

Postproceedings of the 10th Annual International Conference on Biologically Inspired Cognitive Architectures, BICA 2019 (Tenth Annual Meeting of the BICA Society)

Consciousness for AGI

Piotr Boltuć*

*University of Illinois, Departments Philosophy, Computer Science, One University Plaza UHB 3030, Springfield, IL 62703, USA
Warsaw School of Economics, Warsaw, Poland*

Abstract

AI can think, although we need to clarify definition of thinking. It is cognitive, though we need more clarity on cognition. Definitions of consciousness are so diversified that it is not clear whether present-level AI can be conscious – this is primarily for definitional reasons. To fix this would require four definitional clusters: functional consciousness, access consciousness, phenomenal consciousness, hard consciousness. Interestingly, phenomenal consciousness may be understood as first-person functional consciousness, as well as non-reductive phenomenal consciousness the way Ned Block intended [1]. The latter assumes non-reducible experiences or qualia, which is how Dave Chalmers defines the subject matter of the so-called Hard Problem of Consciousness [2]. To the contrary, I pose that the Hard Problem should not be seen as the problem of phenomenal experiences, since those are just objects in the world (specifically, in our mind). What is special in non-reductive consciousness is not its (phenomenal) content, but its epistemic basis (the carrier-wave of phenomenal qualia) often called the *locus of consciousness* [3]. It should be understood through the notion of ‘subject that is not an object’ [4]. This requires a complementary ontology of subject and object [5, 6, 4]. Reductionism is justified in the context of objects, including the experiences (phenomena), but not in the realm of pure subjectivity – such subjectivity is relevant for epistemic co-constitution of reality as it is for Husserl and Fichte [7, 8]. This is less so for Kant for whom the subject was active, so it was a mechanism and mechanism are all objects [9]. Pure epistemicity is hard to grasp; it transpires in second-person relationships with other conscious beings [10] or monads [11, 12]. If Artificial General Intelligence (AGI) is to dwell in the world of meaningful existences, not just their shadows, as the case of Church-Turing Lovers highlights [13], it requires full epistemic subjectivity, meeting the standards of the Engineering Thesis in Machine Consciousness [14, 15].

© 2020 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Peer-review under responsibility of the scientific committee of the 10th Annual International Conference on Biologically Inspired Cognitive Architectures.

* Corresponding author. Tel. +217 206 6600.

E-mail address: pboltu@sgh.waw.pl

Keywords: machine consciousness; AI consciousness, AI thinking, AI cognition, AGI-consciousness; Artificial General Intelligence; phenomenal consciousness; p-consciousness, functional consciousness, f-consciousness, a-consciousness, h-consciousness; hard consciousness; subject-object: pure subject; Thomas Nagel; View from Nowhere: The Engineering Argument in Machine Consciousness; Church-Turing Lovers

1. AI Thinking

Turing was right that, eventually, there would be no cognitive activity that humans conduct better than computers. Lady Lovelace, to whom Turing refers as a potential critic, had a good point that "The Analytical Engine has no pretensions to originate anything. It can do whatever we know how to order it to perform" [16]. Nowadays, it is not the case that all AI engines are 'analytical engines' in that sense. We also have weaker and weaker reasons to view human thinking as a process dominated by predicative logic. For instance, genetic algorithms do create new useful information through the cognitive architecture consisting of a randomizer focused on limited transformations of the image of preexisting reality and an AI 'critic' that selects those permutations of reality (designs) that satisfy functional specifications [17]. There are various designs relying on highly complex stochastic process occurring at the edge of chaos [18]. Just like AI, humans seem to think and create primarily at the unconscious level [19], while language is an evolutionary device helpful in social transmission of science, and less so in individual creativity, including mathematical and scientific discovery, where intuitively grasped paradigms and complex dynamic visualizations play a large role [20].

Some skeptics still search for cognitive domains in principle inaccessible to AI. If one understands that all functions, including the cognitive ones, are guided by algorithms, which, qua algorithms, are necessarily replicable in AI (physical interpretation of the Church-Turing Thesis [21]), it simplifies the debate immensely. There is no good reason to expect any cognitive functionalities unique to animal, in particular human. brains (because they would have to be non-algorithmic, *whatever that would mean*).

One may think of motivations as separate from cognitive functions, but this is flawed. Copious work on robot emotions and motivational structures, as simple as structured rewards, demonstrates such distinction to be irrelevant. Hobbes' explanation of the functioning of mind, including his analogy between magnetic forces and attraction or repulsion in animals [22], explained such mistakes out even before the Babbage computer.

The triumph of those reductive explanation applies to all human cognitive activities (thinking, broadly understood); thus, there is a strong propensity to extrapolate that reductive explanations also apply to the area of consciousness. It does not help that the term consciousness is ill-defined, especially at the level of explanation familiar to the general audience (which in this case means nearly everybody, except for a handful of philosophers of mind). Even the 'experts' tend to disagree, and in vague terms at that. Hence, my first task is to help define consciousness for AI.

2. Consciousness for AI

What is the difference between consciousness, cognition and thinking?

Proposing regulative definitions: It is helpful to view **thinking** as information processing that increases the likelihood of getting from true premises to true conclusions (mistaken thinking is a procedure likely to attain this goal that fails for some formal or contingent reason). Hence, not all thinking has to be conscious; in particular, AI engines do not have to be conscious in order to perform their functions (which often rise to the level of thinking in the above sense.)

Cognition can be defined as processing of sensory inputs, assimilated in cognitive schemes. Cognitive architectures, biological or artificial, that do such processing are cognitive engines.

Consciousness is more elusive since the notion covers several clusters of ideas. In general, it is a high cognitive function, which may involve visualization or similar processes pertaining to different sensory-based imagination [Boltuć 2007], but this is not quite a definition, merely a pointer. Analysis of different cognitive clusters within the idea of consciousness turns out more productive.

The types of consciousness I find focal are F-, P-, A- and H-consciousness. **Functional-consciousness** is whatever performs conscious functions; or, whatever behaves as if it had consciousness. This definition's theoretical weakness is that it refers to the sort of pointer definition I used above. But practically this problem dissipates since we all can single out the set of advanced cognitive functions that are f-conscious. People disagree what some of those functions are, but we are not to involve such level of detail.

Phenomenal-consciousness is “experience; the phenomenally conscious aspect of a state is what it is like to be in that state” [1]. In terms of denotation this definition is identical with Chalmers' definition of consciousness in the Hard Problem of Consciousness [2]. The difference lies in connotations since for Chalmers there is a methodological gap between scientific methods and possible explanation of phenomenal experience (following Leibniz mill [11]), whereas for Block such methodological gap may be illusory, resulting from limitations of past, or also current, science.

Block's attempts to situate phenomenal consciousness as explainable by the sciences has a kick to it. His definition of phenomenal consciousness can be read in strictly functional terms, so that it pertains to cognitive functions of a robot [24], which is correct *de dicto*, but does not seem consistent with Block's intended connotation.

Access-consciousness occurs “when you have an episode of phenomenal consciousness and it is available to your cognitive systems” [1]. One may say, in terms of Baars' global workspace, that it accesses content that is 'globally' broadcast in the system. It is important for Block to emphasize that one may have phenomenal consciousness that is not access consciousness [25], which means not accessed at a given time – some phenomenal experiences may not be attended to. On the other hand, access consciousness is always phenomenal since there is no other content.

Hard-consciousness, the sort of consciousness defined by Chalmers in the Hard Problem [2], and actually in the Reformed Hard Problem, is the basis of h-consciousness [23]. For Chalmers the hard problem of consciousness is the problem of experience and the hard consciousness is the kind of consciousness that has those experiences. Isn't it just Block's phenomenal consciousness? Well, denotation may be the same – denoting the stream of phenomenal qualia. Yet, for Block that consciousness seems to have more epistemic machinery attached to it – the ways we interact with reality, we can investigate it etc., which is why Franklin's reading of it as just first-person functional consciousness is technically correct, though unintended by Block. Yet, such reading of Chalmers' notion of experience resulting in the hard problem of consciousness would disallow Franklin's reading – the epistemic machinery attached to this notion by Block may not change its denotation but it does extend its connotation into some first-person functional domain. Chalmers' notion of the hard problem as the problem of experience, epistemically more detached and conceptually reduced, which brings about the question: What if we reduced this concept even further? Yes, I think we ought to reformulate the Hard Problem, to which we now proceed.

3. The Hard Problem beyond Phenomena

The red herring of non-reductive consciousness are secondary qualities, qualia, phenomena. How so? Think of a phenomenal object: an object in your thoughts – or of your perception, viewed within your stream of consciousness – is an **object** nevertheless. Phenomenal objects can be made objective (*objectivised*). Just like in the Clark-Chalmers extended mind hypothesis where your notebook may belong to your mind extended far beyond one's skull [26], the opposite is true as well. Objects internal to your mind, those *in the skull*, may be externalized in a straightforward way. Take the experiments by Jack Gallant's group of putting images that you see (and recently they whole phenomenal movie you see in your mind) directly from one's visual cortex directly on a computer screen [27]. Your visual cortex is a bit like a biological camera and you can get the recordings directly from it. There is also a developing research on reading thoughts and other non-visual information from various areas of one's brain; *good bye the insurmountable tower of privileged access*. Our brain is an object and so are its products, including internal phenomenal images in our mind. *Nothing's epistemically private anymore*.

Does this lead us towards reductive materialism? Very much so, since those little bastions of epistemic privacy and supposedly non-physical content needed to be gone long ago, and they are getting gone now. This leads towards

reductive materialism insofar as the objects are concerned. So, reductive it is, but not all the way. What's left is the locus of consciousness [3], the very stream of one's awareness in which the content (always built only of objects, phenomenal or otherwise) becomes transparent to the mind. The way to avoid being overly poetic, and vague, about it is to go back to early Nagel, and his view on 'the subjective' [6], but there is a reason why he seems to have abandoned this early, promising project: Such account is very hard to make within the post-Locke, post-Hume Anglo-American tradition. It is easier to formulate in the tradition of classical German philosophy, from Leibniz, through Kant and Fichte all the way even to Husserl [11, 9, 8, 7]; yet, this philosophical tradition has been relegated to the studies of the history of philosophy and, as a living philosophy, it is misunderstood and barely alive.

In his first major book [28] the author reprinted his now very broadly read article "What is it like to be a bat?" [29], which may have given rise both to Block's view on phenomenal consciousness and Chalmers' definition of the Hard Problem [1, 2]. Imagine what it is like to have the other senses, such as echolocation, that we are barely acquainted with. It is impossible to do. In the book this paper [27], chpt. 12 is just an intuition pump aimed to clear the way for the crucial [27] chpt. 14: "Subjective and Objective", which is new to the book. The latter chapter is also the gist of Nagel's *The View from Nowhere* [6], the book viewed for many years as his masterpiece. Later, due largely to the highly skeptical critiques by McGinn, Nagel's main complex argument went out of fashion whereas its propedeutical version – the 'what it is like approach [29] – took the center stage.

4. Complementary Philosophy

Nagel's early philosophy was permeated by the seemingly paradoxical question: How is it possible that I am one of the beings, among all the beings in the objective world? It does not matter whether there are identical beings with me, whether I share continuity and connectedness (physical or mental) with any beings, and so on – even if they and me are very much alike, there is still an open question, which one of them is me. Being me is the only way to experience oneself's phenomenal experiences, even if those could be shared by some objective means with the others. According to Nagel, the problems such as personal identity, mind and body, free will or agent-centered morality cannot be detached 'from the subjective point of view on which they depend for their existence' [28, p. 213].

Nagel may sound a bit like a dualist – e.g. when asking "how one can include in the objective world a mental substance having subjective properties" -- his main focus is on keeping room for the two complementary perspectives, subjective and objective "because the same individual is the occupant of both viewpoints" [28, p. 208]. We live in the two mutually irreducible perspectives, subjective and objective. Both, 'idealism and its objectifying opposite' share 'a conviction that a single world cannot contain both, irreducible points of view and irreducible objective reality' [28, p. 212]. Actually, the early version of Russellian monism, in his analysis of Mind [5] presents this exact same complementarity of the two perspectives, objective and subjective. A few years later, in his Analysis of Matter [30] Russell's view becomes a version of anomalous materialism, and so does Nagel [6]. We may see the object and subject as two different viewpoints on reality, that play a complementary role [14, sec. 4].

Interestingly, Nagel may be avoiding the claim of uniqueness of (or exclusive access to) the content of one's first-person consciousness: "What is more subjective is not necessarily more private", he explains since subjective experiences are, in some sense public property [28 p. 207]. He also makes the following remark: "subjective aspects of the mental can be apprehended only from the point of view of the creature itself (perhaps taken up by someone else)" [28 p.201]. The puzzling point is the parenthetical remark. Did Nagel make a simple point, made multiple times in [6], that we can take somebody else's point of view, just like several pictures can be taken from the exact same location? But this would be his argument for objective aspects of the subjective view. This does not seem to pertain to the 'subjective aspects of the mental'. Perhaps this could be seen as anticipation of something like Chalmers' *dancing qualia argument*: the idea that, provided the right level of neuroscience, we may be able shift observers from the insight of one brains to the other [2]. The argument like *dancing qualia* may be applied to verifiability of the first person experience [31, 32]. This issue is important since non-private settings of the first-person consciousness (and others having the opportunity to pick into it, in various manners) seem like the future of studies on first-person consciousness, where Galant's experiment [27] is just the beginning. Scaffolding of one's brain is as objective as any other objects in the universe; the only puzzling thing is what aspect or part of the brain gives us the first-person feel (and conscious understanding) of being there.

4.1. Section headings

Section headings should be left justified, bold, with the first letter capitalized and numbered consecutively, starting with the Introduction. Sub-section headings should be in capital and lower-case italic letters, numbered 1.1, 1.2, etc, and left justified, with second and subsequent lines indented. All headings should have a minimum of three text lines after them before a page or column break. Ensure the text area is not blank except for the last page.

4.2. Subject not Object

The complementary view on subject and object implies that the subject is not an object of any kind, not even a substance. This amounts to putting all the objects to one side of the equation so as to help us see more clearly what is left. I have discussed this issue in some detail lately. Here let me just mention a few points. “The most reduced definition of pure epistemic subject leaves such subject with no direct predicative features (...) we can predicate *about* it, the way we do in the present sentence, but we cannot predicate *of* it, in the narrow sense of providing a direct description. By predicating *about* the epistemic subject we use a meta-level of reference” [33]. Such subject is only the condition of first-person (non-reducible, I presume) epistemicity – the beginning of the first-person perspective. It can be ontologically explained by accounts such as non-reductive materialism, double aspect theory, neutral monism [5] but hardly through dualism, since the latter requires mental substance. Yet, mental substance is also an object and would violate the criterion that subject is not an object in an ontological sense.

5. AGI and non-reductive subjects

How is the above philosophy of pure subject related to AGI? **First**, it shows how AI consciousness we have is not quite the kind of first-person consciousness that we assume alive people and animals have. **Second**, it points to what would have to happen for robots to have first-person non-reductive consciousness.

5.1. The Engineering Thesis

The Engineering Thesis in Machine Consciousness is the idea that if something characterizes humanoid brains it could, in principle, be engineered (or bioengineered) in AI. We are within the realm of science and more specifically science-based engineering. In the engineering realm we should not be overly particular whether what we can build is accounted for by non-reductive materialism, by panpsychism (where the mental is just one substance about the many substances – in the spectrum from Ann Conway [12] to David Chalmers [2]), complementary view [5, 28 chpt. 14, 6] or double aspect theory. Non-naturalistic worldviews, or those forms of naturalism quite different from contemporary post-materialistic world-view many scientists hold, would impose further ontological or theological conditions. Let us abstract from those conditions, which does not amount to rejecting or being in any way partial for or against such views – those further conditions may be added to our bare-bone model as needed. Here is the bare bone conceptual engineering model:

If human beings and other animals have first person consciousness, then neuroscience or another discipline should eventually discover how it is generated. If so, we should eventually be able to build generators of first-person consciousness [Boltuc 15, 14, 23]. Practical or ethical issues would be important once we reach proximity of realizing such model, but they are premature at this early conceptual phase.

6. Subsymbolic

Symbolic structures can be viewed as more advanced in AI than subsymbolic since they allow full human readability [34]. But subsymbolic structures are incompatibly richer than the standard propositional use of symbolic language. Thus, human readability should not be viewed as the gold standard of information processing [35], though

it is a standard of control in limited capability AI environments. Multifarious gestalts [36] and what I call Spinozian phenomenology able to grasp non-human multitudinal qualities of experience [20] at the subsymbolic level convey incomparably richer information than predicative language.

The same seems true about human thinking since Libet's experiment revealed that consciousness is only the tip of the iceberg of human cognition [19]. Symbolic thinking, especially linguistic communication, is very strongly linked with phenomenally conscious activity. Hence, Libet leads us to question dominance of linguistics in human mind. Though it may still be a convenient way of transferring intellectual culture, especially history and science. Poetry uses language largely in a sort of subsymbolic manner, so do creative music and visual arts. They are largely subliminal and in broad terms can be viewed as subsymbolic.

AI is behooved by using subsymbolic level of information processing as dominant, including direct take on metaphors as subsymbolic permutations of reality – instead of being 'translated' through purely syntactic complex structures that express only language games.

7. AGI and porogress

We should be proud of nearing, slowly but assuredly, origination of Artificial General Intelligence. Just like builders of the pyramids or those who discovered fundamental laws of geometry, dynamics or quantum mechanics would rightly be proud of extending horizons of humanity. Those involved should be proud of transcending the capabilities of un-augmented human mind.

Fear comes from excessive prudence, which is the opposite of success, glory or progress. Fearful roaches would have never evolved into humans, so to say. Posthumanist speculations aside (since this sort of futurology barely ever works in a long run), we should be primarily proud, not primarily worried.

When Mori discovered the Uncanny Valley effect [37], he was preoccupied with quite imperfectly humanoid robots invading human spheres of privacy. Yet, there is a second uncanny valley, the valley of imperfect perfection, which happens when AI transcends human abilities but not to the point of no contest [13]. This second uncanny valley effect is the gist of the singularity fear among the fearful elites trying to slow down or throttle developments in artificial intelligence.

The slowdown of civilizational progress is the opposite of prudence, even in environmentalism [38], but much more so in engineering and cognitive sciences. Not only Adam Smith, but also Karl Marx (in his crushing critique of Proudhon and his endorsement of luddite movement), have understood that attempts to strangle progress, however well-meaning attempts, have always been the opposite of humanism [39].

8. The Importance of findings

It is time for the sum up of this visioning paper since many issues ought to be brought home:

8.1. *How practical would metaphysics of non-reductive subject turn out to be?*

What is subject that is not an object good for?

Pure subjectivity is relevant for epistemic co-constitution of reality, as visible in [8] and to some degree [9]. Without the epistemic level the objects would never cross from potentiality to actual existence since, as Kant shows, objects in themselves have insufficient specifications.

The value of the pure subject of first-person consciousness transpires most easily in second-person relationships with the other consciousness, or epistemic monad [11, 10, 40]. We can see this, a bit more analytically, in the case of Church-Turing Lovers [13]. Imagine that one can have one of the two significant others (companions and lovers), who are identical in any way but one, companion has first-person consciousness and the other does not (being a perfect humanoid machine or a philosophical zombie). Since we have existential/teleological reasons, of our lives making non-solipsistic sense, to care whether our significant have positive first-person experiences in the sense-making

relationship, *ipso facto* we have reasons to care whether they have the experiences at all. This is because we care about them – through love – but also because relationships with significant others are sense-making relations for ourselves.

8.2. Long-shot visioning?

If AGI is to dwell in the world of meaningful existence, it needs to have first-person consciousness. In principle this condition could be satisfied by cyborgization. Yet, this wouldn't do since a machine with a borrowed brain, say of a squirrel, would not be integrated with the borrowed consciousness in the relevant sense. Even advanced brains, coming from *homo sapiens*, would hardly be the true consciousness of AGI. They would just be some consciousness **in** the AGI. If it was to direct it, this would turn AGI into a cognitive gear, not an integrated conscious entity. To have a brain is different for humans than for octopuses, with multiple cognitive centers spreads among their legs and the central unit primarily for coordination. AGI's may very well turn out to be mostly multi-cognitive centers cognitive architectures requiring their own set of experiences. If they were to be conscious, they would be so in their own unique way much more foreign to us than the brain, and first-person feel, of a bat with its echolocation capabilities [29].

If AGI is to have conscious existence, not to dwell in the ontic shadows [41], it requires full epistemic subjectivity. This would amount to meeting and surpassing the standards of the Engineering Thesis in Machine Consciousness [Boltuc 2009, 2012, 2007] – surpassing it, since AGI's mind would need to have the first-person consciousness superior to anything we can first-person imagine and envisage.

References

- [1] Block, N. (1995) "On a Confusion about a Function of Consciousness" *Behavioral and Brain Sciences* 18 (2): 227–287.
- [2] Chalmers, David. (1986) *The Conscious Mind*, Oxford University Press.
- [3] Shalom, Albert. (1989) *Body/Mind Conceptual Framework and the Problem of Personal Identity: Some Theories in Philosophy, Psychoanalysis and Neurology*; Prometheus Books
- [4] Boltuc, Piotr. (2019) "Subject is no Object." in: Mark Burgin; Gordana Dodig-Crnkovic (eds), *Epistemic Basis of Information, Philosophy of Information*, World Scientific 3–39.
- [5] Russell, Bertrand (1921) *Analysis of Mind*, London: George Allen and Unwin; New York: The Macmillan Company
- [6] Nagel, Thomas. (1986) *The View from Nowhere*, Oxford Univ. Press
- [7] Husserl, Edmund. (1931) *Ideas: General Introduction to Pure Phenomenology*; Routledge: Abingdon, UK, 1931
- [8] Fichte, J. G. (1970) *The Science of Knowledge: With the First and Second Introductions*; Heath, P. and Lachs, J., (eds) Cambridge University Press.
- [9] Kant, Immanuel. (1781) *Critique of Pure Reason*; Cambridge University Press: Cambridge, UK
- [10] Buber, Martin. (1970) *I and Thou*, Scribner. New York City
- [11] Leibniz, G. W. (1965) *Monadology* transl. Schrecker P, and Schreckjed A.m. Macmillan Library of Liberal Arts
- [12] Conway, Anne (1996) *Principles of the Most Ancient and Modern Philosophy*. Allison P. Coudert and Taylor Corse (eds) Cambridge: Cambridge University Press
- [13] Boltuc, Piotr. (2017) "Church-Turing Lovers" in K. A. Abney, K. A.; Patrick Lin, R. Ryan (eds.), *Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence*. Oxford: Oxford University Press, 214–228
- [14] Boltuc, Piotr. (2009) "The Philosophical Issue in Machine Consciousness." *International Journal of Machine Consciousness* 1(1), 155–176
- [15] Boltuc, Piotr. (2012) "The Engineering Thesis in Machine Consciousness" *Techné: Research in Philosophy and Technology* 16(2), 187–207
- [16] Turing, A. M. (1950) "Computing Machinery and Intelligence", *Mind*, 59(236), 433–460

- [17] Thaler SL (2012) "The creativity machine paradigm: withstanding the argument from consciousness". *APA Newsletter on Philosophy and Computers* **11(2)**:19–30
- [18] Goertzel, Ben. (2006) *The hidden pattern. A patternist philosophy of mind*. BrownWalker, Boca Raton
- [19] Libet, Benjamin. (2004) *Mind time: the temporal factor in consciousness, perspectives in cognitive neuroscience*. Harvard University Press
- [20] Boltuc, Piotr and Boltuc Marta. "Semantic Beyond Language." in Alexei V. Samsonovich (ed) *Biologically Inspired Cognitive Architectures 2019. Proceedings of the Tenth Annual Meeting of the BICA Society* 36–47
- [21] Deutsch, David. (1985) "Quantum theory, the Church-Turing principle and the universal quantum computer", *Proc. Roy. Soc. A*, **400**: 97–115
- [22] Hobbes, Thomas. (1982) *Leviathan* Penguin Classics
- [23] Boltuc, Nick and Boltuc, Piotr. (2007) "Replication of the Hard Problem of Consciousness in AI and Bio-AI: An Early Conceptual Framework" in Chella, A.; Manzotti, (eds) *AI and Consciousness: Theoretical Foundations and Current Approaches*, Merlo Park, CA: AAAI Press, 24–29
- [24] Franklin, Stan., Baars. Bernard and Ramamurthy, Umma. (2008) "Phenomenally Conscious Robots?" *APA Newsletter on Philosophy and Computers*, Fall, **08(1)**, 2–4
- [25] Block, Ned. (2008) "Phenomenal and Access Consciousness" *Proceedings of the Aristotelian Society*, New Series, **108**, 289–317
- [26] Andy Clark, David J Chalmers. (1998). "The extended mind". *Analysis*. 58 (1): 7–19
- [27] S. Nishimoto, S. and Jack Galant. (2011) "Reconstructing Visual Experiences from Brain Activity Evoked by Natural Movies" *Current Biology* **21**, 1641–1646
- [28] Nagel, Thomas (1979) *Mortal Questions*, Cambridge Univ. Press
- [29] Nagel, Thomas. (1974) "What is it like to be a bat?" *The Philosophical Review*, **83(4)**, 435–450
- [30] Russell, Bertrand (1927) *Analysis of Matter*, Spokesman, Nottingham
- [31] Boltuc, Piotr. (2010) A Philosopher's Take on Machine Consciousness" in Viorel E. Guliciuc (ed) *Philosophy of Engineering and the Artifact in the Digital Age*, Cambridge Scholar's Press, 49–66
- [32] Schneider, Susan 2019 *Artificial You: AI and the Future of Your Mind*, Princeton University Press
- [33] Boltuc, Piotr. (2019) "Subject is no Object." in: Mark Burgin; Gordana Dodig-Crnkovic (eds), *Epistemic Basis of Information, Philosophy of Information*, World Scientific 3–39
- [34] Kelley, Troy Dale (2006) "Developing a Psychologically Inspired Cognitive Architecture for Robotic Control: The Symbolic and Subsymbolic Robotic Intelligence Control System (SS-RICS)." *International Journal of Advanced Robotic Systems*, **3(3)**, 219–222
- [35] Sanz, Ricardo. (2019) "Consciousness, Engineering, and Anthropomorphism", *APA Newsletter on Philosophy and Computers* **19(1)**, 12–18
- [36] Boltuc, Piotr (2018) "Hard Semantic Computing" *Procedia Computer Science* **123**, 98–103
- [37] Mori, Masahiro (2012). "The Uncanny Valley." *IEEE Spectrum*, June 12.
- [38] Bisk, Tsvi and Boltuc, Piotr (2017) "Sustainability as Growth," in Technology, Society, Sustainability, Lech Zacher (ed.) Springer, 175–183
- [39] Marx, Karl. (1998) *The German Ideology, including Theses on Feuerbach* (Great Books in Philosophy, Prometheus Books
- [40] Levinas, Emmanuel (1987) *Time and the Other*, Richard A. Cohen (trans.), Pittsburgh: Duquesne University Press,
- [41] Plato (1961) *The Republic* transl. Shorey, Paul in *The Collected Dialogues of Plato* Edited by Edith Hamilton Huntington Cairns. 575–844