STRANGE REFLEXIVE PROPERTIES AND AWARENESS OF SELF AWARENESS AS ESSENTIAL ATTRIBUTE OF STRONG AI

Dmitriy V. Vinnik
Financial University under the Government of the Russian Federation
Moscow, Russian Federation
dvinstor@gmail.com

Abstract

Awareness of self-awareness (ASA) construed as real high-order reflective type of mental states. This state is rarely the focus of research attention. It is claimed that awareness of self-awareness is not contingent or bizarre state of mind but fundamental cognitive property despite origin of certain intellect, – natural or artificial. The core of such views may be detected in ancient Greek and Indian philosophy, namely in doctrines of great logicians: Aristotle and Dignaga. It is strongly strange from logical point of view that concept of self-awareness is essential for explanation of concept of awareness (consciousness) because more complex notion normally should not be used to define simplest concept. Moreover, the concept of consciousness is often identified with notion of self-awareness ignoring the distinction between levels or ranks of reflection. According to Aristotle the nature of human mind and «first philosophy» is thinking about thinking. His thesis used to explain the autonomous and «external» nature of reasoning as highest part of the soul. Proposed that awareness of self-awareness can be naturalized as specific functional mental state and even the state of human brain in neurophysiological perspective. Raymond Smullyan insisted that awareness of self-awareness is sufficient property for formal intellectual subjects (*Thinkers*) which are able to be be aware about their own type of reflexive rank. Is concluded that if his proof is correct this means that 4th reflective rank may be treated as basic level or «zero» level of Reason. An attempt is made to naturalize Smullian's concepts of normality and stability of reasoning as special reflexive properties. The idea of a psychophysical experiment is proposed, which makes possible to establish the real distribution of the stability property among experts. This experiment may help to answer the question: "Does intuitive belief in the truthiness of statement depends on the real truth of the proposition being proved?" From the standpoint evolutional epistemology high-order mental states could be explained through as a result of social reflective games evolution. Animal have an ability to detect or grasp opponent's reflective range and to detect deception (as behavioral phenomena) or epistemic lie (as propositional phenomena). Probably this is possible due the evolution of empathy ability. Naturalizing of ASA is crucial for designing such type of strong AI as artificial personality.

Keywords: artificial intellect, Gödel theorem, mentalism, machine functionalism, mind, algoritmization, self-awareness, cognitive functions, neurophysiology, polygraph, deception detection

1. Introduction

The concept of artificial intelligence has not been strict since its inception in 1956 at a famous University of Dartmouth workshop [1]. In different fields of studies and computer science schools the meaning of this concept may vary significantly. This is also true for the philosophy of science, which attempts to conceptualize the real subject-matter of AI research that gave rise to this concept. The lack of agreement may be explained that the concept of intelligence itself is deeply polemical, its content depends on a variety of ontological assumptions and conceptual frameworks.

Philosophers commonly distinguish between the concept of strong (general, full, or strong) and weak or narrow artificial intelligence and dispute about the fundamental possibility of strong AI existence [2]. The concept of the strong AI means the possibility of creating an artificial personality with all full-fledged attributes: self-awareness and creative thinking, and, even — the Reason as such. Those who deny the possibility of creating or even spontaneous appearance (emergency) of intelligence in complex cybernetic systems are usually referred to supporters of weak AI. The latter allow only the reproduction of certain cognitive functions by technical utilities without penetrating the essence of humanity and rationality.

The debate about the possibility of full-fledged artificial personality creation was conducted with involvement of arguments from a wide variety of disciplines, from respectable evolutionary biology to the most vulgar sorts of metaphysics. Unfortunately, this dispute rarely goes beyond speculative argumentation. This view is shared by V.V. Tselishchev, who contrast it with the clearer concept of "algorithmic thinking": "First, algorithmic thinking is often linked to a very broad question, which is romantically formulated as the question: "Can a machine think?" In a such formulation, the problem acquires such great uncertainty and metaphoricity that it practically leaves the sphere of the rational. True, the well-known Turing test puts the problem within a certain framework, but even in

this case it acquires a largely psychological character" [3]. Further Tselishchev argues that a wide range of cognitive researches devoted to the creation of brain computer models. They use a number of theories and concepts from mathematics, physiology, theory of complexity, computer research, etc., which are "an almost immense field of action, largely associated with programs for building artificial intelligence systems" [3]. Moving away from such a discourse, Tselishchev raises the question of the necessity to study the fundamental possibility of algorithmic thinking, keeping the distance from practical attempts to study such algorithms. Indeed, the fundamental question of strong AI possibility was posed in the clearest and the most acute form in the philo sophy of logic and directly corresponded to the very foundations of mathematics.

Supporters of so-called mechanism and mentalism holds the opposite answers to this question. Mechanicists follow the reductionist view that the natural mind is equivalent to a computational machine. It means that a natural mind can be formalized. It follows that the various efforts to simulate cognitive functions are apparently meaningful and the creation of a strong AI is theoretically possible. It should be noted that among those who are somehow involved in this issue, mechanicists are the vast majority.

2. «Mechanism» as Computational Reductionism

One of the most famous philosophical forms of mechanism is functionalism the supporters of which include P. Churchland, J. Fodor and N. Blok. The list of most famous mechanists should obviously include early H. Putnam, who suggested the reductionist concept of machine functionalism. According to this concept, mental states are identical to the Turing machine calculation states [4].

H. Putnam insisted that the type of physical carrier involved in functional organization do not have fundamental importance for mental states realization. Therefore, the essense of any mind have entirely structural and logical nature. Computer that is capable to pass the Church test successfully may be considered as a thinking thing in regard to the conjunction of the mental functions, which we will take as a mind definition consequent. Indeed, historically functionalism owes its emergence to the first computers appearance and is commonly called the "computational theory of mind". Functionalism conception is based on two assumptions. The first assumption treated the mind-brain relation as a relation between computer program ("software") and a computer itself ("hardware"). In this case the program must be understood as a universal computational algorithm (the Turing machine). Moreover, any stuff that can be somehow used as a logical machine has to be construed as hardware.

The second assumption, known as the "multiple realization thesis," is an extent of the first but no less interesting. It declares that the computer program is indifferent to its physical implementation type. It means that for the computer program (in form of a universal algorithm) it does not matter on what kind of machine it could be executed. As the sender and the addressee of the transmitted information does not care about the type of information carrier (carved in stone and sent by a parcel post or sent by modulated radio signal), any universal algorithm can be performed on computers created on a different element base (gears, lamps, transistors, thermal valves, quantum switches) and realized in different architectures (serial, parallel). H. Putnam explicated the very idea in a peculiar ironic manner:

"*The question of the autonomy of our mental life … has nothing to do with that all too popular … question about matter or soul-stuff. We could be made of Swiss cheese, and it wouldn't matter*" [5].

Functionalism has interesting implications. In the first instance, calculating devices can be artificial and natural. For the successful computational algorithm work presented in the Turing machine universal form, the physical substrate of the device is of no significance.

Secondly, the same universal algorithm can successfully functionalize both on a computer and in the brain. Mind as a kind of certain algorithm implicitly can be transferred from the brain to a computer of comparable complexity and computational power. The reverse process is also possible.

Concerning functionalism standpoint, the question of a computing device element base is irrelevant. At least any natural object with sufficient number of logical gates and connections between such elements is suitable for computational purposes.

According to radical form of so-called "machine" functionalism, which was developed by H. Putnam (and subsequently abandoned by himself), the question of logical gates physical nature construed as pseudo-problematic. Mental functions are completely indifferent to the way they are embodied. From all appearances, the functionalism standpoint presupposes only one genuine mental function — a universal recursive algorithm or the Turing machine. It can be affirmed that this radical kind of functionalism is consistent with idealistic ontology. The general assumption that mental functions are indifferent to type of substrate (material or immaterial) and, hence cognitive abilities are autonomous is obviously an extremely radical idea.

In 1960, J. Lucas published his famous paper "Minds, Mechanisms and Gödel". He attacked H. Putnam's sort pf machine reductionism, known as machine functionalism [6]. Lucas construed functionalism as a kind of materialism, to which he felt a clear distrust. In modern philosophy of mind functionalism is considered the most progressive version of materialism. However, this is not entirely true. Functionalism was created as a theory of mind that has to go beyond the problems of substantial ontologies as H. Putnam initially conceived it. The success of functionalism may be explained by the fact that it appeared as new type of ontology based on high order logic of predicates. Namely it was an ontology of relations in contrast to classical substantial ontology of things or phenomenal ontology of properties.

3. Mentalism as Algorithmic Nature of Thinking Rejection

Mentalists are rejecting the equivalence principle of human and machine intelligence. They are insisting that human thinking cannot be represented by an algorithm. More than that the fundamental "superiority" of human intelligence over machine intelligence is asserting by them. This assertion should also be understood as the impossibility to automatize mathematician intuition. Only few famous scientists may be considered as mentalists: Oxford professor John Lucas and famous physicist Roger Penrose, who fell under the influence of the former. The core thinker of this field of study is Kurt Gödel, who inspired this polemic by his incompleteness theorem, should also be attributed to mentalist's camp.

Gödel came to the following conclusions in his lecture in honor of Gibbs: "1) The human mind is incapable to formulate (or mechanize) all mathematical intuitions; if they are formulated, new ones inevitably appear; for example, – about consistency. 2) Either the human mind is superior to all machines (more precisely, – it can solve more number-theoretical problems), or there are such number-theoretical problems that are undecided by the human mind. [7] We know that Gödel rejected the second term of the disjunction contained in the second thesis. It is important that Gödel interpreted mathematical intuition in the classical metaphysical style — as a direct access to platonical universum of mathematical objects [8].

The theorem caused an enormous amount of rude ontological and linguistical speculations. Moreover, it was included to "canonicity" of conceptions used to produce pseudoscience alongside quantum mechanics and chaos theory [9]. Though it really raise the fundamental question, which Tselishchev formulates as follows: "If we proceed from the superiority of man the main issue under study is the possibility of algorithmization of thinking. If we proceed from their equivalence of a machine to a man or even the superiority of a machine over a man the main issue is the algorithmic nature of human thinking" [10]. The author notes, that 'thinking' in this case has to be treated as 'mathematical thinking', as reasoning with mathematical definiteness. He insists that Gödel's incompleteness theorems "speak of fundamental constraints in attempting to formalize intuitive knowledge totally" [11].

4. Gödel's Argumentation and Its Heuristic Role

Mechanicists and mentalists discussion revolves around the evaluation of possibilities to construct so-called "Gödel sentence" (an undecidable sentence in a fairly rich formal system) for a machine and for a person. According to Lucas no computer is capable to construct such a sentence, but the human person is. Lucas believed that Gödel sentence is an Achilles' heel of the machine form of intelligence. His idea gave birth to popular belief that the incompleteness theorem is crucial for discussion "machine versus mind" [12].

According to Lucas human person is able to construct a true sentence that never could be printed by a computer. In other words, a Gödel proposal appears as the problem for the computer perfectibility. Lucas also believed that only the human mind is capable to generate ordinal numbers. That is to say that human ability to recognize ordinals superior an ability of any formal algorithm to perform this task.

Lucas's sophisticated argumentation had not been unnoticed by H. Putnam and P. Benacerraf. They paid attention to a number of ambiguities and deliberate dialectical reasons of Oxford professor. Lucas insisted that the human mind have an ability to "see" or "know" the truth of certain Gödel sentence. In this concern, R. Penrose takes into account the phenomenon of mathematical understanding, which cannot be completely reduced to computational methods and to a certain set of rules. According to R. Penrose, understanding must be construed as the function of our consciousness. Acceptance of these epistemological ideas gives us a good reason to believe that conscious perception is an "incalculable" process [13]. Henry Poincaré wrote about different types of empirical mathematical thinking: analytical and geometrical [14]. The thinkers who belong to the second type are more inclined to solve mathematical problems with great respect to their own understanding. The singular case of such mathematician, obviously, was indigenous Indian mathematician Srinivasa Ramanujan. It is well known fact that at the moment his arrival to London he did not understand what a formal proof is and what is the necessity of it.

Such intuitive epistemological modalities may be confusing, especially if they are supported by mystical allusions such as the idea that perception of mathematical objects and truths have transcendental nature.

In regard to Lukas opponents, the main sin of his argumentation consists in acceptance of self-consistency belief. This is the consistency of the person "who see" the truth of the Gödel sentence and also has the ability to distinguish between truth and falsehood. These opponents hesitate in significance of such premises not being supported by evidence of consistency. According to P. Benacerraf, a person may be just the Turing machine, but a contradictory kind of a machine [15].

According to D. McCullough, in order to use Gödel's theorem in obtaining an increasingly general theoretical form of thinking in its own ever-greater formalization, it makes a certain "leap" to the conclusion that this formalization is consistent itself. But if a mathematician formalized too much, including leaps, then the resulting theory should be able to formalize itself. It means that he would jump to the conclusion that his own Gödel sentence is inevitably true. As we know exactly this inference leads to the contradiction. Thus, McCulloch argues that all mathematical thinking faces a fundamental choice: "*So, either (1) a mathematician at some point cannot formalize all his reasoning (in this case, the totality of all the facts that he can prove will be an axiomatizable theory), or (2) he formalizes all his reasoning, and the resulting theory is inconsistent (it will be able to prove its consistency)*"[16]. He comes to an unexpected conclusion, that Penrose's arguments about formalization possibility of our thinking are correct. However, the noted above limitations tell us not about the "inferiority" of machines and the intellectual "superiority" of humans, but about the inherent limitations of our ability to think about our own reasoning process: "*This limitation is not due to a lack of intelligence on our part but is inherent in any system of reasoning that is capable of reasoning about itself*" [16] These arguments demonstrate that for analysis of consistency problem in mathematical reasoning the difference between human and machine is insignificant [17].

This disputable entity is an intrinsic property of human thinking and is known as reflexivity. Mentalists use this property as evidence for superiority of natural intelligence over artificial intelligence, thanks to which one can go beyond formal systems in one's reasoning, abstracting from the object language to the level of metalanguage, formulating sentences in the metalanguage that cannot be proved in the object level, etc. However, Tselishchev shares the point of view that the property of reflexivity also works as a certain kind of internal limitation inherent in human reasoning as such, and this limitation can be seen in operations over transfinite objects. Reflexivity is used as a principle for constructing transfinite sequences for recursive ordinal numbers. This principle is to add the Gödel sentence to obtain a consistent or valid system. It allows one to attach to the theory true but unprovable sentence, which raises the question about formal possibility of reflection principle constant iteration through recursive ordinals. Ultimately this means that "an attempt to obtain guarantees of consistency and validity by formalizing reasoning is doomed to failure" [18].

Lucas parried arguments about the possibility of the human intelligence contradictory nature. He treated it not as a result of ontological inferiority but as a simple and annoying failure in his work, just a common mistake: "*A man's untutored reaction if his consistency is questioned is to affirm it vehemently: but this, in view of Gödel's second theorem, is taken by some philosophers as evidence of his actual inconsistency. Professor Putnam has suggested that human beings are machines, but inconsistent machines. If a machine were wired to correspond to an inconsistent system, then there would be no well-formed formula which it could not produce as true; and so, in no way could it be proved to be inferior to a human being. Nor could we make its inconsistency a reproach to it --- are not men inconsistent too? Certainly, women are, and politicians; and even male non-politicians contradict themselves sometimes, and a single inconsistency is enough to make a system inconsistent. The fact that we are all sometimes inconsistent cannot be gainsaid, but from this it does not follow that we are tantamount to inconsistent systems. Our inconsistencies are mistakes rather than set policies. They correspond to the occasional malfunctioning of a machine, not its normal scheme of operations*" [19].

Tselishchev construed the comparison of human reasoning and the "reasoning" of a computer as a violation of common sense and even a certain paradox. First of all, it was assumed that a brain is computer-like entity, but soon it became clear that our knowledge about computers is much more precise and complete than our knowledge about brains. For this reason, comparisons between brains and computers are unproductive: "*But there is also more radical position, according to which the comparison between a computer and a person is completely wrongful. Because a person has neither completeness nor consistency, the statement that a computer does not possess these qualities also, does not give us anything by trying to compare them*" [20]. Indeed, the presence of common logically negative properties between two entities could not play a sufficient role for a conclusion about their common nature. It can be concluded that the so-called Gödel's argument was adopted by critics of computational reductionism, but it turned out to become a two-edged argument. This adopted argument have been easily using against themselves in a very specific manner, involving the necessity to prove their own consistency. Moreover, this imputation is Jesuitical, since it does not come from demands to throw off the burden of proof, but from the critics' expectation of inevitable paradoxical consequences [21].

5. Smullyan Levels of Machine Awareness and "Zero" Level of Reasoning

Despite the unfinished dispute between mechanists and mentalists and well-founded fears that nature of the dispute is dialectical and eternal, it inspired extremely non-trivial ideas and hypotheses. It has to be reminded that mentalists were criticized for using mathematically obscure epistemic modalities like "see" and "know".

Meanwhile the formal power of reductionism should not be underestimated. Reflecting on Gödel's problems with help of epistemic logic instrumentation, Raymond Smullyan uses the specifically understandable modality "believe in" and the standard tools of propositional logic. He uses the Gödel epistemic operator "B", so the proposition Bp have to be understood as a judgment in which some Reasoner "believes". Reasoner may be treated as a computer or a perfect natural mind from the logic point of view. Furthermore, Bp is not only a certain belief but also a proposition that is provable in the system. It has to be pointed out that identification of act of belief and provability is the weak link constriction. This identity place restrictions on interpretation of subsequent results in concern to natural mind. However, this issue requires extra accurate study [22].

To begin with, Smullyan introduces a formal ranked typology of ideal intellectual subjects - "Reasoners", which should be interpreted as formal systems and, therefore, "intelligent" machines. It makes sense to pay attention to the property of the 4th level (Type-4) Reasoner. For any proposition p, the Type-4 Reasoner believes in $Bp \supset BBp$. The non-trivial conclusion is that whatever we can prove about Reasoners using propositional logic, any Reasoner of type 4 can prove about himself, since he knows propositional logic and knows that he is a Type-4 Reasoner [23]. Any Reasoner of a higher reflective rank (above 4) possesses this epistemic property and does not acquire new ones. If to make an attempt to naturalize this architecture, this epistemic level corresponds to such a reflexive property known as "awareness of self-awareness".

At first glance awareness of self-awareness is an extremely sophisticated and abstract property of mind. The Israeli physicist and philosopher D. Tannenbaum, who died early, paid attention to it along with the self-awareness state that is successfully naturalized in mirror test. Tannenbaum interpreted awareness of self-awareness as a real mental state construed as a result of brain development by training: "*Indeed, awareness of one's self awareness has been one of the central paradoxes of human existence. It has been a central focus in nearly every major religion and philosophy. It is intimately connected with the mind-body problem and the concept of a metaphysical "soul". It may therefore seem strange that a biochemical machine should develop questions that, from a scientific point of view, do not make sense and are therefore unanswerable. However, if we view the emergence of self awareness as a consequence of associative memory and learning in a sufficiently complex brain, then we can rationally speculate on the emergence of this phenomenon*" [24].

This means that genuine intelligence begins at higher reflexive levels than levels of consciousness and even the level of self-awareness. In other words, awareness of self-awareness is the most fundamental form of human reasoning - a kind of "zero" level of rationality, in relation to which even self-awareness states turn out to be a negative level of reasoning [25]. Reflexive ranks below awareness of self-awareness can be considered as negative levels of reasoning, in the sense that, reducing to them, intellect significantly loses completeness.

6. Strange Reflexive Properties and Deception Detection for Improving Human Reasoning

Naturalization of reflexive properties to physical and neurophysiological content seems far from hopeless. We are accustomed to considering truth as fundamental gnoseological concept but we know such a kind of truth that can be construed ontologically. It sounds bizarre but this truth is of an epistemic sort (studied by so called *epistemic logic*), i.e., the truth of the form "*x* believes *p*". From this natural standpoint the truth of a particular belief or knowledge is the certain state of mind. Hypothetically, this real epistemic condition as a state of mind or identical to it state of logic gates can be detected. In practice this task was solved with a satisficed reliability by polygraph.

However, at present time, deception detection is used exclusively to clarify the objective or the first-order truth. But what will happen if the subjects would be questioned by self-referential questions that lead to contradictions and paradoxes? We will get some distribution of answers. Will this knowledge be of value, and if so, what kind of value?

In this concern Smullyan's introduced the two very strange epistemic properties: stability and normality: «*We will call a reasoner stable if for every proposition p, if he believes that he believes p, then he really does believe p. We will call the reasoner unstable if he is not stable, — e.g., if there is at least one proposition p such that the reasoner believes he believes p, but he does not actually believe p … instability is as strange psychological characteristic as peculiarity…We note that stability is the converse of normality. If a normal reasoner believes p, then he believes Bp, whereas if a stable reasoner believes Bp, then he believes p*» [26].

Instability is conceivable but bizarre mental property. Is its natural realization so impossible? Beliefs can arise in metalanguage and at the high reflexive levels of thought. Does it mean that they will "survive" during decline of reflexion rank and in a case of language simplification to an object-notation level? This kind of dramatic process of abandoning original beliefs may be observed in attempts to naturalize high-abstract statements. When someone claims to believe in a Trinity, it usually means that he considers himself a theist and actually believes that he believes in a Trinity. In the case of close examination, it often turns out that this belief has no sufficient theological basis in certain mind and should be considered as false.

Indeed, we may suppose the existence of mathematicians who believe that they believe that Gödel's first theorem is true. It may be turned out that testing such mathematicians on a deception detector would reveal that some of them do not really believe in the truth of this disputable theorem. This is conceivable, since it is known that a person may have a belief at a conscious level supported by many additional conscious assumptions but at the same time does not have deep conviction and experiencing serious subconscious doubts about the truth of his belief.

The stability epistemic property could be efficient in technical expert systems of a mixed type, consisting of two circuits: 1) logically non-transparent perceptron that recognizes and classifies data; 2) classical algorithmic circuit, which carries out logical conclusions based on data of the first circuit and other kinds of rational inputs. One may imagine a situation where algorithmic conclusions come into logical conflict with information at the output of a neural network, which asserts something without transparent rational grounds. This epistemic state of the system may be construed as unstable. If to use this analogy, the human mind should be attributed to the mixed systems of this kind, which was described by Aristotle in the doctrine of three kinds of souls: the nutritive, the sensitive-locomotive, and the rational [27].

It has to be expressed that progress in deception detection opens wide opportunities not only for identifying liars in the interests of the criminal investigations and recruiting. It could also be useful for assessing the sincerity of experts in expert interviews, for example, to increase the reliability of the Delphic method. However, the effectiveness of calibrating expert questions depends on the following theoretical question: does the truth of objective judgments depends on the experts' belief in that they are properly believe in these judgments?

Let's return to the idealizations of normal and stable Reasoners with some modifications. Let's replace Reasoners with Experts.

If Expert is normal: (1) $Bp \rightarrow BBp$

If the Expert is stable: (2) $BBp \rightarrow Bp$

BBp-type propositions are available in public judgments of Experts in the form of expert proofs that p. It may sound strange, but if it is BBp, then the results of the examination are accepted as true that p. This is based on the assumption that Experts make only formally provable statements. Bp statements are the Expert's real inner beliefs. Therefore, let us assume that only a polygraph is capable to detect Bp or $\neg Bp$, i.e., whether the expert really believes in the truth of his belief or not, whether he accepts this belief or not.

These theoretical ideas may be useful to improve trust for weight assignment algorithms in fully connected neural networks [28], to increase logical transparency of deep neural networks outputs [29] and for logical determination of structural parameters of multilayer perceptrons [30].

7. Conclusion

Unfortunately, we don't have knowledge about the distribution of these strange properties in real acts of reasoning. It means that if we have a case detecting the false belief of certain Expert, we may conclude that this Expert is unstable. If we have a case an Expert makes a false conclusion, but with help of a polygraph revealed he believes in truth of this conclusion, it means that Expert is abnormal in the Smullyan sense.

It is difficult to say what is good for the needs of the reliable expert systems in terms of provability requirements. It depends on which version of the philosophy of mathematics we adhere to - intuitive or constructive. From constructivism standpoint it is clear that instability is less negative phenomenon. Is it important for what reasons expert does not believe in his beliefs?

Is it important if he simply can't accept the truth of the axioms used for his inferences? From this point of view, it is much more important that the truth of judgments is reinforced at a higher reflexive rank, into which a powerful formal system of justifying sentences is "installed". The fact that these beliefs cannot be immersed to their very foundations is hardly significant, i.e., corruption of truth after decreasing of the reflection rank is not a critical phenomenon.

In accordance with intuitionism, interpretation should be completely different — the most disgusting property is *instability*, because statements that are not grounded in deep convictions have no value. It is the different matter

when the rank of reflection increases so the subject is not able to translate his beliefs to a higher reflexive level [31]. An epistemic abnormality that prevents to increase the reflection rank of a certain statement would hardly be treated as positive in concern of human intelligence.

References

1. James Moor, "The Dartmouth College Artificial Intelligence Conference: The Next Fifty years', AI Magazine, Vol. 27, No. 4, pp. 87-9, 2006.
2. Ng Gee Wah, Leung Wang Chi, "Strong Artificial Intelligence and Consciousness" Journal of Artificial Intelligence and Consciousness, Vol. 7, No. 1, pp. 63-72, 2020.
3. Tselishchev V.V. Algoritmizaciya myshleniya. Gedelevskij argument Moskow: Lenand, 2021, p.8 (in Russian).
4. Hilary Putnam In: Sidney Hook (ed.) / "Dimensions of Minds', New York, USA: New York University Press, pp. 138-164, 1960. Ibid., p.127.
6. Lukas J. Minds, machines, and Gödel / Philosophy, 1961. Vol. 36. pp.112-127
7. Tselishchev V.V. Algoritmizaciya myshleniya. Gedelevskij argument Moskow: Lenand, 2021, p.25 (in Russian).
8. Davis Matrin. "How Subtle is Gödel's Theorem?" / Behavioral and Brain Sciences. Vol.16, 1993, pp. 611-612.
9. Dawkins Richard "Postmodernism disrobed', Nature, Vol. 394, 1998, pp. 141-143.
10. Tselishchev V.V., p.18. Ibid. Ibid., p.44.
13. Roger Penrose "The Emperor's New Mind Concerning Computers, Minds, and the Laws of Physics", Oxford University Press, 2016, p. 14.
14. Poincaré Henri "Mathematics and Science: Last Essays" (Dernières Pensées), Dover, 1963.
15. Benaceraff P. God, Davil and Godel / Monist, Vol. 51, 1967, pp. 9-32.
16. D. McCullough "Can Humans Escape Gödel? A Review of Shadows of the Mind by Roger Penrose" PSYCHE: An Interdisciplinary Journal of Research On Consciousness, 2:57-65, 1995, p.64.
17. Tselishchev V.V., p.187. Ibid., p.18.
19. Lukas James, pp.120-121.
20. Tselishchev V.V., p.16.
21. Vinnik D.V. Contradiction as a positive property of the mind: 90 years of Gödel's argument // Russian Journal of Philosophical Sciences. 2021. Vol. 64. No. 7. pp. 26-45. (in Russian, URL: https://elibrary.ru/download/elibrary_48735885_33581826.pdf)
22. Raymond Smullyan "Forever undecided. A Puzzle Guide to Godel", USA, 2015, pp. 166-167.
23. Ibid., pp.108-110.
24. Tannenbaum E.D. Speculations on the emergence of self-awareness in big-brained organisms // URL: http://arxiv.org/pdf/q-bio/0701017.pdf
25. Vinnik D.V. 'Awareness of self-awareness as 'zero" level of intellect' / Philosophy of Science. 2015. No 4 (67). pp. 76-96. (in Russian, https://elibrary.ru/download/elibrary_25432170_66653420.pdf)
26. Smullyan Raymond, pp. 166-167.
27. Vinnik D.V. On the possibility of hybrid artificial intelligence // Philosophy of Science. 2022. No 3 (94). pp. 92-109. (in Russian, URL: https://elibrary.ru/download/elibrary_49551668_59031655.pdf)
28. Karande A.M., Kalbande D.R. Weight Assignment Algorithms for Designing Fully Connected Neural Network. IJISA, 2018, 10(6): 68-76.
29. Dharmajee Rao D.T.V., Ramana K.V. Winograd's Inequality: Effectiveness for Efficient Training of Deep Neural Networks. IJISA, 2018, 10(6): 49-58.
30. Hu Z., Tereykovskiy I.A., Tereykovska L.O., Pogorelov V.V. Determination of Structural Parameters of Multilayer Perceptron Designed to Estimate Parameters of Technical Systems. IJISA, 2017,9(10):57-62.

31. Vinnik D.V. Reflective and intentional states of mind. their specific character and relationship // Humanitarian Sciences in Siberia. 2003. No 1. C. 7-10.

About the Author

Dmitriy Vladimirovich Vinnik received D.Sci. degree in 2011 in the Institute of Philosophy and Law, Siberian Branch Russian Academy of Sciences. Current main research interests include philosophy of science, epistemic logic, physical and computational theories of mind, philosophy of mathematics.

♦♦

Self-Awareness – a Review

Self-awareness has been called "arguably the most fundamental issue in psychology, from both a developmental and an evolutionary perspective." Self-awareness theory, developed by Duval and Wicklund in their 1972 landmark book *A theory of objective self-awareness*, states that when we focus on ourselves, we evaluate and compare our current behaviour to our internal standards and values. This elicits a state of objective self-awareness. We become self-conscious as objective evaluators of ourselves. Self-awareness should not be confused with self-consciousness. Various emotional states are intensified by self-awareness. However, some people may seek to increase their self-awareness through these outlets. People are more likely to align their behaviour with their standards when they are made self-aware. People are negatively affected if they do not live up to their personal standards. Various environmental cues and situations induce awareness of the self, such as mirrors, an audience, or being videotaped or recorded. These cues also increase the accuracy of personal memory.

In one of Andreas Demetriou's neo-Piagetian theories of cognitive development, self-awareness develops systematically from birth through the life span and it is a major factor for the development of general inferential processes. Self-awareness about cognitive processes contributes to general intelligence on a par with processing efficiency functions, such as working memory, processing speed, and reasoning.

Albert Bandura's theory of self-efficacy describes "the belief in one's capabilities to organize and execute the courses of action required to manage prospective situations." A person's belief in their ability to succeed sets the stage for how they think, behave, and feel. Someone with a strong self-efficacy, for example, views challenges as tasks to engage in, and is not easily discouraged by setbacks. Such a person is aware of their flaws and abilities and chooses to utilize these qualities to the best of their ability. Someone with a weak sense of self-efficacy evades challenges and quickly feels discouraged by setbacks. They may not be aware of these negative reactions and therefore, may not be prompted to change their attitude. This concept is central to Bandura's social cognitive theory, "which emphasizes the role of observational learning, social experience, and reciprocal determinism in the development of personality."

Courtesy: https://en.wikipedia.org/wiki/Self-awareness