”

Before I show a manner of particularising (i.e. a differentiating fragmentation of content) of this first practical reason, I would like to point out that it constitutes the basis of general axiology. Since existence is valence, it follows – analogically to ontological inferences – that specific existing beings are valuable objects. But because their axiological positivity is constituted by the affirmative primaeval decision of the subject, the values that they are entitled to are not – strictly speaking – their own qualities (such as ontic properties), but their relations to the freedom and happiness of the subject. They are practical relationships, expressed in conceptual discourse with practical categories. Depending on how we logically classify these practical relationships, we will obtain different sets of axiological categories, such as: moral value, epistemic value, aesthetic value, vital value, hedonistic value etc. However, in order for these relations to remain practical (and not ontic), that is, to remain values (and not beings), they must be sustained on the part of freedom via practical rationality. This means that even such unreflective and spontaneous relations as e.g. hedonistic values change from facts into values only once practical rationality affirms them or approves them. It is only when they obtain a practical reason that they become values.[5]

However, while addressing the problem of how the first practical reason is divided into a multiplicity of reasons with a smaller scope, one must already refer to the actual diversity existing in reality. Three circles of diversity are in question here: (a) the multiplicity of subjective desires; (b) the multiplicity of objective things and events; and (c) the multiplicity of types of practical action (this multiplicity is called a multiplicity of practices by e.g. Alasdair MacIntyre [6, pp. 7-10, 340-364]). These three areas of multiplicity concretise and differentiate the content of practical reasons, which makes it possible, firstly, to make specific decisions, and secondly, to apply practical rationality in these decisions.

When describing the nature of practical reasons, it is worth emphasising that their very (previously explained) content is powerless in the sense that thinking it is not able to cause decisions or actions in the subject. This is consistent with the fact that this content is axiologically neutral (a) as long as the subject does not want to continue his or her own existence, and (b) until he or she notices and understands the relationship between the content and his or her own existence. The content of reasons, on the other hand, obtains the power to determine the actions of the subject only when the two aforementioned circumstances change into positive ones – that is, when the subject maintains his or her primaeval decision and when he or she applies practical rationality. It is then that he or she gives the powerless reasons a practical character, and thus the power to direct action. The power of these reasons ultimately comes from the subject’s primaeval decision. In the field of practical rationality, the following principle applies: that for some reason to be able to cause an act of the subject, it must – assuming the rationality of the subject – have a logical connection with the content that the subject already wants. If he or she did not want anything, then no reason would have a causative power for his or her action – and that would mean that it would not be a practical reason. For example, a reason that claims that one should pass another exam at university overcomes the psychological reluctance to learn once the student realises the necessary connection between passing the exam and his or her own desire to obtain a diploma. If this willing was not in the subject, this example of a reason would not move him or her.

Constituting practical reasons by some willing that is more primal than them (it can be called the involvement of the subject in reality) and thus establishing their binding character is what distinguishes them from theoretical reasons. In order for a judgement to be a practical reason, neither its imperative grammatical form nor the logical correctness of its connections with the system of similar judgements is sufficient. In order for imperative judgements of a certain internally coherent system to be practical reasons, at least one of these judgements must be connected with the real subject in such a way that the content of this judgement is both the content of the subject’s act of willing – and it has to be free willing (non-free willing is not willing, but only a causally created fact of desire). Since imperative judgements do not result from judgements about reality, then any their logical system remains non-binding for the subject – that is, it is not a system of reasons – until the content of any of the judgements is the content of the subject’s decision.

However, when a subject wants to decide something in a free way, the freedom of his or her decision depends on whether it is rational – and it is rational when its content is not contradictory both with its more general premises and its more detailed logical consequences. Therefore, for the subject to make one free decision, he or she must at the same time decide on the correctness of an entire system of imperatives which is affirmed by this one decision. If the system contains the imperatives A. B, C, D, etc., then the subject cannot choose for example the imperative C, and reject the others – as then his or her decision, not being rational, would not be a decision.

So if the content of a decision of his or hers (such as the decision to enslave another person) was contrary to the fact that he himself or she herself lives and makes decisions – that is, accepts an imperative affirming the value of subjective, free life – this decision would not be rational. In order to make it rational, he or she would have to cancel his or her decision affirming his or her own life and freedom (i.e., adopt other imperative assumptions) – and then, of course, he or she could not do anything.

In contrast to practical reasons, the binding nature of theoretical reasons is not rooted in the decisions of the subject, but comes either from the principles of logic or from empirical evidence. Consequently, for a theory to be true, the subject's consent is not needed. If the individual subject does not agree with the real theory, then it does not lose out, but he or she does. On the other hand, in the case of validity of imperatives, it is a bit different: as long as inconsistency in the decisions of the subject is fragmentary (local), then this circumstance is devaluing for the subject, not for imperatives (just as in theoretical thinking), but if the subject cancelled his or her primaeval decision – if he or she freely refused to continue his or her being a subject – with such lack of a positive primaeval decision there would not exist any practical reasons that would be binding for him or her. In other words, the possible lack of a positive primaeval decision invalidates, it seems, all systems of imperatives – but it can only happen when this factor occurs, and is the unique factor that can have this effect. Taking the positive primaeval decision as the basis, the difference between practical and theoretical rationality can be concluded as follows: ignoring practical reasons causes in the subject a deficit or complete loss of freedom and happiness, while failing to respect theoretical reasons – as long as they are not related to the personal situation of the subject – only puts him or her in the state of ignorance or cognitive error.

4. The Functioning of Practical Rationality and its Criteria

The area to which practical rationality applies and where it operates is the area of the subject's desires and their references to objective reality – desires are in fact the driving forces of action. The purpose of its functioning is the permanent ordering of relations between desires and reality – namely, such ordering that would be correct in terms of their compliance with the primaeval decision and the subject's endeavours to maintain and increase happiness (i.e. to achieve practical effectiveness). If we treat happiness as a reason of the primaeval decision – and I think that its role in the structure of practical rationality should be understood in this way – then we can define this rational correctness of desires as a reconciliation of their multiplicity with the first principle of practice established by a single primaeval decision. Since the area of concern of practical rationality are the relations of desires to the objective world, it ensues from it that the domains from which it must derive its contents are the sphere of subject's self-knowledge and the sphere of descriptive (scientific and ontological) knowledge about the objective world. Adding axiological reasons – logically consistent with the first practical reason – to both these spheres from the outside, practical rationality transforms the facts occurring in these spheres into valuable objects. Axiology – if I could allow myself an aphorism – is rationality imposed on the real world by freedom, which wants to satisfactorily persist in it.

It can be said that in relation to desires, practical rationality functions in two ways: it appears that it is more often inhibitory and sometimes generative. There are always more such facts as desires in the subject than rational desires – that is why conclusions from practical reasoning hold back most desires and allow only those to function which have gained the approval of rationality. This happens more or less in the same way as when Leibniz's God chose one of the many possible worlds to be created – all "strove for existence" [5, 485] but God did not stop the existence of only that which was considered the best by God's reason. But it also happens the other way round; namely, it is only an imperative conclusion (some individual practical reason) that generates an individual desire which – although it was not present in the subject – after reasoning becomes the motive of the deed. This is always the case when the current empirical situation rationally forces the subject to – for the purpose of achieving the intended general goal – do something that he or she does not actually want at the moment (e.g. making an ill and lazy person take up appropriate exercise in order to save their health). It is worth emphasising, however, that even in the latter case, a specific desire is not produced by a reason (a reason only justifies the need to arouse it), but by that desire with a wider scope for whose realisation this individual desire is necessary. This means that in such cases, practical conclusions serve the purpose of concretising the more general content of desires to the form of individual content – and only in this way shape the act of specific desires.

Practical rationality, as meta-thinking about desires and their objects, does not produce desires by itself, but only logically orders their contents. However, in order for this ordering to be practical, it must – as I have already mentioned – use three types of content: (a) the content of the first practical-axiological premise; (b) the contents of psychological desires, (c) the contents that create descriptive knowledge about the world – including the descriptive content of the subject's self-knowledge.

Therefore, the *criteria* of practical rationality must concern the appropriate composition and interdependencies between these three types of content. Since the criteria are measures of whether given practical reasoning is rational or not, they can be identified with *the principles of practical rationality*. These are the principles without which rationality simply does not function. They are its foundations or constitutive elements.[6] At a high level of generality, these criteria can be characterised as follows:

1) *The criterion of the affirmation of being*. A positive primaeval decision, in which the subject, affirming his or her own freedom and rationality, also affirms the existential conditions of his or her existence, is the criterion of practical rationality in the sense that it provides the first imperative premise without which it could not be decided whether fragmentary practical reasoning and desires themselves are rational or not. The very fact of the occurrence of a desire is not a reason for its legitimacy (rationality). If the imperative premise is necessary for the reasoning to be both logically correct and practical, then it is clear that without it all inferences that would concern action would have to lack either logical correctness or practical character – so they would not be practically rational.

The concepts of axiological rationality which state that the rationality of axiological reasonings (or those concerning goals) consists in their compliance with the values of a given culture or society [3, p. 75], or in compatibility with some facts of nature (e.g. with the human nature or the so-called natural human needs) all fall into the same error. It invariably consists in the fact that such concepts propose recognising as rational such reasoning which logically derives imperatives from descriptive premises – all in all, the invoked measures practical reasoning must be compatible with in order to be rational (whether these are axiological models from a given culture or any qualities of man) are nothing but facts. In view of this kind of concept, the following legitimate question always arises: are the measures which practical reasoning should be compatible with rational themselves?

2) *The criterion of criticism and self-knowledge.* The shortest explanation is that the principle of criticism is to doubt the legitimacy of one's spontaneous desires and, consequently, test their validity or justify them. The psychological fact of the occurrence of a desire is not in itself rational – this desire can only be considered rational as a result of rationally checking the connections that exist between its content and its general surroundings. In particular, it is about establishing the relation of this single content to: (a) the content of the primaeval decision, (b) the content of other desires, and (c) the knowledge of actual reality. The next criterion (3) determines what kind of these relations makes a desire rational. Self-knowledge, on the other hand, is indispensable not only for a reflective realisation of the assumed primaeval decision, but just as much for knowing what makes us happy and to what extent. By revealing the position of an individual desire in the subjective hierarchy of desires, and thus the degree of its effectiveness for happiness, self-knowledge is also a *sine qua non* condition of axiological knowledge of the

hierarchy of values[7] (see criterion 5). Self-knowledge also protects one from erroneous identification with the desires of others, those promoted by ideologies, advertising, social fashion, or environmental pressures. It can be said that criticism and self-knowledge are an irreplaceable antidote against the threats to rationality that have their sources both in irrational social trends and in the brutal biological-emotional spontaneity, particularly intensified in the period of youth.

3) *The criterion of non-contradiction.* While the first criterion concerned only the content of the first axiological and imperative premise, and settled the question of its validity for the subject of action, and the second served the purpose of distinguishing desires that are firmly established in the subject and desires that bring him or her happiness from accidental and "mistaken" desires, the present criterion determines the rationality of all contents of desires. Because the required relationship of non-contradiction may exist (or may be missing) in several different areas, the current criterion requires division into relevant regions. And thus, for the content of a desire to be rational, it must be non-contradictory with: (a) the first axiological premise, i.e. with the content of the primaeval decision; (b) with the content of other specific desires, insofar as those are rational; (c) with the content of real possibilities (it cannot concern what is realistically impossible); this principle can be specified by taking into account particular temporal moments: non-contradictory with the present possibilities (relativised to the current situation) or non-contradictory with the universal possibilities resulting from the laws of nature.

Mentioning the requirement of non-contradiction of desires with the content of the primaeval decision, it is also worth commenting on the question of the practical syllogism. The theory of this syllogism proclaims, as it is known, that between the general imperative premise and the detailed imperative conclusion – after adding a relevant descriptive judgement to the former[8] deductive reasoning takes place. Strictly speaking, such reasoning does not occur for two reasons. First of all, if there is a premise that I want to eat fruit, it does not follow from it that I should eat an apple, but only that I should eat an apple, a pear, a plum, etc. It means that between the general content of the premise and the detailed content of the conclusion there is a gap that the subject fills in with a free choice each time, adapting its content to his or her individuality. And this applies to every stage of concretisation – that is, every transition from any general premise to a specific decision. The second reason for the lack of strict reasoning in the so-called practical syllogism is the nature of descriptive lesser premises. They report, as is known, every empirical situation in which the subject finds himself or herself. Despite the fact that the imperative premise is general, both the subject that employs it and every situation are individual (that is, they differ from other subjects and other situations), and this in turn means that the concretisation of the general premise, which depends logically on the descriptive lesser premises cannot be always identical – that is, it cannot be deductive reasoning. In other words, because an individual subject is a component of the empirical situation, his or her preferences must be taken into account in its description (as long as, of course, they are rational) – which in turn leads to the conclusion that he himself or she herself chooses an adequate description of the situation.

For the aforementioned reasons, neither the rationality of the content of desires nor the rationality of specific decisions can be based on the fact that they are strictly deduced from more general premises (including the absolutely first imperative premise), but only on the weaker relation of non-contradiction. It is also worth noting that if desires and decisions – in order to obtain the value of rationality – were to result by deduction from more general premises, it would lead to a very baffling conclusion about all practical rationality. It would mean that the only act of freedom is the primaeval decision – all the remaining "decisions" would essentially be a stoic consent to the individual links of chains of logical reasoning that, after all, have the attribute of necessity. The human subject would then be in a situation similar to that of Leibniz's God – he or she would indeed choose the existence in the world in which he or she was born, but because of the logical necessities that govern this world, he or she would not be able to choose anything else in it.

4) *The criterion of realism and effectiveness.* In addition to the positive primaeval decision (establishing the first imperative premise) and logical non-contradiction between the contents of desires, rationality of action also requires that both the desires of the subject and their specific selection performed in particular decisions should be shaped with a significant participation of the descriptive knowledge of the real world. Although descriptive knowledge does not justify desires

and decisions, it is a criterion distinguishing the field of desires and rational decisions from the sphere of fantasy, dreams, and choices that "do not take reality into consideration." In the face of all the contents of desires and the arbitrary excess of imperative judgements, descriptive knowledge performs the function of "sifting the wheat from the chaff" – such knowledge exposes desires with unrealistic or mutually excluding contents, as well as unjustified imperatives, as irrational. The functions of descriptive knowledge for practical rationality are in general as follows: (a) only on its basis can we distinguish real possibilities from real impossibilities; (b) effectiveness of action, predicting its effects, or the appropriate counteraction against the incoming threats are not possible without knowledge of causal relations; (c) the subject's self-knowledge (which is also a kind of descriptive knowledge) constitutes the necessary information for him or her about what actions and states of affairs make him or her happy, to what extent, and which of them he or she should prefer over others; (d) and finally, empirical knowledge about each situation and its conditions allows to decide which of the rational desires should be realised and in which situation. The rationality of the desire which justifies a given act is only a necessary condition of the rationality of an act, but it is not a sufficient one. The features of the situation may be such that, for example, some rational desire has no chance of being realised (while in another it can be easily realised) or its effective realisation would have to violate many other values, and as a result, the balance of the results of this realisation would be negative. It may also be that the fulfilment of a rational desire in a given situation excludes a value higher than the value of that fulfilment – while in other situations both values would be reconcilable. In all of these examples of situations (and probably their other variants are also possible), the realisation of a given desire, although it is rational in terms of its logic – that is, it is non-contradictory with the content of the primaeval decision and the content of other desires – is irrational either because of ineffectiveness or because of unfavourable final balance. Both of these reasons, therefore, imply irrationality of action because they indicate a deficit in happiness, that is, a decrease in the very thing that was supposed to be sustained and increased as a result of the action. To sum up, practical rationality requires that its necessary condition in the form of rationality of desires themselves be supplemented with the sufficient condition in the form of rationality of the act. The criterion of the latter rationality comes from (with the former criterion met) the empirical knowledge of each situation.

5) *The criterion of the hierarchy of values.* In the colloquial version it may sound like this: do not confuse the means with the end or the soil with the crop. In the axiological discourse, in turn, it can be expressed as follows: the ontic hierarchy of basic and secondary elements is not identical to the hierarchy of values that these elements are entitled to. In other words, the fact that bodily life is the ontic basis of spiritual acts and creations does not lead to the conclusion that the values of bodily life are higher than those of spiritual life. In the human world, it is the ontically secondary layers of reality, such as the sphere of the spirit and culture, that have higher values, which in the axiological balance of one's life are far more important than the values of the body or social well-being. It is only in this spiritual sphere that a human being can obtain a happy fulfilment of his or her possibilities and the need for the meaning of life. Violating or ignoring this hierarchy in practical endeavours is therefore irrational, because it decreases the level of happiness of individuals and societies, and in the extreme and mass dimension it can even cause social pathologies and anomies (which was discussed not only by Emile Durkheim, but in a broader perspective also by Erich Fromm, or by Max Horkheimer in *Critique of Instrumental Reason*). The question of this hierarchy would require wider axiological explanations concerning a number of topics; the present article, however, has no space to provide them.[9]

Finally, it is worth mentioning the problems that the proposed approach suggests and which should be developed and analysed in detail.

First of all, it is the problem of mutual relations between practical and theoretical rationality. Focusing the discussion on the first one, I only marginally mentioned the most obvious differences between them, basically neglecting their common features and their interpenetration. However, this issue would require a more detailed investigation. It is not only theoretical knowledge – which I have already emphasised – that is necessary for practical rationality, but also vice versa: it seems that the sources and some elements of practical rationality form part of the foundations of theoretical rationality (confirming to some extent Immanuel Kant's claim about the primacy of practical reason over the theoretical). However, at the core of the practice of theoretical argumentation, there are, it seems, some important practical reasons which indicate not only values and goals, but also the conditions of scientific knowledge.[10]

The second problem that I had to omit is the issue of this sphere of practical reasons that creates the field of morality. Although this sphere is a subset and concretisation of the whole domain of axiology (and therefore subject to general rules of practical rationality), the specificity of its area of concern, which is limited to relations between subjects, would require a significant complementation of these general rules with rules governing only this – so to speak – practical moral sub-rationality. The most important difference between the moral sphere and the remaining area of axiology is that the conclusions of practical moral inferences cannot be relativised to the subject's individuality (moral obligations are universal), whereas in the non-moral axiological area, each set and hierarchy of values chosen for realisation are rational as long as they are adapted to each of the subject's individualities. Because, therefore, moral values are universal, and extra-moral values are relativised to subjective differences, the realisation of the latter on the part of every subject would be rational only if it did not violate moral values. I analysed the issue of argumentation required in ethics itself in other articles [8, pp. 211-270]

The third issue, which should – perhaps – be thoroughly examined and developed, is an attempt at logical formalisation concerning the rationality of the act (referred to in criterion 4). This formalisation, if it were successful, would unambiguously establish what kind of logical relations between imperative and descriptive judgements are required to make a given act practically rational in a given situation. It would therefore be an improved version – if it is possible at all – of the practical syllogism.

References

1. Apel, K.-O. *Transformation der Philosophie*, Frankfurt am Main: Suhrkamp, 1973, vol. 2.

2. Chmielecki, A. Rozum i racjonalność – rozważania esencjalne, In A. Chmielecki (ed.), *Rozum i przestrzenie racjonalności*, Gdańsk: Wydawnictwo Uniwersytetu Gdańskiego, 2010, pp. 11-56.

3. Kleszcz, R. *O racjonalności. Studium epistemologiczno-metodologiczne*, Łódź: Wydawnictwo Uniwersytetu Łódzkiego, 1998.

4. Kopciuch, L. *Wolność a wartości. Max Scheler, Nicolai Hartmann, Dietrich von Hildebrand, Hans Reiner*, Lublin: Wydawnictwo UMCS, 2010.

5. Rozprawa metafizyczna, transl. by S. Cichowicz, In G. W. Leibniz, *Wyznanie wiary filozofa*, transl. by various authors, Warszawa: Wydawnictwo Naukowe PWN, 1969.

6. MacIntyre, A. *Dziedzictwo cnoty. Studium z teorii moralności*, transl. by A. Chmielewski, Warszawa: Wydawnictwo Naukowe PWN, 1996.

7. Morawiec, E. *Wybrane filozoficzne koncepcje rozumu ludzkiego i racjonalności*, Warszawa: Wydawnictwo Liberi Libri, 2014.

8. Niemczuk, A. *Filozofia praktyczna. Studia i szkice*, Lublin: Wydawnictwo UMCS, 2016.

9. Niemczuk, A. Kolista struktura praktyki. Rozważania metapraktyczne, In A. Niemczuk, *Filozofia praktyczna. Studia i szkice*, Lublin: Wydawnictwo UMCS, 2016, pp. 126-131.

10. Niemczuk, A. O poznaniu wartości, In A. Niemczuk, *Filozofia praktyczna. Studia i szkice*, Lublin: Wydawnictwo UMCS, 2016, pp. 87-98.

11. Niemczuk, A. *Stosunek wartości do bytu. Dociekania metafizyczne*, Lublin: Wydawnictwo UMCS, 2005.

12. Niemczuk, A. *Traktat o złu*, Lublin: Wydawnictwo UMCS, 2013.

13. Schnadelbach, H. Racjonalność i uzasadnianie, In T. Buksiński (ed.), *Rozumność i racjonalność*, Poznań: Wydawnictwo Naukowe IF UAM, 1997, pp. 37-50.

14. Szarfenberg, R. *Podstawy i granice racjonalizacji polityki społecznej*, http://rszarf.ips.uw.edu.pl/pdf/doktorat_calosc.pdf (12.07.2019), especially chapter 3: Racjonalność i racjonalizacja, pp. 111-196.

15. Sztombka, W. Karla Otto Apla koncepcja dyskursu i uzasadnienia moralności, In J. Ziobrowski (ed.), *Etyka u schyłku drugiego tysiąclecia*, Warszawa: Wydawnictwo Naukowe SCHOLAR, 2013, pp. 122-149.

16. Tałasiewicz, M. O pojęciu racjonalności, part I and II, *Filozofia nauki* 1-2, 1995, pp. 79-100; 3, pp. 39-57.
17. Życiński, J. *Granice racjonalności*, Warszawa: Wydawnictwo Petrus, 1993.
18. Życiński, J. *Teizm i filozofia analityczna*, vol. 1, Kraków: Wydawnictwo "Znak", 1985.

Notes

[1] The identity of the concepts of "rationality" and "functioning of reason" as far as their scope is concerned is assumed, it seems, by Herbert Schnadelbach [13].

[2] I wrote in more detail about practice in [9, pp. 126-131].

[3] I presented a detailed critique of both subjectivist and objectivist understanding of values in [11, pp. 175-200].

[4] I described and commented on this syllogism in [11, pp. 283-288].

[5] Such an example of a reason which is correct and probably most often used is in fact very simple and is as follows: because I have a desire for pleasure X, and there are no practical reasons that would not allow it (i.e. pleasure X does not conflict with more important values), I choose pleasure X.

[6] E.g. Andrzej Chmielecki understands the principles of rationality in the following way: "Like the principles of logic, also the principles of rationality cannot be derived from something else – they are the immanent laws of the functioning of the subject's spiritual acts, belonging to the set of the first principles. They can only be determined a priori, by means of relevant essential analyses. Thus, they are something generally valid, universal, independent of any individual subject. The subject acts rationally if he or she 'participates' in them [...] (if his or her acts are compatible with them); he or she does not need to know them explicitly, however" [2, pp. 47-48].

[7] I wrote about the significance of self-knowledge for axiological knowledge in [10, pp. 87-98].

[8] The suitability of this judgement lies in the fact that its content has to be related with the content of the general premise, e.g. "You should *help* your friends" (greater imperative premise). "Adam is a friend and needs *help*" (lesser descriptive premise). "Adam should be *helped*" (an imperative individual conclusion).

[9] Nicolai Hartmann with Max Scheler were involved in an interesting discussion about the hierarchy of values and dependencies between higher and lower values [4, pp. 77-89].

[10] Broad and significant argumentation in support of the claim about the necessary normative conditions of theoretical knowledge has been presented in modern times by Karl-Otto Apel [1], [14].

Studia Humana
Volume 8:2 (2019), pp. 69—78
DOI: 10.2478/sh-2019-0019

Can the Sense of Agency Be a Marker of Free Will?

Paweł Balcerak

University of Rzeszów,
Rejtana 16c Av.,
35-959 Rzeszów, Poland

e-mail: pbalcerak@gmail.com

Abstract:
In this paper, I will analyse the relation between a sense of agency and free will. It is often proposed that by investigating the former, we can find a way of judging when an action is voluntary. Haggard seems to be one of the authors believing so. To answer if this assumption is correct, I will: 1) analyse the categories of free will and agency; 2) define the sense of agency; 3) describe ways of investigating the sense of agency; 4) describe models of emergence of the sense of agency; 5) analyse the relation between agency and responsibility. I will end by discussing the actual possibility of using the sense of agency measurements (as described in experimental sciences) as markers of free will.

Keywords: sense of agency, sense of ownership, free will, responsibility.

1. Introduction

The question of free will has fascinated humanity throughout its entire history. Minds of greatest philosophers were harnessed to answer this question, and still today this debate is far from being resolved. However, recent years have seen an emergence of research based in psychology, cognitive science, neuroscience, and experimental philosophy that tries to naturalise said problem and find measurable aspects of this phenomenon. In this paper, I will analyse the problem of a sense of agency from the perspective of free will investigations. In the context of free will, we can distinguish free will per se from our experience and a belief in free action. Agency itself is a complex phenomenon, it requires a similar distinction between actual agency and our belief or experience of being an agent in a certain action, thought, etc.

Usually, we take for granted that we possess a body and that we can act upon the world. Parallel to the sense of agency, we can describe a sense of ownership, that is a feeling of mineness that we perceive towards our body, feelings, and thoughts [16]. The sense of agency, on other hand, refers to the experience of initiating and controlling an action [31]. Both experiences seem to play an important role in our life [2]. However, in this paper, I will concentrate only on the sense of agency. As Patrick Haggard writes: “As noted above, a genuine sense of agency clearly requires some internal state of volition, conation, or ‘urge’” [18, p. 196].

How should we understand this ‘volition’, what is it in a metaphysical sense, and can it be found by research using “hard science”?

The choice to concentrate on the sense of agency was made, because the question I am trying to answer is: can the sense of agency be considered an actual marker of free will? The sense of agency, in opposition to free will per se, appears to be measurable and useful for sciences outside philosophy [34]. This category appears, among others, in neuropsychology [7], experimental psychology [42], and cognitive neuroscience [10]. In this work, I will analyse what exactly the sense of agency is in each of these situations and can it really let us measure actual free will.

2. Free Will and Agency

Let us start by analysing briefly what a belief in free will entails and how it connects to the sense of agency. Belief in free will is an abstract idea that people have the ability to act freely. Both by having knowledge of alternative options and by having the ability to choose any of the options without constrains [23], [24].

It appears that most cultures operate on the basis of some belief in free will [39], but, even if that is true, we accept that the degree to which we see ourselves and others as free vary [1]. Scientists performing research in domain of psychology attempt to create tools allowing for measurement of endorsement of the belief in free will. Tests like that usually emphasise different aspects of the philosophical definition of free will. One such test is called The Free Will Inventory [33]. It consists of 29 items divided into two parts. Part one consists of five items designed to measure the strength of a belief in concepts such as: free will, determinism, and duality. Part two consists of statements designed to explore interplay between the attitudes about free will, determinism, choice, the soul, predictability, responsibility, and punishment. In tests like this one, and generally in the experimental approach to free will, we can notice a strong belief in a link between the concepts of choice and free will [9]. I will return to this connection later in this work.

The prevalent belief in free will raises a fundamental question – Why would anyone endorse this idea? To answer this question, let us look at some theories of free will function. On the one side, free will can be seen as a mechanism allowing a person to pursue one's desires, goals, wants, and needs [20]. In that context, free will is only worth having if it allows an individual to follow self-enhancing activities – where self-enhancement is understood as achieving one's goals [8].

On the other side, we have a theoretical position that can be called "action-control perspective." This theory presents free will as a means that evolved to allow the self to coexist with others in society by overriding the biological urge to focus only on personal needs [25]. Impression of free will could have possibly evolved to allow people to deal with a world of complex societal interactions requiring coordination, prospection, planning, and inhibition of self [26], [37].

The close relation between free will and a moral responsibility enforces the view that the concept of free will is strongly embedded in social consideration. This concept may be seen as an explanation to the predicament of associating determinism with inevitability, thus reducing accountability for actions. For instance, Kathleen Vohs and Jonathan Schooler [45] found that inducing a disbelief in free will – using a set of prepared statements about determinism – led to an increase in dishonest behaviour. Based on these observations, we can see the belief in free will as a social tool. After all, a belief that a person could have made a different choice is considered essential in most legal systems to attribute responsibility. Societies usually adjust legal and moral judgement based on the assessment of whether an action of a person was done out of his or her free will. In usual circumstances, that mean a person has to choose to perform a certain action by his or her own volition for that action to be considered a crime.

Simultaneously many, if not most, voluntary actions appear to be "phenomenally thin" [41]. That means we are not aware of most decision processes that lead to our actions. It seems like we perform many of our actions "automatically", even if in reality some kind of mental process is preceding those actions. This "thinness" does not hinder our ability to produce feeling of control over what we are doing. However, this feeling can disappear in certain situations, lets considered Haggard's example:

> (…) a simple example demonstrates the importance and careful construction of the sense of agency. When it gets dark, I may reach out to switch on the lights, perhaps barely aware that I am acting at all. However, if my hand fails to touch the switch, or if the light fails to come on, I will experience a striking conflict and violation of expectations as a result of mismatch between the intended and actual result of the action. In this scenario, the normal experience of fluently controlling the environment is suddenly interrupted as the sense of agency is lost [18, p. 197].

Based on that observation, Haggard argues that criminal and moral responsibility requires not only freedom of action, but in the first place, a sense of agency for a certain action [18, p. 197]. He states that the responsibility requires not only that the agent performs a certain action, but also that they know the nature and quality of said action. This, in his opinion, implies that the agent should experience a sense of agency towards this action.

3. What is the Sense of Agency?

Philosophical reflection upon the phenomenon of a sense of agency allows us to put forth some observations. To begin with, the sense of agency is a complex and non-homogenic structure. Many authors argued that several separate levels of this phenomenon can be distinguished [22], [15]. An influential conceptualisation comes from Matthis Synofzik et al. [40]. According to this theory, the sense of agency has to be described by a two-step account. First level of this phenomenon is the "feeling of agency," it is pre-conceptual and pre-reflective, because of that, it operates on the very edge of consciousness. It may include the experience of intending an action, of choosing to perform this rather than other action, etc. These experiences are cognitive in nature and were linked to processes happening in primary motor cortex that is sending the motor command [36]. Second level is called the "judgment of agency," it reflects a person's judgement on being the author of an action. It hinges on motor information as well as post-hoc recreation of authorship [30]. This typically involves experiences that are associated with bodily movement and is relayed by peripheral somatosensory receptors. What is interesting, the involuntary movements tend to produce this kind of peripheral experience, but not this deeper experience of intent, because of that they are never accompanied by a sense of agency, although they are often accompanied by a sense of ownership.

Another issue is the distinction between the predictive and inferential aspect of the sense of agency. The question here is: what is more crucial for our sense of agency? The first option is, processes associated with action control and predicting possible sensory consequences of said actions – this is a predictive sense of agency [14]. The second option is, the interpretation of actions and experiences happening post factum – this is an inferential sense of agency [28], [47]. In this approach, the sense of agency does not preclude the action but is a consequence of it. Further in this work, I will assume that both aspects are equally necessary to understand the sense of agency, and none alone is enough to fully comprehend this phenomenon.

In philosophical literature, we can find propositions of several components of the sense of agency. We can start by asking if this phenomenon exist jointly with some other? We often find description of this experience as either an experience of being the source of decision or locus of control. This analysis would suggest that acting and controlling an action are intrinsically connected. We can distinguish at least two interpretations for both acting and controlling.

In the case of the former, we have to answer the question – what is this source we are talking about? We can call forth two theories, one authored by Athony J. Marcel [29], other by Nicolas Georgieff and Marc Jeannerod [17]. The first one is based on, the mentioned earlier distinction between a sense of agency and a sense of ownership. He states that in both cases, the sensation we experience is linked to some sense of ownership. In the case of agency, what we experience is an ownership of action. The source in that case is the ownership of action. The second theory is based on the idea of a so-called "who" system. In this theory, we begin with a completely anonymous actions, afterwards, we accredit those actions to us or other people. Then this "who", identified as an agent, becomes the source of action. By accrediting the source of action to ourselves, we constitute our sense of agency [6].

The control of actions can be similarly connected with a sense of agency by some mediating phenomena. It is possible that it is because we experience ourselves as controllers of actions, we have a category of agency. In this situation, the sense of agency can be linked to two different phenomena. In the first place, we can talk about a sense of control over our own body and its movement. It can be connected to control over sensory-motor signals, in that case, the experience of our body and thoughts as being controlled by us would be paramount for the sense of agency [27]. Other possibility is the sense of control over what is not our thoughts, that is we notice a control over aspects of external (physical or social) world. Good example is the experience of control over some machinery like driving a car. This feeling can function on a very primitive level, often pre-reflexive, but is fundamental for our experience of ourselves.

4. Investigating the Sense of Agency

Multiple approaches to studying the sense of agency exist. After James Moore's [30] distinction, we can divide them into two groups: either they use an implicit or explicit method of assessment. Bellow I will briefly describe both of those measurements.

Implicit measurement searches for behaviours or neuropsychological correlates of voluntary actions that can be assessed [30]. In this paradigm, the participants are not explicitly asked about their own experience of agency, instead how their experience looked like is inferred from some measured correlates. These correlates are treated like markers of the sense of agency. Usually, the implicit sense of agency measurement is based on the feeling of agency aspect of the phenomenon. The most widely used implicit sense of agency measurement appears to be the intentional binding [32]. The intentional binding effect is a subjective compression of perceived time between a voluntary action (e.g. voluntary pressing a button) and its external sensory effect (e.g. some king of audio cue). A common result is that the time interval between the action and the effect is underestimated when this action was voluntary, but not when it is involuntary [19] or passively conducted [49]. These findings led Moore and Sukhvinder Obhi [32] to suggest that temporal binding results from an efferent-based prediction system that binds an intent of action with the predicted sensory outcome. With a rise in popularity, this view was challenged by some authors. One objection was that some researchers could not find a difference between self-generated and involuntary actions [35]. Moreover, some studies found temporal binding in a situation of absence of volition [3]. As a result, some authors [3] suggested that a casual inference, rather than an intentional one, leads to temporal binding.

Explicit measurement, in contrast to an implicit one, assesses aspects of the sense of agency directly [30]. To achieve this goal, questionnaires, where participants judge their contribution to a task or describe how intense the experience of agency was during the task, are used. Popular versions of the explicit sense of agency measurements are the "helping hands" experiment [48] and the "I spy" experiment [47]. Both of those experiments will be described below. Another way of explicitly measuring the sense of agency are experiment where participants are asked to perform a motor task which they cannot observe [30]. They are offered some feedback on a screen, but often the movement depicted is not their own. Instead, it is movement of an experimenter or a computer simulation. Basing on that information, the participants are asked to judge whose movement can be seen on the screen.

5. Models of the Emergence of the Sense of Agency

They are multiple models of how the sense of agency appears. In this paragraph, I will attempt to describe the most popular in literature. They will be presented in an order of understanding. That means that the theory that is built upon an earlier one will be presented later.

The first theory I will describe is the comparator model. First fashioned as a theory of motor control, it is used today by authors like Chris Frith [13] and Nicole David [5] to explain the sense of agency. This theory states that the brain has an internal prediction model, it includes an efference copy whenever a new motor command is produced. If this copy matches the sensory input, the movement is perceived as self-caused and a sense of agency is produced. In an opposite situation, efferent does not match reafferent, the sense of agency will not appear. The comparator model as s a model of motor control is well supported by empirical data [5], [40]. Unfortunately, the relation between this model of motor control and mechanisms of how the sense of agency appears is not as clear [30]. One objection is that this model considers only sensorimotor cues neglecting any other that can possibly be relevant for the sense of agency [30], [40], [47]. Another critique is that there exists relevant clinical and experimental evidence of a sense of agency appearing in the absence of reafference, and without it, the comparator mechanism cannot be fulfilled. An example of clinical data, contradicting the comparator model, is the observation of phantom limb patients experiencing voluntary movement in their phantom limb [38]. An example of experimental data, contradicting the comparator model, is Daniel Wegner's "helping hand"-study [48]. In this study, the participants watched themselves in a mirror while another person stands behind them extending and moving his or her arms in such a way that in the mirror, the impression of the participant moving his or her arms is generated. It appears that if in this situation, the participants are verbally informed about the next action; they report a sense of agency arising for said movements [48].

The second theory I will consider is the theory of apparent mental causation [47]. This theory approaches the problem of the emergence of the sense of agency by rejecting a strong involvement of motor systems postulated by the comparator model. Instead, it proposes the sense of agency to be an effect of a purpose inference mechanism, that infers the casual relation for the observed action from the sensory input [32]. The proposed conditions for appearance of a sense of agency are: 1) an intention precedes an observed action; 2) the intention is compatible with this action; 3) the intention is the most likely the cause of this action [32], [47]. Empirical support for this theory comes from, the mentioned earlier "I spy"-experiment [47]. In this experiment, the participants work in a cooperation with the experimenter ally in jointly controlling a computer mouse cursor that can be moved onto a set of pictures displayed on the screen. Their task is to point to one of the pictures and then hold the cursor over this picture for around half a minute. After the task is performed, the participant indicates how big of an impact he or she had, in his or her subjective opinion, on completing the task. An interesting observation was that when the participant is primed with a chosen picture before the trial, he or she tends to attribute more of an impact to his or her actions. This situation is true even if the picture he or she was primed with, was chosen by the experimenter ally and not by him or her. This overestimation of self-agency led Wegner to postulate that the sense of agency is illusionary. He states that conscious willing of an action is not casually involved in performing said action [46].

The next theoretical position, in respect to emergence of sense of agency, is called the multifactorial weighting model. It is an attempt to reconcile the two previous theories. It is achieved by suggesting that the sense of agency is generated based on many different cues, which are weighted according to their reliability in a certain situation. In that way, this theory does not deny the comparator model involvement in creating a sense of agency, but it also allows other processes to play their part in the generation of this experience. Other cues are taken into consideration if, for example, an action does not allow for clear efferent-reafferent comparison. Going back to the feeling of agency and judgment of agency distinction, mentioned earlier in this work, it tends to happen more for the judgment of agency situations. That is the case because for the judgment of agency, social and environmental data provide more reliable indications then the efferent-reafferent comparison. Synofzik [40] provides an example of siting alone in a room when an action happens. He states that we may be ready to ascribe this action to ourselves simply on the basis of believing that we were alone in this room.

Even if the multifactorial weighting model is correct, there still is a question of how the brain assigns the weights to different agency cues. The Bayesian cue integration theory [31] tries to answer this question, and it is the last model of the appearance of a sense of agency I will describe. The background idea behind this theory is the assumption that the brain has access to many different information channels, each giving their own estimation about origins of the action. Those estimations are marked by a high uncertainty, because of that the brain cannot simply rely only on one cue but has to effectively combine all the information coming from different channels. To achieve that, as Moore and Fletchers suggests, the brain creates an estimate out of all agency cues, where importance of each cue is weighted according to every cue precision. The authors' suggestion is that the brain applies a maximum likelihood estimation to all agency cues thus giving an overall agency assessment. This assessment likelihood is much higher than assessment based on any single cue alone [31]. There is significant experimental evidence that the nervous system often integrates multisensory inputs in a maximum likelihood estimation manner [44]. Interestingly, this approach does not require any priori knowledge about which agency assessment is to be expected. However, such a priori knowledge can be added to the model as Bayesian priors [31]. We can notice three important advantages of the mentioned theory. First of all, it provides an effective model of how many agency cues can be integrated in one agency inference mechanism. Secondly, it can explain how the integration of agency cues coming from different modalities is possible. Thirdly, it can integrate the priori knowledge and beliefs into this inference mechanism. An unfortunate aspect of this model is that it cannot answer the question about how many possible cues there are [4].

6. Agency and Responsibility

Haggard adheres to idea that personal responsibility for actions is based in freedom of said actions, and this freedom is judged by the sense of agency. He summarises his views on responsibility in the following way:

> This (personal responsibility) forms the basis for praise and blame, punishment and reward. Individual responsibility depends on the assumption that most, or all, individuals experience a sense of agency over their actions and outcomes. In fact, courtroom pleas of 'guilty' or 'not guilty' are explicit judgements of agency. Few mental states thus sustain such a strong social superstructure as the sense of agency. The 'voluntary act condition' in law insists that an individual can only be criminally responsible for actions that they consciously decided to perform with a reasonable understanding of the likely outcome [18, p. 205].

Examining the problem of responsibility, Polish philosopher Roman Ingarden wrote: :Perpetrator is responsible for an act performed by himself, and its outcomes, if and only if it is his own act" [21, pp. 82-83].

The author follows with observation that, in the first place, we have to answer the question: What does it mean that an act is an own act of someone [21]? He concludes that there are two conditions: 1) the agent has to be conscious and understand his or her actions; 2) the agent has to be able to choose to act. We will not follow the first condition, but we will analyse the second one. Ingarden noticed that the second condition is directly linked to the controversy of determinism-indeterminism. It is like that because, as he states after Nicolai Hartman, free will decision is usually understood as causeless. Often, it is believed that free will cannot be reconciled with the pervasive determinism prevailing in the world. However, after Hartman, he concludes that the lack of cause cannot be a criterion for free action. Causeless action would not be motivated, ergo could not be an action the agent consciously decided to perform. He proposes that free action must mean an action that the cause of has a source only in the agent. That situation happens in two instances: 1) the agent accepts what is necessary, because he or she understands the inevitability of it; 2) the decision comes directly from within the agent without any external impetus. It is very well possible that, in the deterministic material world (and that is the world presented in “hard sciences”), the second criterion cannot be fulfilled, but the first one remains a possibility.

There remains the question of the possibility of free choice in a situation of a lack of alternatives. Can we reasonably assume that the source of action was within us in a situation when we did not have the freedom to do otherwise? The most prominent strategy for defending possibility of this situation comes from Harry Frankfurt [12]. He presented a series of thought experiments intended to show that it is possible for agents to be morally responsible for their actions and yet lack the ability to do otherwise.

Let us consider a Frankfurt-style argument presented by John M. Fischer:

> Imagine, if you will, that Black is a quite nifty (and even generally nice) neurosurgeon. But in performing an operation on Jones to remove a brain tumor, Black inserts a mechanism into Jones’s brain which enables Black to monitor and control Jones’s activities. Jones, meanwhile, knows nothing of this. Black exercises this control through a sophisticated computer which he has programmed so that, among other things, it monitors Jones’s voting behavior. If Jones were to show any inclination to vote for Bush, then the computer, through the mechanism in Jones’s brain, intervenes to ensure that he actually decides to vote for Clinton and does so vote. But if Jones decides on his own to vote for Clinton, the computer does nothing but continue to monitor – without affecting – the goings-on in Jones’s head [11, p. 38].

Fischer goes on to argue that a personal responsibility is not based on the possibility to choose otherwise. If Jones chooses Clinton on his own, Fischer argues, it is his own free action – even if other possibility was never attainable. What matters for the agent’s freedom and moral responsibility is not what might have happened, but how his or her action was actually brought about. Unfortunately, the sense of agency is unable to answer this question. Research on this phenomenon concentrates on how a person decides what the source of the action is. It is not designed to answer how the action was brought about. Because of that, it cannot be used as an actual marker of free will. We can see that in the descriptions of the experimental measurements of the sense of agency. Even the most sophisticated of them, The Bayesian cue integration theory, only answers on what basis we believe that someone was an agent.

7. Conclusions

Research into the sense of agency has an undeniable significance. Moore mentions multiple areas of investigation that can benefit from examining this phenomenon [30]. The mentioned spheres are health and well-being (e.g. research into schizophrenia), human-computer-interaction, issues of free will and responsibility. As much as an importance of this research cannot be denied for first two areas of investigation, Moore himself diagnoses the problem of the research into the third area. He writes:

> Free will is the elephant in the room when it comes to sense of agency research. Researchers tend to sidestep the issue of free will and instead focus solely on uncovering things like the neurocognitive basis of agentic experience. That is, whether or not we have free will, we unquestionably *do* have the experience of agency when we make actions and scientific research has tended to focus on understanding this experience. This evasion of the free will debate is understandable; philosophical debates on free will are often quite complex and confusing, especially for scientists with no background in philosophy. However, I think those of us working on this topic should try to engage more with this debate. In terms of impact, the social and legal consequences of this debate are immense, and our findings should be helping to inform this debate [30].

In this work, the relation between a sense of agency and free will was examined. It is often believed that investigating the former can allow us to find a way of judging when an action is voluntary. An example of a researcher subscribing to this idea is, among others, Haggard. I started by reconstructing why some researchers believe free will requires a sense of agency. Next a description of this phenomenon was provided. Then I described the methodology behind investigating the sense of agency, to follow that with a presentation of the most popular models of emergence of this phenomenon. Finally, I analysed the relation between the responsibility and agency. In conclusion, the sense of agency, in my opinion, fails to fulfil hopes placed in it. It only answers the question of how we ascribe responsibility and not who actually is responsible. After all, as Ingarden noted [21], being held accountable is not the same as actually being accountable.

References

1. Baumeister, R. F. Free Will in Scientific Psychology, *Perspectives on Psychological Science* 3 (1), 2008, pp. 14-19.
2. Blanke, O., Metzinger, T. Full-body Illusions and Minimal Phenomenal Selfhood, *Trends in Cognitive Science* 13, 2009, pp. 7-13.
3. Buehner, M. J., Humphreys, G. R. Causal binding of actions to their effects, *Psychological Science* 20, 2009, pp. 1221-1228.
4. Carruthers, G. The case for the comparator model as an explanation of the sense of agency and its breakdowns. *Consciousness and Cognition* 21, 2012, pp. 30-45.
5. David, N., Newen, A., Vogeley, K. The “sense of agency” and its underlying cognitive and neural mechanisms, *Consciousness and Cognition* 17, 2008, pp. 523-534.
6. De Vignemont, F., Fourneret, P. The sense of agency: a philosophical and empirical review of the “Who” system, *Consciousness and Cognition* 13, 2004, pp. 1-19.
7. Della S., Marchetti, C. Anarchic Hand, *Higher-order Motor Disorders: From Neuroanatomy and Neurobiology to Clinical Neurology,* 2005, pp. 291-301.
8. Dennett, D. C. The Self as a Responding – and Responsible – Artefact, *Annals of the New York Academy of Sciences* 1001 (1), 2003, pp. 39-50.
9. Feldman, G., Baumeister, R. F., Wong, K. F. E. Free Will is About Choosing: The Link Between Choice and the Belief in Free Will, *Journal of Experimental Social Psychology* 55, 2014, pp. 239-245.
10. Firth, Ch. *The Cognitive Neuropsychology of Schizophrenia,* Hove: Psychology Press, 1995.
11. Fischer, J. M. *My Way: Essay on Moral Responsibility,* New York: Oxford University Press, 2006.
12. Frankfurt, H. Alternate Possibilities and Moral Responsibility, *Journal of Philosophy* 66, 1969, pp. 829-39.
13. Frith, C. D. The self in action: lessons from delusions of control, *Consciousness and Cognition* 14, 2005, pp. 752-770.
14. Frith, C. D., Blakemore, S. J., Wolpert, D. M. Abnormalities in the awareness and control of action, *Philosophical Transactions of the Royal Society B* 355, 2000, pp. 1771-1788.
15. Gallagher, S. Multiple aspects in the sense of agency, *New Ideas in Psychology* 30, 2012, pp. 15-31.
16. Gallagher, S. Philosophical Conceptions of the Self: Implications for Cognitive Science, *Trends in Cognitive Sciences* 4, 2000, pp. 14-21.
17. Georgiess, N., Jeannerod, M. Beyond Consciousness of External Reality: A “Who” System for Consciousness of action and self-consciousness, *Consciousness and Cognition* 7, 1998, pp. 465-477.
18. Haggard, P. Sense of Agency in the Human Brain, *Natural Revives Neuroscience* 18, 2017, pp. 196-207.
19. Haggard, P., Clark, S. Intentional action: conscious experience and neural prediction, *Consciousness and Cognition* 12, 2003, pp. 695-707.

20. Hume, D. *An enquiry concerning human understanding*, Oxford/New York: Oxford University Press, 1999.
21. Ingarden, R. *Książeczka o człowieku,* Kraków: Wydawnictwo literackie, 1987.
22. Jeannerod, M. The sense of agency and its disturbances in schizophrenia: a reappraisal, *Experimental Brain Research* 192, 2009, pp. 527-532.
23. Kane, R. *Free Will: New Directions for an Ancient Problem,* Malden: Blackwell Publishers, 2002.
24. Kane, R. *The Oxford Handbook of Free Will,* New York: Oxford University Press, 2011.
25. Kant, I. *Critique of practical reason,* 1797/1967.
26. Laurene, K. R., Rakos, R. F., Tisak, M. S., Robichaud, A. L., Horvath, M. Perception of free will: the perspective of incarcerated adolescent and adult offenders, *Review of Philosophy and Psychology* 2 (4), 2011, pp. 723-740.
27. Legrand, D. Naturalizing the Acting Self: Subjective vs. Anonymous Agency, *Philosophical Psychology* 20 (4), 2007, pp. 457-478.
28. Libet, B., Gleason, C. A., Wright, E. W., Pearl, D. K. Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential), The unconscious initiation of a freely voluntary act, *Brain* 106 (Pt3), 1983, pp. 623-642.
29. Marcel, A. J. The Sense of Agency: Awareness and Ownership of Actions and Intentions, *Agency and Self Awareness*, New York: Oxford University Press, 2003.
30. Moore, J. W. What is the sense of agency and why does it matter, *Frontiers in Psychology* 7 (1272), 2016, doi: 10.3389/fpsyg.2016.01272.
31. Moore, J. W., Fletcher, P. C. Sense of Agency in Health and Disease: A Review of Cue Integration Approaches, *Consciousness and Cognition* 21, 2012, pp. 59-68.
32. Moore, J. W., Obhi, S. S. Intentional binding and the sense of agency: a review, *Consciousness and Cognition* 21, 2012, pp. 546-561.
33. Nadelhoffer, T., Shepard, J., Nahmias, E., Sripada, C., Ross, L. T. The Free Will Inventory: Measuring Beliefs About Agency and Responsibility, *Consciousness and Cognition* 25, 2014, pp. 27-41.
34. Nowakowski, P., Komedzinski, T. Poczucie sprawstwa: ujęcie interdyscyplinarne, In M. Pąchalska, G. E. Kwiatkowska (eds.), *Neuropsychologia a humanistyka,* Lublin: Wydawnictwo UMCS, 2010, pp. 251-261.
35. Passingham, R. E., Wise, S. P. *The neurobiology of the prefrontal cortex: anatomy, evolution, and the origin of insight*, New York: Oxford University Press, 2014.
36. Poonian, S. K., Cunnington, R. Intentional binding in self-made and observed actions, *Experimental Brain Research* 229, 2013, pp. 419-427.
37. Rakos, R. F., Steyer, K. R., Skala, S., Slane, S. Belief in free will: Measurement and conceptualization innovations, *Behavior and Social Issues* 17 (1), 2008, pp. 20-39.
38. Ramachandran, V. S., Hirstein, W. The perception of phantom limbs. The D. O. Hebb lecture, *Brain* 121 (Pt 9), 1998, pp. 1603-1630.
39. Sarkissian, H., Chatterjee, A., De Brigard, F., Knobe, J., Nichols, S., Sirker, S. Is Belief in Free Will a Cultural Universal? *Mind & Language* 25, 2010, pp. 346-358.
40. Synofzik, M., Vosgerau, G., Newen, A. Beyond the comparator model: a multifactorial two-step account of agency, *Consciousness and Cognition* 17, 2008, pp. 219-239.
41. Tsakiris, M., Fotopoulou, A. *Decomposing the Will,* New York: Oxford University Press, 2013.
42. Tsakiris, M., Haggard, P. Experimenting With the Acting Self, *Cognitive Neuropsychology* 22, 2005, pp. 387-407.
43. Tsakiris, M., Haggard, P. The Rubber Hand Illusion Revisited: Visuotactile integration and Self-attribution, *Journal of Experimental Psychology: Human Perception and Performance* 31, 2005, pp. 80-91.
44. van Dam, L. C. J., Parise, C. V., Ernst, M. O. Modeling multisensory integration, In D. J. Bennett, C. S. Hill (eds.), *Sensory Integration and the Unity of Consciousness*, Cambridge, MA: MIT Press, 2014, pp. 209-229.

45. Vohs, K. D., Schooler, J. W. The value of believing in free will encouraging a belief in determinism increases cheating, *Psychological Science* 19 (1), 2008, pp. 49-54.
46. Wegner, D. M. *The Illusion of Conscious Will*, Cambridge, MA: MIT Press, 2002.
47. Wegner, D. M., Wheatly, T. Apparent mental causation: sources of the experience of will, *American Psychologist* 54, 1999, pp. 480-492.
48. Wegner, D. M., Sparrow, B., Winerman, L. Vicarious agency: experiencing control over the movements of others, *Journal of Personality and Social Psychology* 86, 2004, pp. 838-848.
49. Wohlschläger, A., Engbert, K., Haggard, P. Intentionality as a constituting condition for the own self – and other selves, *Consciousness and Cognition* 12, 2003, pp. 708-716.

Studia Humana
Volume 8:2 (2019), pp. 79—83
DOI: 10.2478/sh-2019-0020

On Computers and Men

Tomasz Goban-Klas

University of Information Technology
and Management in Rzeszow,
Sucharskiego 2 Street,
35-225 Rzeszow, Poland

e-mail: tgoban@wsiz.rzeszow.pl

Abstract:
The title of the article was inspired by the novel by John Steinbeck "Of Mice and Men" (1937) and the poem by Robet Burns about the deception of human plans. Even the best of them often lead astray, or their far-reaching negative effects are revealed. As it seems, nowadays nature ("mice") and men (people) are in a breakthrough period – in the geological sense between the old and the new era, the Holocene and the Anthropocene, in the cultural sense – between the analogue and digital era that can be – and it should actually be called a digit. Levi-Strauss in his essay "Raw and cooked" points to the groundbreaking for the emergence of human culture the use of fire in the preparation of food, and therefore the transition from nature to culture, and its foundation – the kitchen [12]. At present, this new phase of transition can be seen in the digitization of interpersonal communication and its current correlation – cross-linking. It was announced by the famous Turing machine (1936), a computer design and layout, which was realized in the 1940s and 1950s, and enter in mass production at its end, networked on a global scale in the 1990s and make mobile in the second decade of the 21st century in the form of a smartphone
Keywords: Turing machine, computer, anthropocene, digital plenitude.

***To a Mouse**,*
But Mouse, you are not alone,
In proving foresight may be vain:
The best laid schemes of mice and men
Go often askew,
And leave us nothing but grief and pain,
For promised joy!
Robert Burns (1786)

The title of the article was inspired by the novel by John Steinbeck “Of Mice and Men” (1937) and the poem by Robert Burns about the deception of human plans. Even the best of them often lead astray, or their far-reaching negative effects are revealed. As it seems, nowadays nature (“mice”) and men (people) are in a breakthrough period – in the geological sense between the old and the new era, the Holocene and the Anthropocene, in the cultural sense – between the analogue and digital era that can be – and it should actually be called a digit. Levi-Strauss in his essay “Raw and cooked” points to the groundbreaking for the emergence of human culture the use of fire in the preparation of food, and therefore the transition from nature to culture, and its foundation – the kitchen [12]. At present, this new phase of transition can be seen in the digitization of interpersonal communication and its current correlation – cross-linking. It was announced by the famous Turing machine (1936), a computer design and layout, which was realized in the 1940s and 1950s, and enter in mass production at its end, networked on a global scale in the 1990s and make mobile in the second decade of the 21st century in the form of a smartphone.

These achievements of human thought – technical and logical – introduced into common human practice change the traditional – seemingly old practices of the analogue era bring them into digital and network forms. More and more extensiveness and strengthening of human capabilities through the media, and new digital tools, including the so-called artificial intelligence, AI. The realization of a vision about which some philosophers and theologians dreamed of, like Theiard de Chardin, is beginning to approximate. Its concise reminder and indication of the basis for its implementation is contained in this article.

1. Noosphere – a New Sphere of Our World

For over a dozen years, seeing the enormous impact of human activity on the global ecosystem, some of them have recognized that humanity has entered a new geological epoch – the Anthropocene. The term was proposed by Paul Crutzen, Nobel laureate. The authors of the “Anthropocene Review” argue that the beginning of the Anthropocene era should be considered half of the twentieth century.

In the Anthropocene – if we accept this distinction of this new geological-biological era – the material and energetic scale of human activity plays a key role, but after all directed – as *homo socialis et comunicans* – by human culture and through inflows and flows of information into the social system and human minds. Therefore, its component is not only the geosphere, biosphere and technosphere (material and energetic artificial basis of functioning of societies), but also the sphere of culture, called the noosphere for centuries.

The Russian scholar Vladimir I. Vernadsky [9], author of an important work “Biosphera” (1926), believed that in the development of the Earth, just as the appearance of life fundamentally changed the geosphere, so – according to Vernadsky – the emergence of people endowed with cognitive abilities will completely transform the biosphere (these views do not were widely accepted in the West) [10]. He completed this developmental line indicating that after the geosphere (inanimate matter) and biosphere (biological life), the noosphere as the “cloak of the mental Earth,” the set of all information and their media on the planet, is its third component

It becomes obvious that the biggest change in human life – individual and social – is in the sphere of ways and means of communication, and – after Jay D. Bolter [10] – it can be expressed in two words: *ubiquity* and *diverstity*. The first one emphasizes the widespread (global) availability of computers of all kinds, including multimedia mobile phones, the second that mediamorphosis maintains and even increases the diversity of media devices, not reducing them to one universal transmission tool.

It develops – using the term of the pope Benedict XVI – *continento digitale*, a digital continent, based on a network, also wireless, and therefore ubiquitous or all-extending [11]. Thus, the noosphere has a material basis – the media apparatus – the space flow of Manuel Castells is real and palpable. Information flows through it, created not only by people, but also by apparatus, processed not so much by human minds but by algorithms. It becomes a key element of management, management, all activity of machines and people. Castells used the term "informationalism," it can be included in the concept of media civilization and define and analyze it as an information and media civilization [3].

After the Second World War, the computer became – as Bolter pointed out – the so-called technology that defines modernity, both realistically and metaphorically ("the Computer Age") introducing humanity into the information society. In his book "Turing's man. Western culture in the computer era" [2], he describes the internal operation and structure of the computer (time, space, language and program) in contrast to old technologies, optically and physically simpler (reel, potter's wheel, clock, steam engine), which today shape the mind of the user and society in its image and likeness. This is an essentially optimistic analysis – "Turing's man" is an expert in their machine and its limitations, wisely and ethically using it.

If you look for a ground-breaking intellectual announcement of the anti era it could be found in the article by Alan Turing from 1936, proposing a scheme of operation on the symbols later called the "Turing machine." The article "On Computable Numbers" or "About computable numbers" described an abstract machine that was able to perform a programmed mathematical operation, i.e. algorithm. In 12 years later, in 1948, in the paper "A Mathematical Theory of Communication," Claude Shannon announced the invention of a transistor, the basis of computerization and digitization [13].

Bolter develops an interesting comparison of the main – dominant – technologies and their metaphors from antiquity to modern times. For the ancient Greeks, according to Bolter, the dominating technological metaphor was a drip spindle, a device for twisting yarn in a thread. Such a metaphor implied technology as a controlled application of power. In Western Europe, after the Middle Ages, the analog to the spindle was first a clock with a load, the triumph of mechanical technology, and then a steam engine, the climax of the dynamics of thermal energy. In the subtly developed observation of Bolter, a computer - as a metaphor defining the present age – is a machine that connects the conceptual ideas of both the clock and the steam engine. However, paradoxically, the computer also represents a return to antiquity in the sense of a certain image of the manual world.

In several well-thought-out chapters on how the computer redefines our concepts of space, time, memory, logic, language and creativity, Bolter makes a comparison in which the computer simultaneously introduces a new wonderful Western technology and turns us back to the idea of ancient Greece. He states that "if the ancient ideals were balanced, proportional and craft (using the spindle), and Western European was Faust's pursuit of power through knowledge (understanding the mechanical universe to achieve the dynamics of the steam engine)," Turing man "combines both ideals" [2, p. 323].

"In a way, a computer man keeps and even extends the Faustian tendency to analyze," concludes Bolter. "But remember, he adds, that the purpose of Faust's analysis was to understand, and this" in-depth, problem, while Turing's man is oriented not so much at understanding, at acting" [2, p.334].

"For a Turing man, knowledge is a process, a skill," just like ancient pottery art. "A man or computer only knows something if it can get the right answer to the right question." Speaking more informative language when an algorithm is prepared. "Faustian depth" adds nothing to the operational success of the program.

Thus portraying the “Turing man,” Bolter seems to refer to the use of a few simple metaphors. However, he develops his arguments with unusual concreteness. If there is any weakness in them, it is included in the range in which he presented a too repetitive and ultimately predictable pattern of computer operation.

Bolter claims that “the computer is the latest and most radical defining technology because it has become the dominant metaphor of the human mind in popular culture as well as in more technical fields such as psychology and neuroscience. This metaphor is essentially a Turing man.” Bolter claims that Alan Turing was right when he predicted that computers would be able to imitate human intelligence perfectly, but “because the machine thinks like a human being, man recreates himself, describes himself as a machine ... as information processor and nature as information for processing.” Trying to build artificial intelligence, we have transformed into artificially intelligent creatures, that explains Bolter’s position.

However, much more important in the “Turing’s man” is to fill the gap between the exact sciences and the humanities. After reading Bolter’s book, the reader finds that the computer is much less mysterious than he thought. It is not a coincidence that the book allows us to understand why computers are not so perfect in mathematics (for example, they cannot use the concept of infinity); but they are helpful in explaining the “Turing test” for assessing artificial intelligence.

The most provocative in the analysis is what Bolter has to say about the political consequences of computer age. Will Turing’s man prove the power of George Orwell’s “Big Brother” instruments as so many observers are afraid of widespread surveillance? It’s very possible that he did not, said Bolter in 1984: “... computer age cannot really produce people who are capable of great good or evil.” Turing’s man is not a possessed soul, as often as a Faustian is, he does not treat himself and his world so deadly seriously, he does not talk about “destiny,” but if the computer age does not produce Michelangelo and Goethe, it is probably less likely to produce Hitler or even Napoleon. Totalitarian leaders were people capable of concentrating the Faustian commitment of the will of citizens to their goals. And what if they lack their strong will? Orwell’s “1984” assumption was to combine a totalitarian goal with modern technology. “But the most modern technology, computer technology, may be incompatible with a totalitarian monster, at least in its classic form,” probably Bolter wrote too optimistically .

2. Computer in the 21st Century – the Same (Turing's Machine), But Not the Same!

After all, 35 years have passed since the writing of “Turing’s Man,” which was very inspiring and still today, and although the logical diagram of the computer has not changed at that time, the computer itself has changed technologically (and programmatically). It is the same, but not the same. And new technology has given him new possibilities (affordance) and creates a new, wonderful and less-than-perfect digital real-life (according to Manuel Castells [3]).

This is perfectly understood by professor Bolter. He in 2019 published a monograph on Western culture in the computer era, titled “Digital Plenitude,” or digital abundance, even excess. excess [1]. Its key idea is to say that our media culture is full of excess. This is the world of products (websites, video games, blogs, books, movies, TV and radio programs, magazines, etc.) and practices (creating all these products together with their remixing, sharing and commenting), is therefore vast, diverse and dynamic which is not understood and understood as a whole. “Excess easily adapts, even absorbs, contradicts the forces of high and popular culture, old and new media, conservative and radical social views. Digital media are an ideal environment for this fullness - for our flattened media culture, in which there are many central points, but there is no single center.”

Yuval Noah Harari in the book “21 lessons for the 21st century” looks even further: “We live in an age when people are hacking. Algorithms are looking at you at this moment. / ... / Based on big data and machine learning, they will get to know you better. And when these algorithms will know you better than you, then they will be able to control you and manipulate you, and you will not be able to do anything about it" [5, p. 342].

Pedro Domingos in his book "The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World" is describing as algorithmic machine learning is remaking business, entertainment, politics, science and military. And he gives a description of the quest to find 'The Master Algorithm' – a universal computer-based learner capable of deriving all knowledge from data. It would be a radical, or better say, total transformation of way as human knowledge is transformed into data, and then, into human life. This vision is not fully optimistic, it contains the seed of terrifying future - full control of humans by machines. In October 2015, a software called "AlphaGo®" became the first computer to beat a professional human Go player in game of Go, more difficult them the chess. It is a clear sign that the Artificial Intelligence mature and is going to master other than games fields of human inventness.

A century and a half ago Karl Marx and Frederic Engels published the "Communist Manifesto" announcing that "A Spectre is haunting Europe – the spectre of Communism." Today, we read more and more philosophical manifestos are warning about perils of artificial intelligence. By 2014, the famous scientist and philosopher Stephen Hawking and business magnate Elon Musk had publicly voiced the opinion that superhuman artificial intelligence could provide incalculable benefits, but also can end the human race if deployed incautiously. One may say: "A new spectre is haunting the world – the Spectre of the Universal Master Algorithm," or – expressing more cautiously – the Spectre of the *Algocracy*. From masters, humans may become slaves. And that it would be an end of Turing's Man.

References

1. Bolter, J. D. *The Digital Plenitude: The Decline of Elite Culture and the Rise of New Media*, Cambridge: The MIT Press, 2019.

2. Bolter, J. D. *Człowiek Turinga. Kultura Zachodu w erze komputera*, trans. by T. Goban-Klas, Warszawa: PiW, 1991.

3. Castells, M. *Społeczeństwo sieci*, trans. by M. Marody et al., Warszawa: PWN, 2008.

4. Domingos, P. *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World*, New York: Basic Books, 2015.

5. Harrari, Y. N. *21 Lessons for the 21st Century*, New York: Spiegel & Grau, 2018.

6. Bolter, J. D. Examining and Changing a World of Media, http://leading-edge.iac.gatech.edu/humanistic-perspectives/examining-and-changing-a-world-of-media./

7. http://movies2.nytimes.com/books/99/01/03/specials/bolter-turing.html.

8._Vernadsky, V. https://www.geochemsoc.org/files/4813/4436/8118/gn125.pdf**.**

9. Wiernadski, W. https://pl.wikipedia.org/wiki/W%C5%82adimir_Wiernadski.

10. Bolter, J. D. https://www.digitalplenitude.net/.

11. Laskowska, M., Marcyñski, K. *Komunikacja społeczna według Benedykta XVI*, Kraków: Petrus, 2016.

12. Lévi-Strauss, C. *Surowe i gotowane,* trans. by M. Falski, Warszawa: Wydawnictwo Aletheia, 2010

13. Shannon, C. A Mathematical Theory of Communication, *The Bell System Technical Journal* 27, 1948, pp. 379-423, 623-656.