YU ISSN 0353-3891 UDK - 1

BELGRADE PHILOSOPHICAL ANNUAL 2022

FILOZOFSKI GODIŠNJAK

35/1

BELGRADE PHILOSOPHICAL ANNUAL/
FILOZOFSKI GODIŠNJAK 35(1)/2022
Institute of Philosophy, Faculty of Philosophy,
University of Belgrade
Belgrade, Čika Ljubina 18–20

Belgrade

Year XXXV

YU ISSN 0353-3891

UDK-1

Belgrade Philosophical Annual is published twice a year and is available online at
http: //www.f.bg.ac.rs/bpa/index.html

Printed by
Službeni glasnik, Belgrade

This issue is financially supported by
Ministry of Education, Science and Technological Development of the Republic of Serbia

The statement on publication ethics can be found at the journal website http://www.f.bg.ac.rs/bpa

BELGRADE PHILOSOPHICAL ANNUAL 35(1)/2022

SELECTION EFFECTS AND THE DEEP PAST

Articles

SELECTION EFFECTS AND THE DEEP PAST

This issue is a result of the presentations at the conference "Sciences of the Origin: The Challenges of Selection Effects and Biases" held at the University of Belgrade in the summer of 2021 and supported by the Ian Ramsey Center for Science and Religion (Oxford) initiative "New Horizons for Science and Religion in Central and Eastern Europe" and Templeton Foundation.

João Barbosa
Center for Philosophy of Sciences
of the University of Lisbon
jlbarbosa@fc.ul.pt

Original Scientific Paper
UDC: 524.852:167
524 Хојл Ф.

113/119

A *THEMATIC* APPROACH TO SELECTION EFFECTS AND BIASES IN COSMOLOGY: FRED HOYLE AND THE REJECTION OF THE BIG BANG IDEA, DESPITE THE EXPERIMENTAL OBSERVATIONS

Abstract: *Despite some important observations and after decades of widespread consensus around the big bang cosmology, Fred Hoyle, one of the proponents of the steady-state cosmology, continued to fight the big bang idea throughout his life.*

We can try to understand this persistent attitude of Hoyle through a Holtonian thematic *approach, by admitting that personal preferences and choices of scientists are conditioned by* themata.

Thematic analysis shows that big bang cosmology is mainly based on a set of themata *consisting of* evolution, finitude, life cycle *(which has a* beginning*), and* change*; the steady-state cosmology is based on opposite* themata: steady-state, infinity, continuous existence*, and* constancy. *Personal preferences seem to have been important in the strong and passionate dispute between big bang and steady-state ideas, and Hoyle is a very illustrative example of a personal commitment remarkably long-lived to some* themata*, in this case to the opposite* themata *of the big bang cosmology. In his personal and persistent struggle against the big bang idea, Hoyle always refused the way how some experimental observations were considered decisive in favor of this cosmology. This is a typical* thematic *attitude: letting some personal* thematic *preferences influence the acceptance or rejection of scientific evidence. In this case, that corresponds to the existence of selection effects and biases regarding important cosmological observations, in order to sustain a persistent rejection of the big bang idea.*

Keywords: *big bang cosmology; experimental observations; Fred Hoyle; selection effects and biases; themata.*

1. Fred Hoyle, a persistent opponent to an idea which, ironically, he helped to consolidate

Big bang cosmology (commonly known as the Big Bang Theory) has dominated all modern cosmology for more than fifty years.

After a strong dispute between the big bang cosmology and its big rival, the steady state cosmology, some important experimental observations in the

1960s, such as the observation of a great abundance of helium in the universe and the discovery of the cosmic background radiation were decisive for the progressive and wide acceptance of the big bang cosmology and the forgetting of the steady state cosmology.

Despite those solid experimental observations favourable to the big bang cosmology, Fred Hoyle, one of the proponents of the steady state cosmology[1] (with Hermann Bondi and Thomas Gold[2]), and the main opponent of the big bang idea (which, curiously, he baptized)[3], never gave up and continued to fight this idea throughout his life, even after decades of widespread consensus around the big bang cosmology.

Ironically, Hoyle was one of the scientists who most contributed to establish one of the strongest evidence in favor of the big bang cosmology: the great abundance of helium in the universe. He describes how it happened:

> Helium makes up about a quarter of the mass of the visible Universe (...). Could the stars alone be responsible for producing such a huge amount of material? Working on this problem in 1964, R. J. Tayler (...) and I reluctantly decided that the answer was no, by a considerable margin. We found ourselves convinced that all the matter in the Universe must have emerged from a state of high density and high pressure, as George Gamow had always maintained. (Hoyle 1983: 175–176)

However, continuing to reject the idea of a single big bang at the origin of the universe which could be responsible for the formation of chemical elements such as helium, Hoyle suggests the idea of multiple "small big bangs" – explosions associated with quasars – as an alternative to explain the great abundance of the helium:

> Nevertheless, even then the case for the big bang was by no means proven. It seemed that matter had passed through an unusually concentrated state, but this might well have happened within the Universe. The material we see in the stars of our galaxy, and in other galaxies, could have originated in events which did not have to call on an origin of the whole Universe. Quasars (...) seemed to be a pointer in that direction. (Hoyle 1983: 176)
>
> It is now widely believed that variations from quasars have a family relationship to the explosions which sometimes occur at the centres of galaxies, explosions which clearly involve matter at high densities and temperatures, just as in the early moments of the proposed big bang itself. On account of this similarity, as a group they are often referred to as "little big bangs" (...). (Hoyle 1983: 178–179)

1 Hoyle 1948.

2 Bondi and Gold 1948.

3 Cf. Kragh 2013: 15–17.

We can see that Hoyle surrenders to the explosive logic of the big bang cosmology but without surrendering to big bang cosmology, because, in his idea, the small and multiple big bangs scattered throughout the universe dispense a big, single, and primordial big bang, that is, dispense a beginning of the whole universe.

Hoyle also refused to accept the cosmic background radiation discovered by Penzias and Wilson in 1965 as an echo of the big bang, arguing that it was too weak for that and proposing an alternative explanation in line with the steady state theory.

In his alternative explanation, Hoyle even proposed that cosmic background radiation could be produced by microorganisms scattered throughout cosmic space:

> To many astronomers it may seem a fantasy to suggest that microorganisms are responsible for the microwave background, but it is not a fantasy that the required particles exist. One can read about them in any textbook or handbook on bacteria. If bacteria really have the universal presence which astronomical observations suggest, I would consider it likely that they are responsible for the microwave background. (Hoyle 1983: 182–183)

Based on the idea of a never observed "universal presence" of bacteria and on the idea of bacteria as microwave sources, this explanation is doubly and deeply speculative. And the fact is that Hoyle never presented evidences capable of supporting his steady state perspective against the big bang idea, even in the context of the new version of the steady state cosmology that he later proposed with some collaborators, the *quasi-steady state* cosmology[4].

2. *Thematic* reasons of Hoyle's attitude towards the big bang idea

If observations seem to be so strongly in favor of the big bang idea and if, over time, the scientific community has generally forgotten the steady state cosmology, it is justified to ask: why this Hoyle's obstinate rejection of the big bang idea?

According to his own words, Hoyle rejected the idea of the big bang even without first examining it in detail. He recognized this in 1952:

> This big bang idea seemed to me to be unsatisfactory even before detailed examination showed that it leads to serious difficulties. (Hoyle 1952: 94)

Hoyle assumes here with great frankness that his bad impression of the big bang cosmology preceded any scientific analysis, that is, any detailed and

4 Hoyle, Burbidge and Narlikar 1993.

objective analysis of its strengths and weaknesses, whether on the theoretical or the experimental level. It is an attitude of denial without a clear logical or epistemological foundation. In this regard, Jacques Merleau-Ponty, who dedicated to the epistemology of modern cosmology, would say:

> The physicist's preference for antithesis [to the idea of a unique origin of the universe] does not, therefore, have a very solid justification; it is not even epistemological, properly speaking; (...) It comes almost from a mental hygiene. (Merleau-Ponty 1965: 343)

Being outside of any solid theoretical or empirical reasons, usually recognized as being fundamental to validating scientific knowledge, this possible Hoyle's "mental hygiene" seems to refer to some dimension of scientific activity more associated with personal conceptions, even if unconscious or not publicly assumed, and this seems to have a *thematic* setting.

Indeed, the *thematic* analysis proposed by physicist and historian of science Gerald Holton identified in the scientific activity a dimension that, even unconscious or not assumed, is nevertheless very important in the work of scientists, in articulation with the theoretical and experimental dimensions of science. This is the *thematic* dimension, constituted by *themata* – concepts, methods, and hypothesis with a metaphysical, aesthetic, logical or epistemological nature, not only associated to the cultural context, but also to the individual psychology of scientists[5].

Adherence to *themata* is, in general, a lasting fidelity, a "loyal dedication" (Holton 1996a: 159), a personal commitment "remarkably long-lived" (Holton 1975b: 334), which can nurture strong convictions and, in some cases, sustain a will to be right even when there is no good evidence to support that convictions. In these cases, it is a question of *thematic* imagination, which consists, in Holton's words, in "letting a fundamental presupposition – (...) a *thema* – act for a time as a guide in one's own research when there is not yet good proof for it, and sometimes even in the face of seemingly contrary evidence" (Holton 1996a: 96).

According to Holton, a *thema* can function at an individual level as a "guide" (1996a: 96), a "deep conviction" (1996a: 59), an "attachment" (1996a: 158), a "preference" (Holton 1996b: 201), a "preconception" (1996b: 201), a "predisposition" (1996a: 153), a "belief" (1996a: 96; 2005: 145), which can be "obstinate" (1996a: 96), a possible "enchantment" (1996a: 101). As a result, *themata* are elements that establish a "conceptual and even emotional support" (1996a: 159) capable of strongly determining the scientific work, although they are often unconscious for scientists themselves.

In practice, *themata* can be expressed through personal preferences and choices which guide the individual and collective work of scientists. In such work orientation, which can be more emotional than rational,

5 Cf. Holton 1975a: 54–58.

rarely conscious and even more rarely admitted, there could be a strange relationship with truth and falsity: if some objective knowledge (analytically and/or experimentally achieved) does not fit into the *thematic* matrix (a set of personal *themata*) of a scientist, it is quite possible that this scientist is suspicious of that knowledge, seeing it as false or, at least, as incomplete or imperfect, and imagining a truth not yet discovered, but conforms to his personal set of *themata*.

Therefore, the acceptance or rejection of a certain idea may be conditioned by personal *themata* and something which is accepted by a scientist as a proof of that idea may be considered by another scientist, who defends different or even opposites *themata*, as something that may or should be interpreted in other way.

The big bang cosmology is mainly based on a set of *themata* consisting of *life cycle* (with a *beginning*), *evolution*, *finitude*, and *change*. The steady state cosmology is based on opposite *themata*: *continuous existence*, *steady-state*, *infinity*, and *constancy*[6]. The passionate controversy that these cosmological views carried out is part of an old cosmological opposition: the *thematic* opposition between an evolutionary view of the world (traditionally associated to Heraclitus) and a stationary view (traditionally associated to Parmenides)[7].

The relationship with *themata* always has very personal contours and modern cosmology seems to have been no exception. In addition to theoretical issues and experimental observations, adherence to one or the other cosmological current also had to do with adherence to one or the other *thematic* matrix of cosmology, which brought an important personal dimension to the dispute. This was, in fact, what Jacques Merleau-Ponty proved, although without knowing or using the concept of *thema*, when he showed how, in the 20th century, the answers to important cosmological questions that were difficult to answer were driven by the «uncertainties of opinion and belief» (Merleau-Ponty 1965: 107).

According to Helge Kragh, "scientists can have emotional preferences for a theory for all sorts of reasons" (Kragh 1996: 267). And, being a territory especially open to personal preferences, cosmology allowed disputes that occurred, not only theoretically and experimentally, which are traditionally considered scientific, but also, and very intensely, in other domains traditionally and considered not scientific, such as the domain of philosophical (especially aesthetic and metaphysical), religious and even political motivations and options (Kragh 1996: x, 220–232, 237, 249, 251–268).

Other authors, like Yuri Balashov, also showed how "philosophical considerations have been essentially involved in the origin and development

6 Cf. Holton 1975a: 62; Cf. Barbosa 2021: 9.

7 Cf. Holton 1975a: 45; Cf. Merleau-Ponty 1965: 300.

of the steady-state cosmological theory" in explicit and implicit ways (Balashov 1994: 933).

Right in the introduction to the 1948 paper in which he presented his steady-state theory, Hoyle wrote:

> The following work (...) arose from a discussion with Mr T. Gold who remarked that through continuous creation of matter it might be possible to obtain an expanding universe in which the proper density of matter remained constant. This possibility seemed attractive, especially when taken in conjunction with aesthetic objections to the creation of the universe in the remote past. (Hoyle 1948: 372)

It is remarkable how the starting point of a scientific paper is an assumed aesthetic preference, with Hoyle presenting the creation of the universe in the remote past as an aesthetically repulsive idea, as opposed to the idea of continuous creation of matter.

From a *thematic* point of view, we can say that it is an aesthetic preference for the *thema* of continuous existence over the *thema* of life cycle, for the *thema* of steady-state over the *thema* of evolution, for the *thema* of constancy over the *thema* of change.

This preference that Hoyle classifies as aesthetic is linked to a fundamental metaphysical issue related to causality, as Hoyle immediately adds in the following sentence:

> For it is against the spirit of scientific inquiry to regard observable effects as arising from "causes unknown to science", and this in principle is what creation-in-the-past implies. (Hoyle 1948: 372)

This metaphysical issue related to causality was an important question to the founders of the steady-state cosmology. Bondi, whom Hoyle thanked the contributions to his paper of 1948 and to "many discussions on the general problems of cosmology" (Hoyle 1948: 372), also expressed his preference for a theory "in that the problem of the origin of the universe, that is, the problem of creation, is brought within the scope of physical inquiry" (Bondi 1960: 140), which, according to him, was not the case of the big bang theory.

So, we could say that some aesthetic and metaphysical preferences seem to have somehow influenced and even conditioned important scientific ideas about the universe, namely in the construction of the steady-state theory. In the case of Hoyle, we can say that, from a *thematic* point of view, some observations were difficult to fit or even seemed to contradict the personal set of *themata* to which he was faithful, especially the *continuous existence*, which dispenses *a beginning* of the whole universe.

In his personal struggle, Hoyle refused throughout his life the way how some experimental observations were considered decisive in favour of the big bang idea, always looking for alternative explanations for those observations.

Considering the *thematic* features in question and because it was a persistent and long-lived attitude, we can say that the obstinate rejection of the big bang idea is a Hoyle's *thematic* attitude.

It should be recognized that, apparently, this attitude has somewhat changed throughout Hoyle's life. Indeed, recognizing that each of the theories (big bang and steady-state) has its strengths and its weaknesses, Hoyle stated in 1983:

> I have always tried to hold a balanced point of view between several possibilities, whereas some scientists often seem to feel the need to declare themselves unequivocally for one theory or another, rather as if they were supporting a political party or a football club. (Hoyle 1983: 179)

If from the 1940s to the 1960s Hoyle was a radical and fierce opponent to the big bang idea, some decades later he had a more moderate position, materialized in theoretical proposals like the "small big bangs" idea or the quasi steady-state theory (proposed with Burbidge and Narlikar). From a *thematic* point of view, these Hoyle's later ideas conciliate in some way continuous existence with life cycle, steady-state with evolution, constancy with change. However, if we can recognize a softening of attitude towards the big bang idea and its *thematic* matrix, we also must recognize that, in the most essential, Hoyle's fidelity to continuous existence, steady-state and constancy has remained, as we can see in these words of *The Intelligent Universe*, about two decades after the defeat of the steady-state theory and a few pages away from those in which he talks about small big bangs:

> The present orthodox concept of a Universe as a kind of island in time [the temporal finitude advocated by the big bang idea] is all too reminiscent of the erroneous older conception of the Universe of stars as an island in space. The mistake is essentially the same, and it springs not from objective scientific reasons but from sociological and cultural prejudices. (Hoyle 1983: 166–167)

3. *Thematic* preferences, selection effects and biases

This *thematic* attitude expresses a personal *thematic* preference. But scientists' *thematic* preferences are, at bottom, valuations (or overvaluations) of certain aspects to the detriment of others. Indeed, if *themata* can guide the individual and collective work of scientists and, therefore, conditionate the acceptance or rejection of a scientific idea, we can recognize that personal *thematic* preferences of scientists can influence the personal assessment of theoretical or empirical scientific elements: such preferences can compel a scientist to overvalue elements that fit well in his *thematic* matrix and to devalue, or even to ignore, elements that are difficult to explain by a theory belonging to his *thematic* matrix.

Another way of saying this is to recognize that the conditionings that *themata* and *thematic* preferences operate in the ideas and work of a scientist or a scientific community materialize in the form of selection effects and biases.

According to Hoyle, the defenders of the big bang idea devalue alternative possibilities, such as the little big bangs, and ignore the difficulties of the idea they defend:

> Most astronomers and physicists do not like the idea of attributing such great significance to little big bangs, even though there are evidently very many of them. (...) The majority of astronomers and physicists seem to prefer to commit themselves to the idea of the big bang, although by doing so a number of serious difficulties have to be ignored, swept under the rug, difficulties which indeed it may never be possible to resolve from within this particular theory. (Hoyle 1983: 179)

In other words, we could say that, from Hoyle's point of view, the acceptance of the big bang cosmology is, after all, a problem of selection effects and biases concerning some cosmological observations.

However, the defenders of the big bang idea could legitimately say the same thing about Hoyle: his interpretations and alternative explanations about the observations widely regarded as important evidence of the big bang idea "suffer from the same problem" of selection effects and biases. It is a problem easily unconscious and difficult to recognize for those who "suffer" from it, because *thematic* preferences are often unconscious and strongly defended even facing insufficient evidence or contrary evidence.

References

Balashov, Y. (1994). Uniformitarianism in Cosmology: Background and Philosophical Implications of the Steady-State Theory. *Studies in History and Philosophy of Science* 25B: 933–958.

Barbosa, J. (2021). Why Big Bang is so Accepted and Popular: Some Contributions of a *Thematic* Analysis. *Axiomathes*. https://doi.org/10.1007/s10516–021-09533-y.

Bondi, H. and Gold, T. (1948). The Steady-State Theory of the Expanding Universe. *Monthly Notices of Royal Astronomical Society* 108: 252–270.

Bondi, H. (1960). *Cosmology*. Cambridge: Cambridge University Press, 2nd ed.

Holton, G. (1975a). *Thematic Origins of Scientific Thought. Kepler to Einstein*. Cambridge, Massachusetts and London, England: Harvard University Press.

Holton, G. (1975b). On the Role of Themata in Scientific Thought. *Science*, Volume 188, April 25: 328–334.

Holton, G. (1996a). *Einstein, History, and Other Passions*. New York: AIP Press, Woodbury.

Holton, G. (1996b). On the Art of Scientific Imagination. *Daedalus*, Vol. 125, No. 2: 183–208. MIT Press, http://www.jstor.org/stable/20013446.

Holton, G. (2005). *Victory and Vexation in Science – Einstein, Bohr, Heisenberg and Others*. Cambridge, Massachusetts and London, England: Harvard University Press.

Hoyle, F. (1948). A New Model for the Expanding Universe. *Monthly Notices of Royal Astronomical Society* 108: 372–382.

Hoyle, F. (1952). *The Nature of the Universe*. Oxford: Basil Blackwell.

Hoyle, F. (1983). *The Intelligent Universe – A new view of creation and evolution*. London: Dorling Kindersley Book.

Hoyle, F., Burbidge, G. and Narlikar, J. V. (1993). "A quasi-steady state cosmological model with creation of matter". *The Astronomical Journal* 410: 437–457.

Kragh, H. (1996). *Cosmology and Controversy: The Historical Development of Two Theories of the Universe*. Princeton, NJ: Princeton Univ. Press.

Kragh, H. (2013). What's in a name: history and meanings of the term "big bang". https://arxiv.org/ftp/arxiv/papers/1301/1301.0219.pdf.

Merleau-Ponty, J. (1965). *Cosmologie du XXème Siècle. Étude épistémologique et historique des théories de la cosmologie contemporaine*. Paris: Éditions Gallimard.

Milan M. Ćirković
Astronomical Observatory of Belgrade,
Future of Humanity Institute (Oxford University)
mcirkovic@aob.rs

Original Scientific Paper
UDC: 004.8:167/168
004.8:121

WHO'S REALLY AFRAID OF AI? ANTHROPOCENTRIC BIAS AND POSTBIOLOGICAL EVOLUTION

Abstract: *The advent of artificial intelligence (AI) systems has provoked a lot of discussions in both epistemological, bioethical and risk-analytic terms, much of it rather paranoid in nature. Unless one takes an extreme anthropocentric and chronocentric stance, this process can be safely regarded as part and parcel of the sciences of the origin. In this contribution, I would like to suggest that at least four different classes of arguments could be brought forth against the proposition that AI – either human-level or superintelligent – represents in itself a credible existential threat to humanity in either nearby or distant future. Part of the same argumentation is applicable to the general notion of postbiological evolution, which has caused even more unnecessary confusion and uproar in both laymen and philosophical circles. While the due attention should be given to the risks and dangers of the transitional period, there are many reasons why we should openly support and cheer for humanity's transition into the postbiological epoch.*

Keywords: *evolutionary theory, bioethics, artificial life, artificial intelligence, astrobiology, postbiological evolution*

> Whenever the word 'origin' is used, disbelieve everything you are told.
>
> Sir Fred Hoyle

> I think we agree, the past is over.
>
> George W. Bush

1. Introduction: the existential fear of AI

There are people who prefer not to see the link between artificial intelligence (AI) in its multiple forms and the sciences of the origin. This is unfortunate from both theoretical and practical points of view. First, and rather obvious, point is that "sciences of the origin" are not necessarily about the past. Unless one wishes to return to the pre-Darwinian or even pre-Copernican thinking, we should not regard ourselves and the world we are living in as pinnacles of creation and evolution. Quite to the contrary, we have little reason to think ourselves special in any way, including our position in the hierarchy of complex systems, i.e., in the abstract design space. To some extent, this applies to our location in cosmic time as well. Although we have reasons to doubt the unthinking temporal Copernicanism (Ćirković and Balbi 2020), this still does not argue against futures origins events, especially when they concern close future and not the distant cosmological one. All in all, there should be nothing special about those kinds of events/processes we denote as "origins".

If we think about conventional origins as about the emergence/evolution of classes of complex systems, we are thinking about e.g.,

- (our) universe;
- (our) stellar system, i.e., the Milky Way;
- (our) planetary system, i.e., the Solar System;
- (our) planet Earth;
- (terrestrial) life;
- (human) mind;
- (human) language/culture.

Why stop there, however? The answer "because we don't see other classes of complex systems around us" is not only unphilosophical – it is also dangerously parochial and misleading. Not only should we contemplate things we are not in empirical touch with (otherwise, we would have never understood the atomic structure of matter, among other things), we should look more carefully around us. We have at least one process we actually are living through, which certainly qualifies as the origin: the emergence of various kinds of ultracomplex systems which go under the often too narrow label of artificial intelligence (henceforth AI). While the conservative understanding of AI suggests that, even when we achieve artificial *general* intelligence, it will be akin to human intelligence, this is not necessarily so – and might indeed be a dangerous conceit.

Further origins could as well lurk in our future – that is, located along the future temporal axis. We could also have origins which are removed from us in space, rather than in time. The formation of extrasolar planetary systems obviously belong to this category, as does the origin of extraterrestrial life, with the added spice in form of the panspermia hypothesis and other ways of introducing correlations between those origins. Those "other origins" deserve their own philosophical exploration, which is beyond the scope of the present paper. For the moment, we shall focus upon what is actually going on in the real world and what is putting the groundwork for the very next important "origins" – the key role of the emerging AI.

Clearly, human relation to AI is quite a complex and multifaceted topic. Detailed and proper analysis of it will constitute a large fraction of science and philosophy in the rest of this century, and likely in the centuries to follow. This aspect of the sciences of the origins will yet to be discussed and elaborated. Here, we focus on the very narrow aspect of the discussion which, unfortunately, often stunts and impedes the really fruitful debate: correct construals of the concept of "AI risk" and our philosophical attitude toward it.

In the rest of this paper we shall first consider the physical eschatological argument for rejecting chronocentrism (Section 2), before we briefly survey the AI risk scene (Section 3) and outline classes of arguments to the effect that the AI risk is purely instrumental (Section 4). Some provocative tentative conclusions are given in the concluding section.

2. Physical eschatological argument

In order to properly assess our perspective on the very concept of "origins", we need to actively reject the bias of chronocentrism, defined as "the belief that one's own times are paramount" (Fowles 1974). There are multiple ways of doing so and my subjective preference for the arguments following from our global, cosmological knowledge should not occlude the fact that there are other, perhaps better, lines of attack.

One way of showing how outright ridiculous chronocentrism really is, leads through physical eschatology (Ćirković 2003). In brief, this young branch of physical science tells us that the future is way larger than the past. Some of the relevant future timescales are as follows (Adams & Laughlin 1997):

- Sun's future lifetime: 7.7×10^9 yrs;
- star-formation in the Galaxy continues for another $(5–10) \times 10^{10}$ yrs;
- longest-living stars existing today will exist for further 6.5×10^{12} yrs;
- the future lifetime of the Galaxy: cca. 10^{19} yrs.

Take just the last datum as a convenient placeholder. If it is correct – and there is no clear astrophysical reason at present why it should not be – *we are living in the first millionth of the first percent of our stellar system history!*

What's past is prologue. It's very brief, though. The fact that we are living so early in the course of the universal evolution – again, the first 0.000001% – of the lifetime of our stellar system, should give us a pause and perhaps instill some humility regarding our temporal position and epistemic pretensions. Plus, it should arguably motivate a reassessment of our research priorities, which has so far been extremely, staggeringly past-oriented rather than future-oriented.

Consider a man who has lived 80 years (not so rare these days, at least in developed countries) and all variables of his life. He might have been a sailor, a scientist, a criminal, a mailman, or whatever. He might have never married or might have been married three times with children. He might have been a passionate angler or biker or ecological activist or an atheist. Etc. etc. etc. Now, consider how much of this information on a rich and complex life could you get by observing the first *5 minutes of his life*. Certainly not much. That is exactly the fraction of the total lifetime of the Galaxy we study as history this far. Therefore, it is exactly within this minuscule fraction that we have defined the origins.

Of course, the richness of the phenomena is not a linear function of the time elapsed. There are clear physical reasons to expect that the universal increase in entropy mandated by the Second Law of thermodynamics will make far-future epochs much slower as far as the evolution of various systems is concerned. Therefore, long future epochs of gradual decay of structures are much less interesting than the early epochs like the one we are living in. But even if we accept this effect of diminishing returns, interest-wise, the future is still so much larger than the past, even if we neglect plethora of very slow processes which have been unimportant so far but will become important in the future. We also need to take into account more bizarre and speculative cosmological futures we are currently vaguely aware of, such as the possibility of future vacuum phase transition at very low temperatures (e.g., Kusenko and Langacker 1997). In addition to all this, the actions of intelligent and intentional beings bring an entirely new dimension to our studies of the future, making some of our favorite inductive strategies largely irrelevant (Walker and Ćirković 2021). In fact, it is likely that only the development of AI will enable us to make sufficiently rich and detailed models of future evolution of complex systems including future human/posthuman societies; thus, the topic has both aspects of inevitability and self-reference.

3. AI fear-mongering: philosophy and marketing

If we allow for future origins, it is only natural to conclude that the first impending of such events is the emergence of AI and the accompanying onset of *postbiological evolution*. It is immaterial for the present argument which of the many construals of postbiological is the best or even acceptable; as a placeholder, we may refer to the convergence of nanotechnology, biotechnology, information technology and cognitive science (NBIC) to improve human performance (e.g., Canton 2004; Bainbridge and Roco 2006). So, while specifics may vary, we may expect that the NBIC convergence describes satisfactorily well human/terrestrial approach to the postbiological era. One important note must be made here: the locution “evolution” too often connotes blind Darwinian processes. While it may certainly be true that the classic Darwinian selection processes will continue to be in play among cybernetic posthumans or within the digital substrate, there are all reasons to expect the *mechanisms* of change to be dramatically different. If anything, postbiological evolution as a macroevolutionary trend could be expected to be Lamarckian, where intelligent agents consciously and deliberately choose their own evolutionary course, starting from an inherited state.

That would be the end of the discussion... if not for already more than a century of anthropocentric fear-mongering, starting with Karel Čapek's magnificent drama *R.U.R.* back in 1921. The spectre of "machine revolt" as the true *Phantom Menace* of the third millennium is certainly a global cultural phenomenon. One could trace its origins to myths and remythologized ideas such as Golems, Frankenstein monsters and in the modern context – robots. Moreover, the current phantasm implies not only that *Eeeevil Machines* are threatening us with existential disaster, so we ("obviously") have to "do something". If primitive bacteria in some pond 2 billion years BC had been able to think, they would have thought something along the lines *look at those pesky eukaryotes, they're threatening us, we Have To Do Something!* And some philosophers happily joined in the fear-mongering festivities, on a wide front from Robert Sparrow and Phil Torres to Toby Ord and Nick Bostrom.

This does not mean that there is no AI risk; the most comprehensive serious meta-survey up to 2014 is given by Sotala and Yampolskiy (2014). We need to be realistic about its magnitude and structure, however. There is no reason – at least not *a priori* – to expect the AI risk to be outside of the domain of the classical risk analysis, or to somehow circumvent the usual research approaches relevant for its assessment and management (cf. Scheessele 2021).

In that spirit, consider the distinction between substantive and instrumental risk in risk analysis. This is a distinction applicable to the anthropogenic/technogenic risks, such as the risks from nuclear weapons, synthetic biology or nanotechnology. Clearly, there are technologies which are carriers of *substantive risk* in and of themselves; nuclear weapons are almost perfect example in this regard. While some contrived peaceful applications of nuclear weaponry have historically been suggested (e.g., canal-digging), usually by people having vested interest in improving their public image, even that was not really assumed to be risk-free.[1] So, the concept of substantive risk applies to any kind of use. In contrast, instrumental risk is associated with the technologies which have legitimate risk-free uses. Instrumental risk goes under the vernacular terms such as „misuse", „abuse" and so on. Its inherent property is that it requires malevolent actors – and its key feature is that it removes nonsensical labeling of entire technologies as inherently risky or dangerous. One can obviously misuse all technologies, starting with the fire and the wheel, but it is on an entirely different level than the threat of nuclear weapons or synthetic pathogens.

1 Obviously, there is no risk-free activity *sensu stricto*. However, there is a clear and rather discrete difference between digging of a canal using shovels and digging of a canal using nuclear warheads, discernible even to philosophers.

4. Realism about AI risks and the arguments for it

At long last we come to the main thesis of this essay, which highlights the key feature of the incoming origins process: *the real AI risk is 100% instrumental risk*. Obviously, the qualification "real" is necessary here, since we need to assess what is truly at play here: clearly not what is sometimes hysterically *claimed* as the AI risk, but what can be realistically expected from the deployment of specific AI technologies. We should think about wild cards, of course: currently unimagined scenarios which increase the risk. We should not, however, allow for those wild cards to dominate the discourse.

There are several classes of arguments for the main thesis, which we shall outline in the following subsections.

4.1. Postbiological evolution

The most important argument comes from evolutionary biology and it is rather surprising that it has been very little discussed in the literature. Since the early days of the Modern Synthesis (in particular the work of J. B. S. Haldane which prefigured later results of George Williams and William Hamilton), if not from Darwin's *Descent of Man* in 1871, it was obvious that human behavioural traits are evolutionary grounded (Tattersall 2009). Therefore, as a matter of principle, the transition to postbiological phase obviates most, if not all, biological motivations. The very definition of ecology and the relevant ecological needs and imperatives changes, leading to significant changes in other fields which have been traditionally linked to the evolutionary processes.

As an example, the imperative for filling the complete ecological niche in order to maximize one's survival chances and decrease the amount of biotic competition is an essentially biological part of motivation for any species, including present-day humans. It would be hard to deny that this circumstance has played a significant role in colonization of the surface of the Earth. But expanding and filling ecological niches are not intrinsic properties of life or intelligence – they are just consequences of the predominant evolutionary mechanism, i.e., natural selection (e.g., Tinbergen 1968; Wilson 1978; Trivers 1985; Heying and Weinstein 2021). It seems logically possible to imagine a situation in which some other mechanism of evolutionary change, like the Lamarckian inheritance or genetic drift, could dominate and prompt different types of behaviour. The same applies for the desire to procreate, leave many children and enable more competitive transmission of one's genes to future generation is linked with the very basics of the Darwinian evolution. Postbiological civilization is quite unlikely to retain anything like the genetic lottery when the creation of new generations is concerned. In addition, the easiness of producing and retaining copies of postbiological organisms in the digital substrate is likely to dramatically change the meaning of terms such as "maturation", "adulthood", "parenthood", "kin", etc. Thus, we need to make an additional step symbolically represented as the analogy:

biological evolution → postbiological evolution
sociobiology → "post-sociobiology"
biologically rooted behaviors → behaviors based on postbiological factors.

Clearly, we need much more research and thinking in order to establish what exactly could "post-sociobiology" be, but as a provocation we may suppose that it will deal with "stable ingredients" (to use the expression of Arnold Toynbee; see Toynbee 1966) of postbiological development. In the case of (post)human evolution, one may argue that this will encompass "posthuman nature" in the same manner as authors like Fukuyama (2002) invoke "human nature" as an explanatory device. It is very hard to conceptualize such a dramatic change – but we still ought to think as hard as possible about its outcomes since, among other things, some very early decisions can have long-reaching consequences (Bostrom 2003b).

In brief, the biological evolutionary baggage such as aggression, teritoriality, tribalism, possessiveness, etc. etc., probably responsible for 99% of problematic features of human culture and history will be getting *weaker*, not stronger, with the advent of AI. No biological evolution whatsoever means no baggage! And, clearly, a postbiological-based ethics, whatever that is, should substitute for the current, biological-based ethics.[2]

4.2. The design space non-ergodicity

It is very reasonable to assume that postbiological design space is much, much bigger than the biological morphospace; after all, a part of the motivation for transition to the postbiological realm is exactly to mitigate the weaknesses and insufficiencies of the biological. If a cyborg with transplanted human brain or an uploaded mind can survive and thrive on the lunar surface, for example, this means that a new wide subspace of the postbiological design space will have opened, with no analogs in the space of biological forms. (Assumed, of course, that there are no biological forms which could survive on the lunar surface, which – while very plausible – cannot yet be regarded as definitely proven.)

This difference in generalized space size cannot come for free; the no free lunch axiom applies here as well. An important consequence of the difference in size is that trajectories are much more complex to navigate within larger

2 For those worrying excessively about the extinction of humanity as a consequence of "new and better" model of intelligent beings coming online, we should note that even within *purely biological* evolution, "nature red in tooth and claw", the emergence of new taxa (even high ones!) did *not* mean the removal of the old ones. Thus, we still have the sponges (*Porifera*), the comb jellies (*Ctenophora*), and other ancient phyla and other taxa.

than within smaller space. If one wishes to arrive at a desired point – or a desired region of the parameter space – one has to steer much more carefully in a larger space than in a smaller one. That much is a clear mathematical truth, not something abstract, hypothetical, or conjectural.

What is somewhat conjectural, but nevertheless highly plausible, is that taking evolution in our own hands is going to be extremely complex affair, which can hardly be calculated without a new, powerful numerical models. And this applies to any and all aspects of the transition, including the ecological ones. Postbiological minds are simply likelier to navigate the complex sustainability trajectories. (Which, parenthetically, apply to the extraterrestrial civilizations and the crossroads between astrobiology and the futures studies; cf. Frank and Sullivan 2014.)

4.3. "Wisdom of the ages"

It is often claimed, especially among conservative thinkers, that while *intelligence* of AI systems – or indeed many fellow humans – is measurable and indubitable, their *wisdom* is at the very best limited (cf. Fukuyama 2002). Now, this argument would have more merit if it were accompanied by an operational definition of wisdom. Unfortunately, this is usually not the case, so one has to improvise irrespective of whether one accepts the argument or not.

Suppose that we construe wisdom as a set of beneficial insights and statements tested by history and ages; beneficial, presumably, for both individual and the community she lives in. Clearly, in the human experience so far, it is very, very difficult to develop new or expand the existing wisdom, since the human history and the history of each culture and community, as well as life history of any individual are unique and do not allow for controlled experiments of the kind we have in most physical sciences. That very circumstance, that *new wisdom is hard to come by*, can be regarded as a major (or even *the* major) tenet of conservatism and perhaps the reason why conservatism is an anthropological constant over all ages and all cultures.

AI brings an entirely novel element in this perennial story. History is not perfectly immutable set of records any more. One can reasonably experiment with history via detailed simulations, potentially with arbitrarily wide scope and high resolution. The simulations of the kind envisioned here are extremely complex and similar to the concept of the "ancestor simulations" of Bostrom (2003a). There are all kinds of auxiliary problems and issues related to the existence of such simulations, but they are immaterial for the present goal. What is crucial is that such historical simulations enable experimenting with various evolutionary trajectories – and hence *maximizing wisdom*!

Not only that: the process of optimization could, in principle, be sped up. Postbiological "effective age" is a function of the processor clock rate. Therefore, experimenting with history will enable much *quicker* acquiring of wisdom (as operationally defined above). Multiple internal simulations will likely bring insight into all kinds of moral dilemmas – which is the key ingredient in any kind of construal of wisdom. All in all, it could be persuasively argued that it is in fact likelier that future advanced AIs will be endowed by true wisdom than what we could say, for example, for our present political and cultural leaders.

4.4. Semantic dysfunction of "nature" vs. "artifice"

Ever since Turing proposed "a heretical theory" (Turing 1970) the specter of the "artificial Other" has been haunting human thinking. Insofar we accept physicalism about minds, the distinction between "natural" and "artificial" minds becomes blurry at best. There is no "natural" trade mark which could be stamped on a mental state if it is equivalent to a physical state of matter, governed by the universal laws of physics. There is no difference between properties of the water molecule produced in a laboratory by joining two hydrogen atoms with one oxygen atom and the H_2O molecule taken from, say, a river. Historically contingent origin of even those hydrogen and oxygen atoms themselves (hydrogen being produced in the Big Bang nucleosynthesis, oxygen mostly in low– and intermediate-mass Population II stars) does not impact their properties here and now. This applies, rather obviously, to their *yet undiscovered* functional properties, for instance, their participation in a not-yet-synthesized complex pharmaceutical compound. In more general terms, physicalism at least vaguely implies insensitivity to historically contingent initial conditions (even without taking into account specific physical *processes* like nonlinear dynamical evolution, which erase information on the initial states). *Why would, then, anyone assume the difference between properties of a mind produced in an AI lab and properties of a mind evolved by biological evolution?*

Whoever accepts physicalism and is still afraid of AI *in the substantive sense* is a bit of a hypocrite. Worrying that something especially bad or corrupt will happen because AI is labeled as artificial is dangerously similar to anthropocentric justification of crimes and corruption perpetrated by those minds labeled as natural.[3] One could speculate that the erroneous demarcation between the „natural" and the „artificial" is an evolutionary consequence in itself, as are most of other persistent irrational beliefs (e.g., religious ones). This should not discourage us, however, to problematize and criticize it at any point. In fact, this kind of argument is the true point of both Čapek's *R.U.R.* and Ridley Scott's celebrated movie *Blade Runner*: discrimination against robots or replicants on the basis of their allegedly "unnatural" origins is a dangerous anthropo-chauvinistic nonsense, which cannot end well.[4]

3 Obviously, non-physicalism about minds open a host of other possibilities, including those in which non-physical ingredient prevent artificial minds from being "true" minds, which is perhaps the closest to the folk understanding of AI in most of the world.

4 Notice the absent argument in all of the above, which has been sometime heard in transhumanist circles: that the AI risk is not really a risk, since even the adverse outcome will have net positive value. (This has been circulated much informally, but rarely in

5. Conclusions

We have so far, in spite of much hand-waiving, obtained not a single strong reason to believe that the AI risk is anything but instrumental. This especially pertains to the alarmist positions which argues that the instrumental fraction of the AI risk tends to zero or some very small quantity. Those positions betray self-serving, antievolutionary and anthropocentric attitude of their proponents and tell us little about the reality (or otherwise) of true AI, be it of human or superhuman level.

Philosophical failure to put the relevant concepts into the truly Copernican, non-anthropocentric, non-chronocentric evolutionary context should worry us enormously. It shows how empty our oft-verbalized proclamation of inclusiveness, universality, and brotherhood of the mind are in reality. In the times when astrobiologists prepare to search in detail for biosignatures on extrasolar planets, and the new generation of search for technosignatures (previously known as SETI), it is indeed deplorable that we are seemingly not able to adequately conceptualize other minds, *even if they are of our own making*. What to expect, then, of future contact with the independently-evolved extraterrestrial life and intelligence? If fear and paranoia dominate our thinking about minds we have built ourselves, what hope is there that we will be able to empathize with complex beings we have no phylogenetic relation to whatsoever? Even more, won't such fear and paranoia lead us to totalitarian oppression of our own kind to preserve "essential human dignity", "human essence", "neural purity" (cf. Reynolds 2006), and such demagogic, ideology-laden nonsense?

All what has been said should not be construed as stating that the transition to postbiological will be a cozy ride. On the contrary, there are all kinds of indications that – similar to the previous global revolutions such as the agricultural and the industrial revolution – it will be accompanied by much upheaval, societal chaos, and risk in general (Bostrom 2014; Ćirković 2017; Ord 2020). It is paramount, however, to understand where does that risk mostly come from: anthropocentric human institutions, from political and judicial systems, via economic structures, all the way to media, institutions of culture, etc. For instance, it has already become obvious that automatization and robotization are bringing about a surge in unemployment, as well as gradual obsolescence of many cherished human professions and trades. As we

print; perhaps the closest are formulations of Moravec 1988, 1998.) This is nonsense not only because the definition of risk can never be given entirely objectively which is a foundational principle of risk analysis since its emergence as a discipline (e.g., Byrd and Cothern 2000), and always refers to a set of subjects. It also manifests a pathology which could be justly dubbed the "utilitarian Stockholm syndrome": incorporating parts of the ethical mindset and values of one's captors on the basis of allegedly correct utilitarian calculus. Detailing fallacies inherent in this attitude would require a separate paper and is not of great interest for us here.

have seen in recent years, this has already caused social unrest in many of the affluent societies on the planet; we need to be wary and extremely cautious regarding further stages of this process, to avoid sliding into totalitarianism. The advent of the true general AI is likely to bring about now unconceived problems, true "unknown unknowns" we need to be agile in perceiving and analyzing.

All those, however, stem from the instrumental use of AI and related technologies and do not present us with something necessarily dangerous in and of itself. While putting AI in hands of dictators, supreme religious leaders, unscrupulous spymasters, mafia bosses, white-collar criminals, or similar unsavory characters is certainly a very, very bad idea, the badness is just a function of negative value assessment of those characters. Putting *anything* of value in hands of dictators et al., is a very bad idea as well; this is valid irrespectively of the specific nature of AI. While AI can magnify the evil-doing of evildoers by a large factor, the root cause is clearly the existence of evildoers among humans. While all this is embarrassingly trivial, the fact that many authors and even more people among the general public find it easy to slip into the "AI is evildoer itself" mode of *2001: A Space Odyssey* (1968) is quite disturbing.

For the very end, whoever finds scientific books and papers boring should read some *belles-lettres* about the postbiological evolution. Almost 30 years ago, the great Australian master of hard science fiction, Greg Egan, has published a majestic novel, *Permutation City*, which does not only deal with the postbiological evolution, but whose protagonists – at least in the second part of the book – are indeed cellular automata (Egan 1994)! Egan's novel not only highlights the capacity of such a transformative technology, but brings a dire warning of problems we may have with our own intellectual progeny – if we continue along the anthropocentric road. It is a common knowledge how difficult successful science and technology public outreach is. Therefore, one should use and celebrate each and any lump of gold encountered along the way.

Acknowledgements. The author is happy to express gratitude to the organizers of the *Sciences of the Origin: The Challenges of Selection Effects and Biases* conference (June 3–5, 2021), to an anonymous reviewer for useful comments, as well as to Vojin Rakić, Barry Dainton, Mark Walker, Srdja Janković, and Ana Katić for helpful discussions.

References

2001: A Space Odyssey [1968] 1997. DVD, dir. S. Kubrick, Metro-Goldwyn-Mayer, Beverly Hills, California.

Adams, F. C. and Laughlin, G. 1997, "A dying universe: the long-term fate and evolution of astrophysical objects," *Rev. Mod. Phys.* **69**, 337–372.

Bainbridge, W. S. and Roco, M. C. 2006, *Managing nano-bio-info-cogno innovations* (Springer, Dordrecht).

Blade Runner [1982] 2007. DVD, dir. R. Scott, Warner Bros., Burbank, California.

Bostrom, N. 2003a, "Are You Living in a Computer Simulation?" *Philosophical Quarterly* **53**, 243–55.

Bostrom, N. 2003b, "Astronomical Waste: The Opportunity Cost of Delayed Technological Development," *Utilitas* 15, 308–314.

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

Bostrom, N. and Ćirković, M. M. (eds.) 2008, *Global Catastrophic Risks* (Oxford University Press, Oxford).

Byrd III, D. M. and Cothern, R. C. 2000, *Introduction to risk analysis: a systematic approach to science-based decision making* (Rowman & Littlefield, Lanham).

Canton, J. 2004, "Designing the future: NBIC technologies and human performance enhancement," *Annals of the New York Academy of Sciences* **1013**, 186–198.

Ćirković, M. M. 2003, "Resource Letter PEs-1: Physical Eschatology," *Am. J. Phys.* **71**, 122–133.

Ćirković, M. M. 2017, "Enhancing a person, enhancing a civilization: a research program at the intersection of bioethics, future studies, and astrobiology," *Cambridge Quarterly of Healthcare Ethics* **26**, 459–468.

Ćirković, M. M. and Balbi, A. 2020, "Copernicanism and the Typicality in Time," *International Journal of Astrobiology* **19**, 101–109.

Dick, S. J. 2003, "Cultural Evolution, the Postbiological Universe and SETI," *Int. J. Astrobiology* **2**, 65–74.

Egan, G. 1994, *Permutation City* (Orion/Millennium, London).

Fowles, J. 1974, "On chronocentrism," *Futures* **6**, 65–68.

Frank, A. and Sullivan, W., 2014. Sustainability and the astrobiological perspective: framing human futures in a planetary context. *Anthropocene*, *5*, pp.32–41.

Fukuyama, F. 2002, *Our Posthuman Future: Consequences of the Biotechnology Revolution* (Farrar, Straus & Giroux, New York).

Heying, H. and Weinstein, B. 2021, *A Hunter-Gatherer's Guide to the 21st Century: Evolution and the Challenges of Modern Life* (Portfolio, London).

Janković, S., Katić, A., and Ćirković, M. M. 2022, "Gaia as Solaris: An Alternative Default Evolutionary Trajectory," *Origins of Life and Evolution of Biospheres*, in press (DOI: 10.1007/s11084–022–09619–6).

Kusenko, A. and Langacker, P. 1997, "Is the vacuum stable?" *Physics Letters B* **391**, 29–33.

Moravec, H., 1988. *Mind children: The future of robot and human intelligence*. Harvard University Press.

Moravec, H. 1998, *Robot: Mere Machine to Transcendent Mind* (Oxford University Press, Oxford).

Ord, T. 2020, *The precipice: Existential risk and the future of humanity* (Hachette Books, New York).

Reynolds, A. 2006, *Galactic North* (Gollancz, London).

Scheessele, M. R. 2021, "The hard limit on human nonanthropocentrism," *AI & SOCIETY* **37**, 49–65.

Sotala, K. and Yampolskiy, R. V. 2014, "Responses to catastrophic AGI risk: a survey," *Physica Scripta* **90**, 018001.

Tattersall, I. 2009, "Charles Darwin and human evolution," *Evolution: Education and Outreach* **2**, 28–34.

Tinbergen, N. 1968, "On war and peace in animals and man," *Science* **160**, 1411–1418.

Toynbee, A. 1966, *Change and Habit* (Oxford University Press, Oxford).

Trivers, R. L. 1985, *The Social Evolution* (Benjamin Cummings, Menlo Park).

Turing, A. 1970, "Intelligent Machinery, A Heretical Theory: Report, National Physical Laboratory," *Machine Intelligence* **5**, 3–23.

Walker, M. and Ćirković, M. M. 2021, "Induction Ain't What It Used to Be: Skepticism About the Future of Induction," *Journal of Ethics and Emerging Technologies* **30**(1).

Wilson, E. O. 1978, *On Human Nature* (Harvard University Press, Cambridge).

Andrej Jeftić
Institute for Philosophy and Social Theory
University of Belgrade
andrej.jeftic@ifdt.bg.ac.rs

Original Scientific Paper
UDC: 575.85:167/168
575.85 Гулд С. Џ.
575.85 Конвеј Морис С.

CONTINGENCY AND CONVERGENCE IN THE THEORY OF EVOLUTION: STEPHEN JAY GOULD VS. SIMON CONWAY MORRIS

Abstract: *Debating the interpretation of the Burgess Shale fossil records, Stephen Jay Gould and Simon Conway Morris have formulated two conflicting theses regarding the nature of evolutionary processes. While Gould argued that evolution is essentially a contingent process whose outcomes are unpredictable, Conway Morris claimed that the omnipresence of convergence testifies that it is in fact deterministic, leading to predictable and inevitable outcomes. Their theses have been extensively researched from various perspectives. However, a systematic parallel analysis of the core arguments each of them offered in support of their thesis has been lacking. I argue Conway Morris has successfully exposed the core weaknesses of Gould's thesis and offered a comprehensive account in favor of the major role of convergence in evolutionary history. On the other hand, I will point out some of the weak points in the latter's arguments supporting the deterministic view of life's evolution. Although Conway Morris has been more successful in arguing for the deterministic nature of the evolutionary processes, both theses could be improved if their shortcomings are taken into consideration.*

Keywords: *evolution, contingency, convergence, Stephen Jay Gould, Simon Conway Morris"".*

Introduction

This paper analyzes arguments in favor of the contingent or convergent character of the historical path of life's evolution on Earth as they were presented by two figures: Stephen Jay Gould and Simon Conway Morris. The role of contingency (Wong 2020; Hopster 2017; Ramsey and Pence 2016; Ćirković 2014) and that of convergence (Currie 2012; Losos 2011; Harmon et al. 2005; Futuyma 2010; Stayton 2015a, 2015b) in evolution has been discussed extensively in the literature. The way these factors were used to develop the *contingency thesis* as promoted by S. J. Gould (Blount, Lenski, and Losos 2018; McConwell 2019; Turner 2011; Dresow 2019; Beatty 2006a, 2006b) and the *convergence thesis* as promoted by S. Conway Morris, has also been researched. In addition, the debate between the two has been approached from different perspectives (Baron 2011; Bowler 1998; Baron 2009; Mcshea 1993). However, what is lacking is a parallel philosophical analysis of the core arguments each of them offered in support of their thesis. This paper aims to fill that gap.

The task is not to present how the historical debate on these matters occurred, culminating in the clash of (re)interpretations of the Burgess Shale findings, which took place in the '80s and '90s. Instead, the focus will be on systematic analysis of the arguments favoring one position over another. At the same time, their history will be visited only sporadically. Finally, these arguments will be analyzed from the philosophical perspective and not that of paleontology, microbiology, cladistics, genetics, and other sciences. However, their results will inevitably be considered to some degree.

The topic before us more or less came into existence when Gould took upon himself to systematize for a broader public the research done by a group of scientists from Cambridge University, which consisted of Conway Morris, Harry Whittington, and Dereck Briggs. The group of three (we will call them the Cambridge Team) did entirely new research on the Burgess Shale, which Charles Walcott discovered in 1909. Whittington was not satisfied with Walcott's classification and description of the fossils (Whittington 1985, xiv), so he embraced the offer to reevaluate the fossils and the shale itself. The results of their work were published in a series of studies and articles published between 1971 and 1985. Gould recognized the importance of these findings and spectacularly presented them in his book *Life's Wonderful Story: The Burgess Shale and the Nature of History* (1989). Gould thought that these new insights had profoundly changed the picture of the history of life and the processes that affect the evolution of organisms. In this context, he formulated a thought experiment that will become well known, i.e., a replay of the tape of life. Paradoxically enough, he did so by elaborating the ideas first brought up by a man who would later become the most prominent opponent of the results of Gould's thought experiment – Simon Conway Morris.

I will first offer a review of Gould's thesis about the *contingent* nature of evolution and of the arguments by which he had supported it. Then I will explain Conway Morris' view on the role of *convergence* in evolutionary processes. In the end, I will confront their standpoints to establish their strengths and weaknesses. The following analysis will be predominantly based upon two monographs published by these authors in which they offered the most comprehensive elaboration of their respective theses: Gould's *Life's Wonderful Story* and Conway Morris' *Life's Solution: Inevitable Humans in a Lonely Universe* (2003).

Stephen Jay Gould's replay of the tape of life

In a nutshell, Gould formulates his thesis in the form of a thought experiment in the following manner:

> Wind back the tape of life to the early days of the Burgess Shale; let it play again from an identical starting point, and the chance becomes

> vanishingly small that anything like human intelligence would grace the replay (Gould 1990, 14).

According to Gould, the nature of the historical trajectory of evolution is in itself contingent. This term is central between the two terms that express the opposite viewpoints: necessity and chance. According to the first one, the history of life on Earth is essentially deterministic. Therefore, any replay of the life's tape would yield the same results. On the other hand, according to the opposite standpoint, history evolves through a line of mutually independent random events. What Gould is offering is a third solution: contingency. In his view, the theory of evolution is a historical science, just like geology and cosmology. Therefore, it attempts to offer an explanation of events which are essentially contingent. This means that its

> ...historical explanation does not rest on direct deductions from laws of nature but on an unpredictable sequence of antecedent states, where any major change in any step of the sequence would have altered the final result. This final result is therefore dependent, or contingent, upon everything that came before – the unerasable and determining signature of history (Gould 1990, 283).

The concept of contingency that Gould opts for signifies a viewpoint that event E is a necessary consequence of a line of events: D, C, B, and A. However, these events might not have occurred at all, or they might have occurred in a different manner, which would have resulted in the non-existence or essentially different shape of the event E. Therefore, E is a neither necessary nor random event: it is a contingent event (Gould 1990, 51).

Since evolution is a contingent process, its outcomes are neither random nor determined from the very beginning. These include the existence of humans, but they could have been entirely different, and they would have indeed been such if we would replay life's tape.

What is the foundation of Gould's thesis? We could summarize Gould's argument in favor of the contingent nature of evolution in the following way.

1. Reinterpretation of the Burgess Shale has shown that our common notion of evolution as a cone of increasing diversity and a ladder of progress is wrong. The number of body plans was maximal at the beginning (maximal initial proliferation), and with the passing of time, some of them survived while others became extinct. The cone is upside-down, *diversity* has decreased through time, and the tree of life is not spreading but narrowing, like a Christmas tree.
2. Selection of the surviving body plans was not a deterministic process. Their *decimation* was most probably a consequence of a lottery. Mechanisms of natural selection had not played a key role in it; this process was *random*.

3. *Prediction* of these outcomes would have been impossible for a hypothetical paleontologist who would be granted the opportunity to have a glimpse of living Burgess fauna. This claim is also supported by the phenomena of *massive extinctions,* which are also random events themselves.
4. Since none of the critical points in the history of life's evolution could have been anticipated, it would have been impossible to predict which body plans of a higher taxonomic rank would survive or become dominant. The evolution of the eukaryotic cell, the disappearance of the Ediacara fauna, the development of the terrestrial vertebrates conditioned by the contingent evolution of a particular skeleton among fish, the adaptive radiation of mammals after the extinction of dinosaurs, and the survival of a tiny African population of *Homo Erectus* – are all events that could have never happened. Consequently, it would have been impossible to predict the existence of conscious beings, i.e. anatomically modern humans even from the middle of Pleistocene, let alone the beginning of Cambrian.
5. The history of life results from unpredictable, partly random, and contingent processes whose outcomes bear the same features of unpredictability and contingency. Therefore, the replay of this history, that is, the replay of life's tape, would each time produce different results. The chance that among those results one could find the existence of the human species is close to none.

Gould's book *Life's Wonderful Story* can be seen as his critique of the adaptationist program. One of the prominent figures of this evolutionary program is the second author that we are dealing with here, Simon Conway Morris.

Simon Conway Morris' Inevitability of Humans

Gould celebrated the achievements of the Cambridge Team throughout his book, especially Conway Morris, "the young and radical man of ideas who developed a revolutionary interpretation and dragged everyone else along" (Gould 1990, 157). Although he was the first to introduce the thought experiment of the replay of life's tape, Conway Morris changed his mind later on. Gould was surprised that he never mentioned that he ever agreed with Gould's standpoint, but Baron (Baron 2011) has convincingly shown that they have always had different starting theoretical frameworks. Conway Morris explained his disagreement with Gould in a book published almost a decade later than Gould's. In *The Crucible of Creation,* he strongly opposed Gould's interpretation of the Burgess fauna and proposed a thesis that evolution is not a contingent but rather a *convergent* process. He developed this thesis a couple of years later in his *Life's Solution*. Contrary to Gould, he states that

> evolution is the outcome of stochastic and deterministic processes. As such, should the tape of life be replayed, undoubtedly, there would be many differences, but there would also be a very significant number of similarities (Conway Morris 2003, 272–73).

He rejects Gould's thesis and the results of the thought experiment, stating:

> Rerun the tape of life as often as you like, and the end result will be much the same. On Earth, it happens to be humans (Conway Morris 2003, 282).

His starting point for the rejection of Gould's contingency thesis is the all-presence of convergence. Broadly defined, it denotes the independent evolution of similar features in multiple species or clades (Losos 2011). For example, birds, bats, and butterflies all have wings that have evolved independently within different clades and were not inherited from a common ancestor. Although a well-documented phenomenon, its significance has not been completely acknowledged by scientists, Conway Morris notes. He argues that it does not represent an exception but rather the rule itself which tells us something important about the nature of evolution. What it tells us is that evolution constantly develops the same solutions to tasks posed by the environment. Therefore, these solutions are inevitable and consequently predictable. One of those solutions is the existence of humans which is declared by Conway Morris to be inevitable, once the proper conditions, like the ones that are present on Earth only, are all set. Humans are inevitable, but they evolved only once in the whole universe.

What are the arguments he offers in support of this thesis? They could be summarized in the following way:

1. *Convergence* is an all-pervasive phenomenon. It is present on the level of (1) molecules, (2) body plans, (3) the structure of the body systems, (4) organs, (5) higher characteristics, and (6) behavior. Particular proteins, e.g. hemoglobin, have evolved independently several times while certain amino acids in the protein chain have also been replaced independently which enabled the emergence of color vision in different clades. Secondly, animals that share the same environment, e.g. those living underground do, tend to develop similar *body plans*. Thirdly, certain *systems* have evolved in different clades. For example, this is the case with the hearing system as well as the very basis of the nervous system, the sodium canal. Fourthly, different groups of animals have developed the same *organs*, such as camera eye or wings. Fifthly, *higher biological characteristics*, such as intelligence, have developed separately among fish, arthropods, and mammals. Finally, specific *forms of behavior* appear in distant clades. We find social organization among insects and primates, while agriculture and even military arming is found both among ants and humans.

2. Presence of convergence on all previously mentioned levels, proves that life possesses particular chemical, physical, historical, and ecological *constraints* imposed by the environment. Chemical constraints are imposed by the chemical organization of molecules. Secondly, physical constraints limit the spectrum of possible biological characteristics. Thirdly, there are constraints imposed by a particular evolutionary history of a species that determine its possible evolutionary future. Finally, specific constraints are set by different environments, be it water or underground, which determine what characteristics are realistically possible. This means that the number of solutions to the tasks posed by the environment is limited, and consequently so is the spectrum of (realistically) possible life forms. The 'hyperspace' of possible protein structures, anatomic plans, and combinations of biological characteristics (morphospace) is enormous, and yet only a fraction of it (0.1%) is occupied by life on Earth. However, this hyperspace is occupied *entirely* in terms of real and not hypothetical possibilities. These are determined by what is structurally possible and what holds a positive adaptive advantage.
3. The best solutions to the tasks presented by the environment represent adaptive peaks which can be considered as 'attractors' of functionality that navigate the evolutionary process, directing it towards themselves. If a particular solution is especially good, be it hemoglobin, wings, or fission-fusion societies, it will repeatedly be achieved through evolution.
4. The existence of attractors speaks in favor of evolution's *directionality*. It is a process that develops in a particular direction, more precisely, in a limited number of directions. *Trends* and *progress* are not absent from the evolutionary process. Trends signify the fact that certain traits keep evolving independently. Progress itself should not be understood as a gradual increase in the number of body forms or species but as an increase in complexity. Conway Morris understands complexity to be a phenomenon that is hard to define, although everybody has a tacit notion of what it refers to. For him, it simply denotes the fact that "Once there were bacteria, now there is New York" (Conway Morris 2013, 136). Since it leads to a gradual increase in complexity, evolution is a teleological process.
5. Directionality of evolution confirms the principle of *inherence*: if certain building blocks are present, each organism will use them to build the best solution to a given problem, i.e. it will reach the attractor or at least evolve towards it.
6. Because evolution is a directional and progressive process, it would be possible to *predict* its outcomes with greater probability. One of those outcomes, and for Conway Morris, the most important one, is *humans*. Humans have been inevitable at least since the Cambrian era.

7. If life exists on another planet, it will most certainly evolve under the same constraints and directions that we recognize here on Earth. Therefore, life on another planet would undoubtedly lead to the appearance of humans or very similar beings. However, this cannot happen – life does not exist anywhere else and could not exist since its appearance on Earth is dependent on a vast number of contingent factors. The possibility of all of them being repeated anywhere else is close to zero. Humans are an inevitable outcome of the evolutionary process that could have happened only once.

Gould vs. Conway Morris: The Face-off

In this section, I will explore the two authors' arguments that support their theses to present both their strong and weak points.

1. Cone of increasing diversity

The core of Gould's contingency thesis is based on the reinterpretation of the Burgess fossil record, which has shown that the total amount of phyla right after the Cambrian explosion was greater than it is today.

In *The Crucible of Creation,* Conway Morris attempted to refute this argument and, thereby, Gould's contingency thesis. He blamed Gould for making the same mistake that he accused Walcott of – using the ideological shoehorn to classify fossils so that they match the presupposed notion of the evolution of life. However, he never mentions that Gould took this shoehorn from himself and the rest of the Cambridge Team and that Gould's interpretation of the Burgess fauna was originally theirs. This has not passed unnoticed by Gould himself (Conway Morris and Gould 1998/1999) and others (Fortey, Briggs, and Wills 1996, 24–25).

The research of the Cambridge Team was done, as is shown by Conway Morris (Conway Morris 1998, 171–176), within the theoretical framework based upon the research of Sidnie M. Manton. She believed that arthropods have a polyphyletic origin. According to Manton, every group of arthropods represents a different phylum that evolved from a worm-like form separately (Manton 1977). Since they were unable to classify them in any of the existing phyla, the researchers of the Cambridge Team believed that some of the wondrous Burgess creatures are representatives of a separate, thereby unknown phyla. However, later research – especially one based on cladistic classification – has shown, Conway Morris believes, that both Manton and the Cambridge Team were wrong. Arthropods do not have a polyphyletic origin and are more closely related than previously thought.[1]

1 It is worth noting that by denying the polyphyletic origin of arthropods, Conway Morris has rejected one of the compelling arguments in favor of his convergency thesis. However, I don't think this should make us question his sound mind, as Peter Bowler does (Bowler 1998, 475), but rather respect his intellectual honesty.

He also argues that the increase or potential decrease of the number of phyla should be analyzed within a particular group of animals separately. The analysis would in fact show that the number of body plans within arthropods has increased. Although he didn't compare the total number of phyla in Cambrian and today – which might be considered a telling silence – the research has indeed shown that this comparison doesn't suit Gould's thesis either. The number of phyla today is the same or even greater than in the Burgess (Fortey and Briggs 2005; Budd and Jensen 2000).

Conway Morris also criticizes Gould's thesis regarding the possible causes of the Cambrian explosion characterized by unprecedented adaptive radiation. With much caution and hesitation, Gould offered a hypothesis that the genome had more plasticity at the time. Therefore, the jumps in its restructuring during the Cambrian could have been done more easily than it would have been possible today. These jumps resulted in the emergence of many different anatomic plans in, geologically speaking, a short period (Gould 1990, 230–32). Many scholars, including Conway Morris, tended to interpret what he presented as a hypothesis to be a dogmatic standpoint. Gould suffered severe critique from the scientific community. Conway Morris believes that the causes of the Cambrian explosion should be searched for in ecological conditions, i.e. a large number of empty ecological niches, as well as in the sudden increase of food resources and phosphate, and not within the genome itself (Conway Morris 1998, 153–65).

Going back to the core argument of Gould's contingency thesis, it should be noted that it consists of two mutually dependent premises: (1) maximal initial proliferation, and (2) randomness of the decimation process. The first premise, taken by itself, could also be used to support the deterministic view of life's history. If the number of phyla initially was greater than it is today, that does not tell us anything about the nature of the process which resulted in the decrease of this number. The reduction in the number of phyla could result from the natural selection pressure that favored the more adapted phyla over others. In that case, every tape of life's replay would yield the same results regarding the survival of certain phyla and the extinction of others. Therefore, to support the contingency thesis the argument from maximal initial proliferation requires the other one – on the random nature of decimation.

Conway Morris refutes the first premise and demonstrates that Gould's basic framework regarding the initial radiation of body plans was flawed. The number of phyla has indeed increased over time rather than decreased. However, he does not deal with the second premise at all, nor does he seem to recognize the mutual dependency between the two. But he does not have to, either. Since he has shown that the number of phyla has not decreased over time, he is not obliged to prove that the otherwise inexistant process of decimation was not governed by random but rather deterministic factors.

Therefore, we may conclude that by refuting one of the two basic premises, Conway Morris significantly undermines the contingency thesis, at least in the way that Gould has formulated it. Since false premises may lead to conclusions that are themselves true, it is conceivable that the contingency thesis might be formulated in a more plausible manner, not relying on the assumption of maximal initial proliferation.

It is worth noting in this context that unlike Gould, Conway Morris is keen to defend the idea of progress in evolution. A decrease in the number of phyla diminishes the idea of progress as it is defined by Gould (as an increasing diversity), but not as it is defined by Conway Morris (as an increasing complexity). Therefore, for Conway Morris, maximal initial proliferation, even if it took place – and he shows that it has not – does not call into question the idea of progress. The increase in complexity, which he takes as too obvious to require explanation, bares the fact that progress is real. While it is not impossible that the two would agree on what complexity is and that it has in fact increased over time, it is evident that they would disagree on the matter of whether that constitutes progress or not.

On the other hand, I believe that Conway Morris' effort to reject Gould's speculations regarding the causes that fueled the Cambrian explosion misses the point. By refuting them he does not harm the contingency thesis. Whatever one might consider being its root cause – genomic flexibility or ecological factors – it could complement either the contingency or convergence thesis. It is entirely imaginable that the unique process of adaptive radiation in the Cambrian was possible thanks to the flexibility of the genome. At the same time, its outcome – the perseverance of one group of phyla over the other – was caused by deterministic processes, e.g., natural selection. The complete opposite could also be the case – that the maximal initial proliferation was caused by ecological factors, as stated by Conway Morris, while contingent factors determined the later trajectories of evolution. I believe that the reason why Conway Morris insists so much that Gould's hypothesis regarding the genome's flexibility is false is because he wants to stress the fact that very well-known processes based in the Darwinian paradigm were in place back then as they are now.

2. Prediction

As we have seen, Gould's premise on initial maximal proliferation is dependent on the one regarding decimation. Moreover, the argument from decimation is more critical than the first one. This argument is entirely based on the impossibility of making reliable predictions of the success of specific body plans. Gould believes that the theory of evolution, *in order to be scientific*, has to be capable of producing predictions and not only retrospective explanations (Gould 1990, 236). If a particular trait is designated as adaptive by an evolutionary biologist, then by doing so, one predicts that the organisms which possess this feature will have a greater chance to survive and produce descendants. Therefore, the hypothetical paleontologist enabled to get a glimpse of live Burgess fauna would have to be able to offer a prediction of the survival of specific phyla over others. However, as we have seen, it would have been impossible for him to do so. Since this prediction would have been impossible, Gould concludes that the decimation was utterly random. Therefore, evolution is to be considered a contingent process. I believe, however, that Gould set up this argument wrongly.

First, it seems as if Gould himself was not entirely convinced of the randomness of the decimation. Whenever he speaks of the critical importance of the lottery, he adds ‘maybe’ or ‘probably’ (Gould 1990, 276, 288, 301). He also states that we cannot be sure that this process was indeed a lottery (Gould 1990, 239, 302). In addition, he seems to be unsure about his concept of contingency and conflates it with randomness (Gould 1990, 50–51; n. 5, n. 6; Blount, Lenski, and Losos 2018).

Secondly, even if the hypothetical paleontologist would have been unable to predict the survival of one group of phyla over the other, this does not mean that this outcome was random. At most, it means that the paleontologist could not have insight into all the factors that caused this outcome. And these factors could be deterministic or random. The critical role could be equally played either by natural selection or pure randomness. Gould seems to jump to a conclusion, implying that the current knowledge about these factors equals all that could be known. There may be certain deterministic factors that played a vital role in the process of decimation that are entirely unknown to us. The fact that these are still unknown *to us* does not mean that *they do not exist.* Gould has simply identified epistemic indeterminism with ontological indeterminism.

Thirdly, Gould postulates a criterion for the theory of evolution to be scientific and then proceeds to show that it can never be met. Some scholars think that even if it were unable to make valid predictions, the theory of evolution would still be scientific (Wasserman 1981). However, Gould was clear about it: making valid predictions is a necessary condition for a theory to be considered scientific. However, he does not offer a single example of a valid prediction that *could,* in fact, be made on a macro-level of evolution. Moreover, the predictability of evolution is only possible if we accept that deterministic processes guide its outcomes on macro levels. For this reason, the idea of evolution’s predictability is tied closely to the adaptationist program (Sober 2000, 122). Making long-term predictions regarding an outcome of a process essentially affected by randomness and lottery would have been impossible. And this is precisely the sort of prediction that Gould expects. Therefore, it turns out that the only scientific model of the theory of evolution is precisely the one he rejects, i.e., the adaptationist model. Since he (1) failed to explain whether it would be possible to make a prediction that will not be based on the idea of predominance of natural selection; and (2) failed to offer some other criteria which the theory of evolution has to meet to be considered scientific, Gould has left it outside of the demarcation line of science.[2]

Although he offered a convincing image of life's history that *supports* the contingency thesis, Gould failed to provide sufficient arguments in its favor. By challenging the basic premise of maximal initial proliferation, Conway Morris has indeed undermined the very basis of Gould's contingency thesis.

On the other hand, Conway Morris also believes that evolution is predictable (Conway Morris 2010).[3] As a proponent of the adaptationist program, establishing a basis for predictability in the predominant role of natural selection in life's history does not pose a problem for him. However, predictability for Conway Morris is based on the phenomenon of convergence. Therefore, one might conclude that in Conway Morris' mind, the hypothetical paleontologist could not have predicted which of the anatomic plans would survive, but that is not important at all. He could have predicted with considerable certainty that some biological features would be developed in the post-Cambrian future in any case, regardless of the survival of specific phyla. This prediction would, therefore, be based on the all-presence of convergence.

3. Convergence

Conway Morris' idea of predictability is, as we have seen, based upon the all-presence of convergence in all levels of life. If life forms converge toward identical solutions, i.e. the attractors, then the process of evolution is under constraints that determine its trajectories and outcomes. As such, it is predictable and its inevitable result: humans. What follows will point to some weak points in Conway Morris' convergence thesis.

First, Conway Morris does not point to the direction in which one should look for the root causes of convergence. One has to admit that phenomenon of convergence is rather complex, and mapping its ubiquitous presence and importance for the understanding of life's evolution is a significant endeavor on its own merit. However, even if it would be too demanding to ask for a definitive explanation of its root causes, one might expect that Conway Morris would at least offer a clear direction in which it is to be sought for. However, this seems to be lacking. One is led to believe that the primary cause should be located in the environment that acts through the mechanism

---

2 Molnar (Molnar 2008) makes a similar remark.

3 It might be worth noting that for both of them predictability implies knowledge that we have already acquired with the passing of the history of life's evolution. The hypothetical paleontologist is not imagined as someone who lived at the time of the Cambrian fauna, but rather someone who *came back in the past* with the knowledge of final outcomes.

of natural selection. However, on multiple occasions he points to examples of convergence in radically different environments. For instance, he shows that the crab-like form has evolved independently numerous times within the arthropods in entirely different surroundings (Conway Morris 2003, 130). But he does not discuss the possibility that in these cases convergence is unrelated to natural selection (Wake 1991).

On the other hand, he *does* point to the similar features of rather different environments. He states, for example, that although fission-fusion societies have evolved among the primates and dolphins in different surroundings, "there is a deeper constraint imposed by the patchiness in space and time of food resources in both ocean and jungle" (Conway Morris 2003, 249). In this case, *the same properties* of *different surroundings* are viewed as the causes of convergence. It seems, however, that this interpretation is relatively weak since the mentioned property (the patchiness in space and time of food resources) can be viewed as almost universal to any given surrounding. Therefore, it can be argued that convergence is not the result of the constraints imposed by the environment, or at least not in this case. Furthermore, constraints cannot be interpreted as a result of the physical or chemical limitations either since, as we know, different anatomic plans are realized e.g. among the arthropods. If the environmental, physical, and chemical limitations are not the causes, the only place to look for them is in the evolutionary trajectory of each species. But if one would identify the causes for convergence in this domain, it would be unclear how to interpret the phenomenon of convergence.

If we understand it as a process of producing good solutions to the tasks imposed by the environment, as Conway Morris seems to be suggesting, it remains unclear how to interpret the fact that the solutions continue repeating in different environments. If convergence is not an exclusive product of the force of the environment acting through natural selection and secondarily a byproduct of other types of constraints, it is unclear how one should interpret it.

Secondly, Conway Morris views convergence as an act of reaching the adaptive peaks and approaching the 'attractors' of functionality. When it comes to particular adaptive features, the existence of attractors is well demonstrated. Sabre-teeth, wings, and echolocation are all individual features that have evolved independently multiple times in different clades. Therefore, their evolution was, as Conway Morris puts it, inevitable. But he goes a step further, claiming that the evolution of certain *groups of features* characteristic of certain groups of organisms was also inevitable. He attempts to demonstrate that 'mammal-ness' was one such group of features whose emergence was inevitable. He proves his point by analyzing the traits of 'the honorary mammal,' the Kiwi bird. He seems to be jumping to a conclusion here. Here is why.

'Mammal-ness' as a group of traits is demonstrated through one or two broadly defined traits: nocturnal life and non-mammalian ovoviviparity. Therefore, Conway Morris' predictions regarding the evolution of such groups are not specific in any meaningful way. For example, he states that the evolution of a whale could not have been predicted at the time of Cambrian, but that the evolution of a fast sea animal that feeds by filtering water indeed could have been known (Conway Morris 1998, 202). However, such a definition of a group of traits is wildly unspecific. Let alone the fact that this type of nutrition was made possible by closing off access from mouth to lungs, an evolutionary byproduct common for mammals (Foote 1998, 2069). Similarly, one could argue that the existence of the bipedal terrestrial carnivore has been predictable since Cambrian. However, the problem is that this portrayal of a 'featherless biped' is a sort of Platonesque definition which fits the description of both *Alioramus* and humans.

On the other hand, Conway Morris decomposes the organism and views it as a set of traits, which is precisely what he previously criticized when he stated that 'organisms are more than the sum of their parts' (Conway Morris 1998, 9). Even if one accepts that the evolution of a particular group of traits was inevitable, it is not the same as saying that the whole and unique set of the same characteristics, as found in specific organisms, was also inevitable. A group of traits cannot be mistaken for the whole composition of characteristics. The evolution of particular traits, like intelligence, depends entirely on the emergence of the entire composition of traits, not only on its sum. If one would take the agriculture of ants, the vocal capacities of a parrot, the carnivore diet of a cat, the bipedal movement of a dinosaur, the warm-bloodedness of a vulture, the sense of smell of a mole, the camera-eye of an octopus, etc., it is hard to imagine that one would end up with an organism capable of abstract thinking.

These weaknesses in Conway Morris' arguments affect his thesis on the inevitability of humans as well.

4. The inevitability of humans?

Conway Morris claims that intelligence is also one of the convergent traits. Besides primates, it is present among sea mammals, birds, and cephalopods. Although he claims that these types of intelligence are not the same (Conway Morris 2003, 156; 264), he never explains the difference between them. Therefore, he remains open to the objection that human intelligence is actually not a convergent property.

He also points out that certain constraints can stop the further evolution of a species and its advance towards the adaptive peak. These constraints are set by the evolutionary history of a species. Therefore, Conway Morris affirms the existence of 'the burden of history' while he simultaneously limits its effects:

> … the constraints of past 'decisions' that guide, restrict, and perhaps even interfere a phylogenetic 'career'. That such constraints exist is undeniable, but what is far more interesting is the way in which organisms repeatedly 'get round' these problems, which is why convergences are ubiquitous (Conway Morris 2003, 302)

However, he fails to mention that there are counterexamples in which the historical constraints are so strong that evolution was unable to get around them. For example, he has previously stated that the dolphins have 'hit the wall' of evolution since they are unable to overcome the constraints of their environment. Although intelligent, they cannot use tools and develop technologies since they live in the sea. Therefore, the evolution of intelligence in this species couldn't advance further toward the adaptive peak (Conway Morris 2003, 260). The same goes for the octopus, another example of the convergence of intelligence. These examples demonstrate that the adaptive peaks may become unreachable through particular evolutionary trajectories. They become discovered but not attained.

If we take human-like intelligence to be one of those adaptive peaks, it turns out that its emergence was entirely dependent on the historical trajectory of hominid evolution. And this is where we come back to Gould's experiment of life's tape. Even if we consider that the emergence of human-like intelligence was possible in all the descendant lines of *Homo Erectus*,[4] it turns out that it was conditioned by the survival of a tiny African population of *Homo Sapiens*. If it had become extinct as well, as other species of Homo did, one could not have hoped for the evolution of consciousness. However, if the emergence of primates conditioned the evolution of consciousness, their existence was conditioned by mammals' survival and adaptive radiation after the extinction of dinosaurs. This is where things start to get slippery for Conway Morris. He thinks that mammals' survival and adaptive radiation were inevitable since they were better adapted to the cold climate, which was about to get colder. In this case, the appearance of humans on the world stage would still have happened, although it would have been delayed by a couple of million years. However, he does not proceed to prove this bold thesis. Still, he proceeds by demonstrating something else, namely, that the evolution of *mammal-ness* (which is not to be mixed with the mammal, as he points out) was inevitable (Conway Morris 2003, 222–23). He points to the Kiwi bird as an 'honorary mammal.' Therefore, the claim for the inevitability of mammal-ness is based upon the existence of a single species whose mammal-ness is quite dubious. This is where we once again get back to Gould: if the asteroid

4 Gould believes that this was not the case, while Conway Morris thinks that in fact it was and offers more convincing evidence. For example, Gould (Gould 1990, 320) believes that the Neanderthals did not possess the capacities of abstract thinking and numerical reasoning. Conway Morris, on the other hand, refers to the studies which demonstrate that they indeed were capable of such things and that they did not acquire these capacities by mimicking humans, but developed them independently (Conway Morris 2003, 276–81).

didn't hit the Earth and wipe out the dinosaurs, would the adaptive radiation of mammals and the evolution of mammal-ness have happened? If the answer to this question is negative, then the same goes for the potential evolution of human-like intelligence and consciousness.

However, was it possible for the human-like intelligence to evolve within an entirely different group of animals that would not possess 'mammal-ness' at all? If it represents one of the adaptive peaks, should we expect that it could have been attained through different evolutionary trajectories?

Conway Morris is unclear about his opinion in this regard. He gravitates between two claims. Claim number one is that humans would have inevitably evolved as mammal-like creatures. As such, they are the final outcome of evolution. Once this outcome is accomplished, one should not expect it to happen again. Along these lines, he points to the structural constraints of other species (which are results of their own evolutionary history) in which intelligence has evolved that limits its further evolution. Also, he explicitly states that

> Even acknowledging the realities of convergence is not to imagine that every organism is 'trying' to evolve into a human (Conway Morris 2003, 302).

On the other hand, according to the second claim, human-like intelligence is one of the adaptive peaks scarcely achieved in evolutionary history. Still, it does not mean that things will not be different in the future (Conway Morris 1997, 14). This means that the emergence of human-like intelligence could have been achieved through other animal groups.

In the first case, Conway Morris would be forced to admit that the historical trajectory of our evolution determined the emergence of humans. For example, it could have been the case that the dolphins became the most intelligent species, which in fact, they were 1.5 million years ago (Conway Morris 2003, 247), but that they were unable to achieve human-like intelligence due to the constraints set by the environment, as stated before. Similarly, if the dinosaurs had not become extinct, it could have happened that the adaptive radiation of mammals or the evolution of 'mammal-ness' never occurred. Conway Morris' counterarguments against this possibility do not look convincing. In the second case, he would be forced to renounce the claim that 'not every organism is trying to evolve into a human,' or at least the claim that the human-like beings would have been necessarily mammal-like.

It can be concluded that Conway Morris successfully demonstrates the existence of 'attractors' of functionality, i.e. adaptive peaks that are reached repeatedly through life's evolution. They indeed might be 'the property of the system' (Ray 2006), which will be inevitably achieved. However, he fails to convince us that the whole set of traits, such as 'mammal-ness' or specific species, such as humans, should be regarded as such adaptive peaks. Even if the ’attractors’ indeed represent the property of the system, the trajectory through which they are achieved is not. If history plays a role in the evolution, and Conway Morris admits it does, one could have expected the emergence of behavioral flexibility characteristic of intelligence, but consciousness and its emergence within the mammal-like animals perhaps not.

Conclusion

I analyzed Gould’s convergency and Conway Morris’ convergence thesis regarding the nature of life’s evolution and the arguments they used in their support, respectively. Thereby I attempted to show the strengths and weaknesses of their arguments. I believe that the power of Gould’s thesis is not in what he thought was sufficient evidence for the contingency thesis but instead in a vivid and illustrative review of life’s history. In it, he pointed out some of the turning points that were entirely dependent on contingent factors and directly impacted the emergence of the human species. On the other hand, Conway Morris has successfully demonstrated core weaknesses of Gould’s basic premises and demonstrated convincing arguments in favor of the existence of what he calls the ’attractors’ of functionality. However, he failed at proving that mammal-ness and humans are one of those attractors. Both Gould and Conway Morris have shaped particular models of understanding the evolution of life which could be further developed if their shortcomings are taken into consideration.

Bibliography

Baron, C. (2009). Epistemic values in the Burgess Shale debate. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *40* (4), 286–295. https://doi.org/10.1016/j.shpsc.2009.09.008

Baron, C. (2011). A Web of Controversies: Complexity in the Burgess Shale Debate. *Journal of the History of Biology*, *44*, 745–780. https://doi.org/10.1007/s10739–010–9248–2

Beatty, J. H. (2006a). Replaying Life’s Tape. *The Journal of Philosophy*, *103* (7), 336–362. https://doi.org/10.5840/jphil2006103716

Beatty, J. H. (2006b). The Evolutionary Contingency Thesis. In E. Sober (Ed.), *Conceptual Issues in Evolutionary Biology*, 2017–2248. Massachusetts Institute of Technology.

Blount, Z. D., Lenski, R. E., & Losos, J. B. (2018). Contingency and determinism in evolution: Replaying life’s tape. *Science*, *362* (6415). https://doi.org/10.1126/science.aam5979

Bowler, P. J. (1998). Cambrian Conflict: Crucible an Assault on Gould's Burgess Shale Interpretation. *American Scientist*, *86* (5), 472–475.

Budd GE, Jensen S. 2000. A critical reappraisal of the fossil record of the bilaterian phyla. *Biological reviews of the Cambridge Philosophical Society 75* (2): 253–295. https://doi.org/10.1017/s000632310000548x.

Ćirković, M. M. (2014). Evolutionary contingency and SETI revisited. *Biology and Philosophy*, *29*, 539–557. https://doi.org/10.1007/s10539–013–9397–8

Conway Morris, S. (1998). *The Crucible of Creation: The Burgess Shale and the Rise of Animals*. Oxford University Press.

Conway Morris, S.; Gould, S. J. (1998/1999). Showdown on the Burgess Shale. *Natural History* 107 (10), 1–12.

Conway Morris, S. (2003). *Life's Solution: Inevitable Humans in a Lonely Universe*. Cambridge University Press.

Conway Morris, S. (2010). Evolution: like any other science it is predictable. *Philosophical Transactions of Royal Society*, *365*, 133–145. https://doi.org/10.1098/rstb.2009.0154

Conway Morris, S. (2013). Life: the final frontier for complexity? In C. H. Lineweaver, P. C. Davies, & M. Ruse (Eds.), *Complexity and the Arrow of Time*, 135–161. Cambridge University Press. https://doi.org/10.1017/CBO9781139225700.010

Currie, A. M. (2012). Convergence, contingency & morphospace. *Biology & Philosophy*, *27* (4), 583–593. https://doi.org/10.1007/s10539–012–9319–1

Dresow, M. (2019). Gould's laws: a second perspective. *Biology and Philosophy*, *34* (5), 1–20. https://doi.org/10.1007/s10539–019–9698–7

Foote, M. (1998). Contingency and Convergence. *Science*, *280* (5372), 2068–2070. https://doi.org/10.1126/science.280.5372.206

Fortey, R. A., Briggs, D. E. G., & Wills, M. A. (1996). The Cambrian evolutionary "explosion": decoupling cladogenesis from morphological disparity. *Biological Journal of the Linnean Society*, *57* (1), 13–33. https://doi.org/10.1111/j.1095–8312.1996.tb01693.x

Futuyma, D. J. (2010). Evolutionary constraint and ecological consequences. *Evolution*, *64* (7), 1865–1884. https://doi.org/10.1111/j.1558–5646.2010.00960.x

Gould, S. J. (1990). *Wonderful Life: The Burgess Shale and the Nature of History*. W. W. Norton & Company.

Harmon, L. J., Kolbe, J. J., Cheverud, J. M., & Losos, J. B. (2005). Convergence and the multidimensional niche. *Evolution*, *59* (2), 409–421. https://doi.org/10.1111/j.0014–3820.2005.tb00999.x

Hopster, J. (2017). Evolutionary chance and contingency: in search for systematics. *Metascience*, *26* (3), 481–485. https://doi.org/10.1007/s11016–017–0213-y

Losos, J. B. (2011). Convergence, adaptation, and constraint. *Evolution*, *65* (7), 1827–1840. https://doi.org/10.1111/j.1558–5646.2011.01289.x

Manton, Sidnie, (1977). *The Arthropoda: Habits, Functional Morphology and Evolution*. Oxford University Press.

McConwell, A. K. (2019). Contingency's causality and structural diversity. *Biology and Philosophy*, *34* (2), 1–26. https://doi.org/10.1007/s10539–019–9679-x

Mcshea, D. W. (1993). Arguments, Tests, and the Burgess Shale – A Commentary on the Debate. *Paleobiology*, *19* (4), 399–402.

Molnar, R. E. (2008). "Life's Solution, Inevitable Humans in a Lonely Universe" by Simon Conway Morris. *Journal of Vertebrate Paleontology*, *28* (2), 586–587. https://doi.org/10.1671/0272–4634(2008)28[586:LSIHIA]2.0.CO;2

Ramsey, Grant, and Charles H. Pence, eds. 2016. *Chance in Evolution*. The University of Chicago Press.

Ray, T. S. (2006). "Life's Solution: Inevitable Humans in a Lonely Universe" by Simon Conway Morris. *Artificial Life*, *12*, 453–456. https://doi.org/10.1162/artl.2006.12.3.453

Sober, E. (2000). *Philosophy of biology*. Westview Press.

Stayton, C. T. (2015a). The definition, recognition, and interpretation of convergent evolution, and two new measures for quantifying and assessing the significance of convergence. *Evolution*, *69* (8), 2140–2153. https://doi.org/10.1111/evo.12729

Stayton, C. T. (2015b). What does convergent evolution mean? The interpretation of convergence and its implications in the search for limits to evolution. *Interface Focus 5* (6). https://doi.org/10.1098/rsfs.2015.0039

Turner, D. D. (2011). Gould's replay revisited. *Biology and Philosophy*, *26* (1), 65–79. https://doi.org/10.1007/s10539–010–9228–0

Wake, D. B. (1991). Homoplasy: the result of natural selection, or evidence of design limitations? *American Naturalist*, *138* (3), 543–567. https://doi.org/10.1086/285234

Whittington, Harry B. (1985). *The Burges Shale*. Yale University Press.

Wong, T. Y. W. (2020). Sources of evolutionary contingency: chance variation and genetic drift. *Biology and Philosophy, 35*, 36 (4). https://doi.org/10.1007/s10539–020–09752–4

Petar Nurkić
Institute for Philosophy
Faculty of Philosophy University of Belgrade
E-mail: petar.nurkic@f.bg.ac.rs

Original Scientific Paper
UDC: 575.8:167/168
113:167/168
573.5:1

DARWIN MEETS DR. FRANKENSTEIN: USING THE DRAKE EQUATION TO CALCULATE THE PROBABILITY OF VOLCANIC LIGHTNING'S IMPACT ON CHEMICAL EVOLUTION

Abstract: *Horizontal gene transfer (HGT) has been a paramount mechanism of interest in recent literature addressing the origins of biological evolution. However, research on lightning-triggered electroporation represents the innovative and still insufficiently grasped approach to HGT (Kotnik, 2013). On the other hand, prebiotic synthesis is a fundamental process for chemical evolution. Recently, the effects of volcanic lightning on nitrogen fixation and phosphate reduction have also been considered (Navarro-González and Segura, 2004). This paper aims to present a top-down approach to the question of the origin of life on early Earth. By considering the conditions necessary for the emergence of biological and chemical evolution, emphasizing electrostatic discharges, we will attempt to link previous theoretical and experimental research. Furthermore, we will present a recent endeavor at applying the Drake equation to calculating the probability of volcanic lightning impact on the prebiotic synthesis and derive a similar use in estimating the contribution of lightning to HGT (Weaver, 2013). We will also display that choosing a type of probability appropriate for the context of life sciences is not necessarily a quantitative issue. Finally, we will show that significant conceptual constraints, like determining the relevant factors and sources of uncertainty when considering the origin of life on early Earth, are fundamentally philosophical issues. We hope that the results of our research – deriving Drake's equation in the domain of chemical evolution and considering Bayesian and counterfactual types as potentially more suitable candidates for calculating probabilities in the evolutionary framework – will contribute to developing new discussions in life sciences.*

Keywords: *biological evolution, chemical evolution, horizontal gene transfer, prebiotic synthesis, volcanic lightning, Drake equation.*

1. Introduction

A wide variety of scientists interested in questions about the origin of life believe that the first living cells were created by a natural process called chemical evolution. However, when we talk about evolution, we focus mainly on its biological aspects: reproduction, variation, and selection (Sober, 2006: 11–16). To examine chemical evolution and how it differs from biological evolution, we need to define key concepts in chemistry and biology and demonstrate how chemical evolution creates complex systems from simple molecules that form a similar structure that we can find in living cells[1].

Evolution, simply put, is change over time. This change is focused primarily on biological organisms that can reproduce. Change over time in biological evolution places great emphasis on survival. If an individual can survive, it also gets the opportunity to reproduce and make „copies of itself" while entire populations develop new traits and abilities. Reproduction, variation, and selection as the basis of biological evolution can be illustrated by focusing on Plant X and Y (Obeso, 1997). Suppose Plant X has leaves with smooth edges, while Plant Y, which belongs to the same species, has spines that allow it to vary. The natural habitat of Plants X and Y are deciduous forests where herbivores are abundant. Plant Y is harder to eat because the thorns are not particularly pleasant for chewing and digestion and therefore have a higher probability of survival and reproduction. Nature sometimes quite arbitrarily places barriers to survival and selects opportunities for reproduction and the transmission of traits to new generations. In the case of Plants X and Y, the spiny leaf mutation represents a new trait that provides a distinct survival advantage.

The problem underlying biological evolution is the necessity of reproduction to function properly[2]. Complex reproduction process gives rise to the question: how did evolution initially evolve? To answer this meta-question, researchers turn to chemical evolution (Ruiz-Mirazo et al., 2017). In chemical evolution, we can observe changes in organisms that cannot reproduce. These changes are relatively simple, such as iron corrosion when it comes into contact with water. Nevertheless, no matter how minimal their effect, simple chemical changes lead to the formation of organisms capable of reproducing and acquiring new traits. Thus, the main difference between biological and chemical evolution is that reproduction is replaced by a more straightforward process – repetitive production (Rauchfuss, 2008: 21–29).

As we have indicated, this paper aims to take a top-down approach to questions about the origin of life on the early Earth, with particular emphasis on electrostatic discharges, or lightning, as triggers of natural processes

---

1 Our paper does not aim to show the influence of lightning on the occurrence of chemical or biological evolution but evolution in general. Although it can be more challenging to follow, we will provide key aspects of chemical and biological evolution in parallel. The rationale for this approach is the continuous retention of the general characteristics of evolution as an umbrella phenomenon. With this parallelism, we will single out the similarities of the origin of life on different explanatory levels and more easily illustrate the key philosophical problems of calculating probability in evolutionary domains.

2 For recent research considering the possibility of biological evolution with natural selection but without reproduction, see Papale (2021). To achieve the goals of our paper, we will consider the more traditional assumption of reproduction as a necessary condition for evolution.

responsible for the emergence of evolution. Therefore, the second chapter focuses on biological evolution, in which we describe horizontal gene transfer and electroporation as central phenomena responsible for the exchange of intercellular material between unicellular organisms. In the third chapter, we describe the process of prebiotic synthesis and the elements necessary for chemical evolution – nitrogen and phosphorus. For nitrogen and phosphorus to participate in prebiotic synthesis, they must be converted to ammonia and hydrophosphates through processes of nitrogen fixation and phosphate reduction. The stimulating environment of volcanic gas and ash clouds that generate volcanic lightning is the necessary condition for these conversion processes. In the fourth chapter, we present an attempt to apply the Drake equation to calculate the probability of the influence of volcanic lightning on the origins of chemical evolution. Finally, we will attempt to apply the same equation to calculate the probability of the lightning-triggered electroporation influence on horizontal gene transfer and the origin of biological evolution. In the final chapter, we will also examine whether the Drake equation is the most appropriate probability model to apply to natural processes underlying the origin of life on early Earth. By considering the relevant parameters necessary for the origin of life and the sources of uncertainty in calculating the probability of the impact of lightning on prebiotic synthesis and HGT, we will try to show that this is not a purely quantitative issue but a fundamental philosophical undertaking of different possibilities.

2. Warm little pond

But if (and oh what a big if) we could conceive in some warm little pond with all sorts of ammonia and phosphoric salts, light, heat, electricity etcetera present, that a protein compound was chemically formed, ready to undergo still more complex changes (Browne, 1978).

Did evolution take place in a warm little pond? In his 1871 letter to Joseph Hooker, Charles Darwin hints at an idea that is still very relevant today. Darwin's „little pond" is a matter of a suitable environment for the origin of life, whether it is a pond, fresh or saltwater, soil, or some other environment. More importantly, all of the parameters that Darwin listed are still under intense scrutiny. Our paper also considers the necessary ranges of electricity, heat, nitrogen fixation, and phosphate reduction required for evolution to occur.

The possibility of lightning-mediated horizontal gene transfer (HGT) contributing to the origin of evolution represents the Darwinian extension of the „warm little pond" question. Various researchers from different fields have attempted to respond to Darwin's inspiring astonishment, but none has done so in such detail as Tadej Kotnik (2013).

2.1 Horizontal gene transfer (HGT)

In the introductory part, we mentioned that one of the main features of biological evolution is variation. This subsection will explain why HGT is an essential contributor to prokaryotic genetic variability. HGT is the transfer of genetic material between unicellular or multicellular organisms or DNA transfer between „ancestors“ and „descendants“. Of particular importance to HGT is the phylogenetic tree, which shows the common roots of bacteria, archaea, and eukaryotes.

Research on the transmission of genetic information is not new. It dates back to Griffith's 1928 experiment, which showed that bacteria could exchange intracellular content (Blokesch, 2016). Several mechanisms enable HGT, but we can single out three natural pathways by which DNA transfer is enabled (Kotnik, 2013: 355): (i) bacterial conjugation, (ii) natural bacterial competence for DNA uptake, and (iii) viral transduction. The fourth natural mechanism that enables horizontal gene transfer is (iv) membrane electroporation[3]. Each of these mechanisms evolved at a particular stage of evolution, but the question of a possible fourth mechanism also answers the problem of the existence of HGT prior to each of these mechanisms.

Membrane electroporation is the result of atmospheric electrostatic discharges, i.e., lightning. Several theoretical and experimental considerations support the thesis that HGT is a process triggered by lightning (Golberg, 2013; Weaver and Chizmadzhev, 2018). Before turning to the chapter on electroporation and gene transfer, it should be noted that HGT is a vital process for variability in biological evolution mentioned above. Without the mechanisms of HGT, Plant Y could not mutate successfully and would not have the advantage of spiny leaves over Plant X, which is easily digested by herbivores. HGT represents, in a sense, a primordial mechanism of natural selection. Cells whose genes acquire a proper function have a significant advantage over cells that are not so „lucky“.

2.2 Electroporation and HGT[4]

As we have already defined earlier, electroporation exposes membranes (unicellular or multicellular organisms) to electric fields of sufficient strength

---

3 Kotnik (2013) and numerous other researchers cite electrofusion as a natural HGT mechanism. However, for the purposes of this article, which focuses on probability calculations in the final section, we will focus exclusively on electroporation as a more likely mechanism that requires fewer assumptions and *ad hoc* hypotheses.

4 Behind the thesis that electroporation is a successful mechanism that enables HGT is a fascinating story about the Louisville Water Company (LWC). In 1896, the LWC investigated various methods of purifying water by killing microorganisms. On this occasion, they used high-voltage electrical pulses and discovered that this process contributed to the leakage of intracellular material, including DNA (Benton, 1896).

and conductivity (Kotnik, 2013: 355). We will explain how membrane electroporation enables the exchange of genetic material and what role water molecules play in this process. That means that electroporation and HGT depend on the aquatic environment in which they occur. The voltage ranges, measured in kV/cm, required to form transmembrane electric fields should also be considered[5]. We will also explain exactly how the pores through which HGT occurs are formed by the electric fields.

The ranges of electric fields can be divided into low, intermediate, high, and very high. In the low ranges of electric fields, formed pores are too small to allow molecular transport through the membrane. In the intermediated range, the pores only provide a temporary HGT pathway. However, they clog very quickly and interrupt HGT. In high ranges, the cells electroporate irreversibly, and the pores do not close, allowing an efficient exchange of intracellular material. However, the reason why high ranges are not optimal for HGT is also why very high ranges are optimal. This reason is thermal damage, allowing molecules to be released and DNA to melt. In very high ranges of electric fields, the temperature is high enough to cause thermal damage, and at the same time, electroporation is irreversible (*Ibid*: 357).

Thermal damage occurring in very high ranges of electric fields is characterized by pulse friction. The shorter the electrical pulse (as in the case of lightning in milliseconds), the more thermal damage is present, and electroporation successfully leads to irreversibility[6]. In aquatic environments affected by atmospheric electrostatic discharges, i.e., lightning, DNA exchange occurs between electroporated organisms, e.g., two prokaryotes without cell walls.

---

5 The field strength required for electroporation depends on the cell type; for bacteria, it varies between 3–24kV / cm, for mammalian cells, 0.25–3 kV / cm, and for plants, 3–12 kV / cm (Kumar et al., 2019).

6 Reversible electroporation is also considered an efficient mechanism for eukaryotic transformation (Kotnik, 2017). However, for the same reasons that we focused on electroporation rather than electrofusion, we will focus on irreversible rather than reversible electroporation. To calculate the probability, we will attempt to minimize the number of assumptions necessary for the possibility that HGT is a lightning-triggered process.

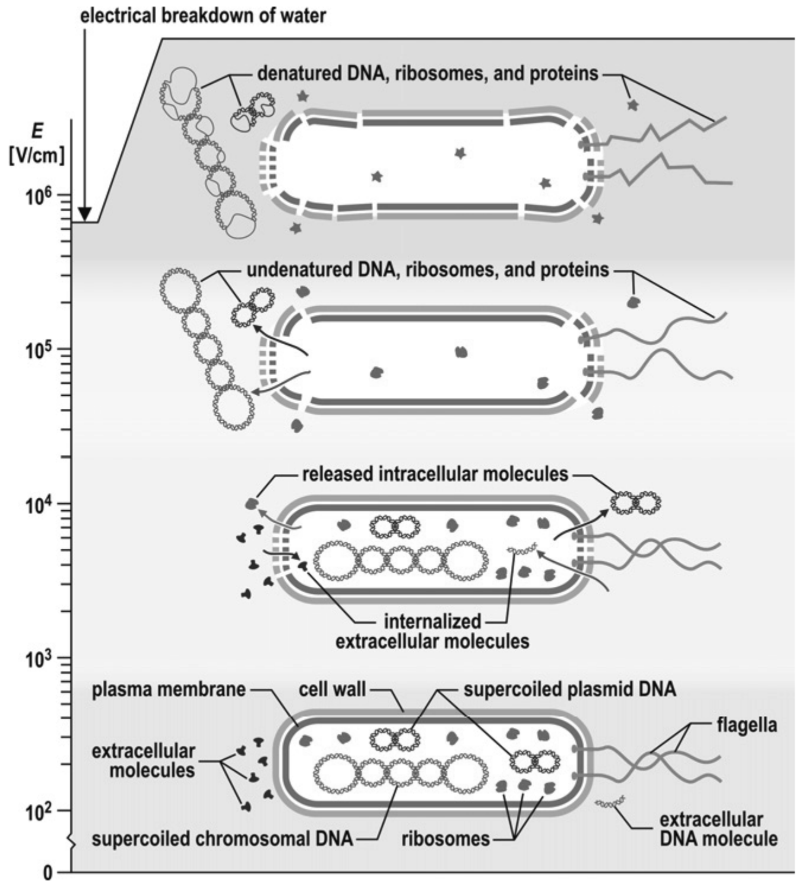

Figure 1. Electroporation-mediated molecular transport as a function of the external electric field – lightning (Kotnik, 2013: 356).

It should be noted that the vast majority of lightning is negatively classified, while only one-tenth are positive. Negative lightning occurs in sequences where the first stroke is most vigorous and originates mainly from the lower parts of the clouds, while positive lightning consists of a single stroke from the upper part of the cloud. Nag and collegues (2015) shows that the electrical currents range from 80 to 250 kA for 5–20 microseconds. That indicates that the first negative lightning strike could meet the required duration and electric current levels for lightning-triggered HGT to occur without DNA melting or rapid cell pore closure.

Considering rainwater, rivers, oceans, and lakes, it is necessary to single out the aquatic properties that HGT requires to be successful due to lightning strikes. When we talk about aquatic environments, we should pay attention to the type of water in question. Different aquatic environments have different electrical conductivities. Saltwater stimulates conductivity, but not the same degree as shallow and small ponds. In addition to conductivity, shallow and small ponds have more affluent populations of prokaryotes (Kotnik, 2013:

364). That favorable aquatic environments, electric field ranges, lightning strikes, and thermal DNA denaturation are not merely theoretical speculations is shown by the studies of Park and collegues (2013) on *Escherichia coli* and *Salmonella typhimurium* and Lee and collegues (2021) on *Pseudomonas aeruginosa*.

To conclude this chapter, it is essential to point out that theoretical considerations about HGT as lightning-triggered are supported by empirical research on the same phenomenon. Experimental conditions that simulate the environment of early Earth conditions must be finely nuanced to demonstrate the plausibility of theoretical hypotheses that would otherwise be mere fantasies at the long stick. One indicator of this research dynamic is the Miller-Urey experiment, which we will discuss in more detail in the next chapter.

Eruption of evolution

If we remove ancient cosmology from its anecdotal-mystical context and synchronize it with paradigm shifts in the life sciences, astrobiology, and prebiotic chemistry, we will find that Anaxagoras and Empedocles were not so far from the truth. Many pre-Socratic philosophers were guided by the idea that the entire universe, including our planet, is in constant pulsation. Natural events such as the tides, the alternation of day and night, and the recurring eruptions of volcanic geysers constantly give rise to new molecules and chemical systems through the processes from which they emerge. Molecules and systems become more complex and develop new properties as they interact with the environment. Simply put, they create life!

3.1 Fatty acids: bridging the gap between chemical and biological evolution

For illustrative purposes, we can take a fatty acid that consists of carbon, hydrogen, and oxygen atoms. As a group of atoms arranged in a specific pattern, fatty acid is just one of the complex molecules essential to living cells. Scientists believed that only cells could build fatty acids, but experiments conducted under controlled laboratory conditions provide new insights (Morigaki and Walde, 2007).

When simple gases, such as carbon monoxide and hydrogen, are heated with minerals in the earth's mantle, more complex carbon molecules such as fatty acids begin to form. That implies that living cells are not necessary for the emergence of chemical evolution but that there is a possibility that life may have originated in subsurface chambers using volcanic magma as a heat source (Van Gaever et al., 2009). As pressure increases, molecules rise into a water basin above the subterranean chambers, where a simplified version of natural selection takes control, as nature „decides" which molecules remain in the aquatic environment and which sink. Volcanic magma heats water in which fatty acids accumulate due to the attraction between oxygen and water molecules. On the other hand, the carbon in fatty acids repels water molecules. This dynamic of repulsion and attraction causes the fatty acids to collide with their „tails“ in the procession, forming a sphere. The result of this process is a stable, hollow vessel that resembles a membrane and represents an entirely new environment in which chemical evolution takes place[7].

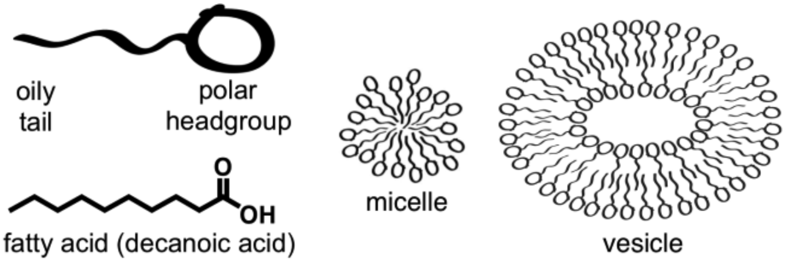

Figure 2. Fatty acid self-assemble into micelles and bilayer vesicles that resemble cell membranes (Black and Blosser, 2016: 4).

It is essential to point out that membranes formed from fatty acids cannot be classified as „living“ because, unlike living cells, they do not have the possibility of reproduction. That is very important because it shows that chemical evolution can, through further development, create new properties, environments, and systems that are fully reproducible. This transition from inanimate to living membranes or cells bridges the gap between chemical and biological evolution and represents a significant advance in life sciences.

3.2 Nitrogen and phosphorus

Nitrogen and phosphorus are fundamental elements for the origin of life, as they enable prebiotic synthesis. Nitrogen is found in proteins, enzymes, ribosomes, ATP, RNA, and DNA (Vitousek et al., 2002), while phosphorus is also found in ribosomes, ATP, RNA, and DNA molecules (Tiessen, 2008).

Prebiotic synthesis, or abiogenesis, combines molecules into more complex structures capable of independent reproduction. For that reason, prebiotic synthesis is crucial for chemical and biological evolution. Nitrogen, an essential element for processes such as cell division and morphogenesis, must be converted to hydrogen cyanide, ammonia, or nitrate to participate in prebiotic synthesis. For this to happen, nitrogen fixation, which allows nitrogen as an inert gas to chemically interact with other elements, must

7 The first membranes are composed of simple amphiphiles, i.e., fatty acids consisting of tails (hydrocarbon chains) and heads (carboxyl groups) forming water vesicles. Prebiotic fatty acids are spontaneously assembled into compartments that look like cells and have the capacity for growth (Black and Blosser, 2016: 4).

occur (Vitousek et al., 2002). On the other hand, phosphorus is essential for cell structure and its functions and must be reduced to hydrophosphites or phosphites to be eligible for prebiotic synthesis (Tiessen, 2008).

Atmospheric models show that significant amounts of nitrogen were present in the early Earth's atmosphere in the form of the inert gas – molecular nitrogen. Also, phosphorus was naturally present in the form of, both, the insoluble mineral *apatite* and the unusual mineral, *schreibersite*, found in meteorites. Experiments have revealed that schreibersite can disband in water to create hydrophosphites, which in turn react to assemble organic molecules essential for the origin of life (Walton et al., 2021).

3.3 Frankenstein's monster from the volcano

We have already mentioned that volcanic magma is a vital source of energy in underground chambers where chemical evolution occurs. Alongside that, even more interesting is the multifaceted influence of volcanoes on the origin of chemical evolution and life on the early Earth.

Navarro-González and Segura (2004) show how the impact of lightning can produce hydrogen cyanide and ammonia through nitrogen fixation. And not just any lightning, but electrostatic discharges from volcanic clouds. The same goes for the phosphate reduction that results from fulgurite formation due to the impact of volcanic lightning. Fulgurite is a hollow, glassy tube that forms in quartz sand (fossilized lightning). Significant amounts of schreibersite and other reactive phosphorus minerals can be found in fulgurite (Pasek and Block, 2009).

Clouds of volcanic ash and gas, where all the prebiotic components necessary for the origin of life are present, are suitable environment for synthesizing organic molecules (Navarro-González and Segura, 2004: 139). Volcanic ash contains minerals with sufficient surface area and catalytic properties, while volcanic clouds produce sufficient temperature and electrostatic discharge force as efficient energy sources. Both experimental studies and theoretical considerations suggest that lightning from volcanic clouds during high-explosive eruptions could be an essential source of reactive nitrogen and phosphorus, and thus chemical evolution (Navarro-González and Segura, 2004: 140).

Thermohistorical models show that volcanic activity of the early Earth was very intense, producing large amounts of lava and pyroclastics annually (Herzberg et al., 2010). Volcanoes emit physical properties in the form of gases, liquids, and solids, all of which are very important for prebiotic synthesis. Magmatic processes, where volatiles dissolve and magma is decompressed, and hydromagmatic processes, where liquid or ice comes into contact with lava, create environments where large amounts of gases and hot solid fragments are present. This further contributes to the formation of exhaust gases and airfall fragments that culminate in intense lightning activity from explosive volcanoes.

During these eruptive episodes, the pyroclastic and magmatic gases generate strong electric fields and photochemical processes that simulate a natural chemical reactor (Navarro-González and Segura, 2004: 141–143).

Parameters such as the individual variety of gases present in a volcano, the temperature and pressure, the strength of the electric fields, and the range of energy dissipation lead to the generation of different types of lightning: intracloud, cloud-to-ground, ground-to-cloud, and air discharges. Focusing on cloud-to-ground volcanic lightning and its properties, we can explain the origin of nitrogen fixation and phosphate reduction. Assuming that the early Earth's atmosphere was 80% carbon dioxide and 20% molecular nitrogen, Navarro-González (1998) showed that lightning fixes nitrogen, in experimental studies of Hawaiian volcanoes[8]. On the other hand, Pasek (2008) showed that the heating of apatite minerals and other minerals from volcanic environment, by lightning strikes, leads to polyphosphate formation.

To conclude this chapter, we must point out that theoretically severe considerations and conceptual barriers accompany the above experimental studies. In the next chapter, we will show what parameters and sources of uncertainty can be considered when calculating the probability of the influence of lightning on HGT and the prebiotic synthesis.

4. The truth is out there? Guidance from the Drake equation

> As I planned the meeting, I realized a few day[s] ahead of time we needed an agenda. And so I wrote down all the things you needed to know to predict how hard it's going to be to detect extraterrestrial life. And looking at them it became pretty evident that if you multiplied all these together, you got a number, N, which is the number of detectable civilizations in our galaxy. This was aimed at the radio search, and not to search for primordial or primitive life forms.
>
> —Frank Drake

In 1961, American astronomer and astrophysicist Frank Drake formulated an unusual equation. Although the equation calculates the number of interplanetary civilizations capable of communicating within the Milky Way, it was initially a semi-parody or approximation. Drake intended his equation to promote scientific dialogue at the first SETI[9] conference. Drake's equation can be represented as follows:

8 Due to the considerations that will be mentioned later in the paper, we must note here that the assumed composition of the Early Earth's atmosphere in the Miller-Urey experiment is completely different. In this experiment, the composition of the Early Earth's atmosphere is greatly reduced. The Miller-Urey experimental error will also be significant for dealing with key philosophical issues of parameter choice when calculating probabilities in evolutionary domains.

9 Search for extraterrestrial intelligence, or SETI is the collective term for the search for extraterrestrial life. The need to communicate with extraterrestrial organisms is as old as the idea of the possibility of life existing „somewhere out there." However, with advances

$$N = R_* f_p n_p f_l f_i f_c L,$$

where, N represents the number of civilizations that are detectable by electromagnetic emission, R_* is the formation rate of the corresponding stars, f_p is the fraction with planetary systems, n_p is the number of planets in such systems suitable for life, f_1 is the fraction at which life actually develops, f_i is the fraction with the planets on which intelligent life can occur, f_c is the fraction of the planets that developed a civilization with the necessary signal detection technology and L is the time frame during which these civilizations emit such signals (Ćirković, 2004).

A glance at the literature shows that Drake's equation has received much more criticism than scientific and reasoned support. On the other hand, with the same insight, we come to a vast number of papers and research that show that the promotion of scientific debate, as Drake's original motivation, was a great success. We can talk about the usefulness of the Drake equation from different perspectives. Some have taken it seriously, others as a parody. Nevertheless, the fact is that Drake's equation has a wide application (or at least attempts to do so) outside the field in which it originated. It also has an application within the subject of our work, which we will present in the following subsection.

4.1 Drake equation and the lightning-triggered electroporation

In the second chapter, we presented a detailed theoretical and experimental contribution by Tadej Kotnik on lightning-triggered electroporation as a possible contributor to HGT. At the same time we previosly hinted at the application of the Drake equation outside the astrophysical domain from which it originated. One such application is James Weaver's (2013) attempt to replicate Kotnik's assumptions and calculate the probability of HGT as a lightning-triggered process.

At the risk of repeating ourselves, it is necessary to highlight the main parameters of Kotnik's considerations to understand the basis on which Weaver formulated a specific application of the Drake equation. The primary phenomenon is the generation of electric current density J in the aquatic environment. A widely distributed J results in electric fields that can electroporate prokaryotic membranes. Based on pulse strength, duration, and repeatability, there are three possible outcomes: lethal, HGT possible, and no EP effect (Weaver, 2013: 374). The near-ideal candidate for the necessary energy source is lightning, which has not changed significantly over time in evolutionary terms. Other parameters include global lightning and the

---

in technology and means of electromagnetic radiation detecting distant civilizations, it has also become somewhat institutionalized (Dick, 2020). As a curiosity, we can highlight Nikola Tesla's unusual idea in 1896 to make contact with Mars using a wireless electrical transmission system.

evolutionary time frame in which HGT was important. Weaver finally arrives at the following reformulation of the Drake equation:

$$N_{HGT} = R_{LST} n_{BAC} V_{EPZ} f_{EPT} f_{SIG} L_{HGT},$$

where N_{HGT} is the total number of evolutionarily significant changes due to lightning, R_{LST} is the rate of lightning strikes on the early Earth, n_{BAC} is the bacterial concentration in the environment, V_{EPZ} is the volume of the zone with successful electroporation, f_{EPT} is the fraction of successfully electroporated bacteria, f_{SIG} is the test value for a fraction of porous cells undergoing evolutionarily significant HGT, and L_{HGT} is the time during which evolutionarily significant changes occur (*Ibid*: 375).

Attempts to salvage Drake's equation as plausible and practical consist of conceptual-statistical arguments. Drake's equation is a mere average rate, but it can help identify essential parameters and highlight sources of uncertainty in the calculation of some evolutionary problems. This case can help isolate important processes for the lightning-triggered HGT and the rather sensitive N_{HGT} value as a source of uncertainty. From our case's average rate statistic nature, it is clear that the total number of evolutionarily significant changes caused by lightning will be pretty large. Weaver correlates N_{HGT} and f_{SIG} to reduce the enormous number of evolutionary changes and counterargument himself. The test value for a fraction of porous cells participating in evolutionarily significant HGT would be really-really small for only a few significant HGT occurrences (Weaver, 2013: 375). Furthermore, since experiments readily show how electroporation triggered by lightning can cause HGT, one can conclude that N_{HGT} remains large valued. If there is no conceptual problem (which we will discuss in the next chapter) with the choice of parameters for applying Drake's equation, then Weaver successfully shows a high probability of the influence of lightning on the HGT emergence.

4.2 Drake equation and lightning-triggered prebiotic synthesis

In the previous part of the paper, we explained the symmetrical similarities between the origin of chemical and biological evolution. We presented Kotnik's experimental and theoretical arguments for lightning-triggered HGT and Navarro-González's considerations on the influence of volcanic lightning on the occurrence of prebiotic synthesis. We then presented Weaver's attempt to apply the Drake equation to parameters that Kotnik elaborated as significant for lightning-triggered HGT. The similarity between Kotnik and Navarro-González in isolating lightning as a potential contributor to evolution and life on the early Earth and Weaver's application of the Drake equation in calculating lightning-triggered HGT probabilities provide an argumentative basis for applying the Drake equation to the origin of prebiotic synthesis. Our reformulation of Drake's equation would be as follows:

$$N_{PBS} = R_{VLS} n_{DNC} n_{PVS} V_{SOM} f_{NFX} f_{PHR} f_{SIG} L_{PBS},$$

where N_{PBS} is the total number of evolutionarily significant changes caused by volcanic lightning, R_{VLS} is the rate of volcanic lightning strikes on the early Earth, n_{DNC} is the concentration of molecular nitrogen in the early atmosphere and n_{PVS} is phosphite saturation in volcanic minerals, V_{SOM} is the volume zone of successful synthesis of organic molecules, f_{NFX} is a fraction of successful nitrogen fixation and f_{PHR} is a fraction of successful phosphate reduction, f_{SIG} is the test value for a fraction of molecules undergoing evolutionarily significant prebiotic synthesis, and L_{PBS} is the time during which evolutionarily significant changes occur.

We emphasized earlier that the Drake equation, in statistical terms, is just an average rate. That means that N_{PBS}, with its enormous value, will be a source of uncertainty even in this case. To make our assignment a bit harsher, we will again correlate N_{PBS} and f_{SIG}. The test value for a fraction of molecules undergoing evolutionarily significant synthesis would be minimal for just a few significant volcanic lightning occurrences. Numerous theoretical considerations, followed by Navarro-González's experiments on Hawaiian volcanoes similar to the primordial magmatic environments of the early Earth, offer solidly backed-up evidence that prebiotic synthesis is caused by volcanic lightning. We can conclude that the value of N_{PBS} will remain relatively high and that the Drake equation also proved successful in solving the evolutionary problems of our case.

However, our „Drake" chapter is under a big „if". The ground of our „probabilistic project" lies on a conceptual slide of proper selection of parameters relevant to the origin of life on the early Earth and efficient determination of the sources of uncertainty. A thorough examination and methodological tuning are necessary prerequisites for calculating the probability of the influence of lightning's impact on chemical and biological evolution. In the next chapter, we will address these conceptual obstacles and attempt to argue that the problem of probability in the life sciences is fundamentally a philosophical issue and not a purely quantitative project.

5. Concluding remarks

In the previous chapter, we presented the basic features of the Drake equation (DE), Weaver's application of DE to the probability of HGT as a lightning-triggered process, and our application of DE to the probability of prebiotic synthesis as a volcanic lightning process. We highlighted the conceptual elements in DE, such as selecting relevant parameters for the emergence of chemical and biological evolution on the early Earth and determining the sources of uncertainty in calculating probabilities in the life sciences. In this chapter, to further clarify these conceptual hurdles, we will introduce the Miller-Urey experiment and Hess's (2021) estimation of the amount of *fulgurite* on the early Earth. Finally, we will show why relevant parameters and sources of uncertainty are problems that philosophers can answer and provide furder guidelines for using different types of probabilities in life sciences.

5.1 About parameters: Miller-Urey experiment

Several recurring phenomena link Kotnik's and Navarro-González's research. In addition to lightning, which both consider a cause of evolution, there is also part of their research on fatty acids and lipid bilayers and a reference to the Miller-Urey experiment, which is a vital starting point for both authors.

Stanley Miller, a chemist at the University of Chicago, conducted the first experiment in the 1950s to produce amino acids and protein building blocks from inorganic molecules and electricity (McColl, 2013). Miller's students discovered many new organic molecules based on their mentor's experiment and showed that the standard experiment performed by Miller, although never published, offered the best guidance to the origin of life on Earth before 4 Gyr.

The classic Miller-Urey experiment uses a mixture of gases and water in a proportion that Miller assumed was present on the early Earth. This mixture was later subjected to a specific temperature and electrical current fluctuations to simulate a lightning strike. In this way, Miller could generate and identify five different amino acids (Ibid: 207–210). The secondary setting of this experiment is called the „volcanic apparatus," in which new 22 amino acids were generated and identified (Parker et al., 2014).

The volcanic apparatus differs only in minor details from the classical experiment, although these details make a big difference. Narrowing one of the glass tubes increases the flow of water vapor through which the electrical current flows. This slight variation and reconfiguration of the experiment results in a more decadent combination of amino acids and produces new amino acids that had not been discovered in any other simulated early Earth experiment. In addition, many of the newly discovered amino acids have hydroxyl groups, making them more reactive and prone to forming new molecules over more extended periods (*Ibid*: 2–5).

One of the main criticisms of Miller's experiment is that he did not use all the relevant parameters related to the atmosphere of the early Earth (Ibid: 5–8). The initial conditions in Miller's experiments did not simulate the entire surface of the early Earth, so replication is questionable. From this, it is clear why the conceptual hurdle in selecting relevant parameters is fundamental to research in the life sciences. A small error can lead to significant discrepancies in results and replication. However, the case of the Miller-Urey experiment is fascinating because it was this oversight that steered research in an entirely new direction.

Although it is questionable whether Miller's experiments faithfully replicate the „entire" surface of the early Earth, it undoubtedly simulates conditions that could be found in some smaller regions of the planet. Miller's ratio of gasses to water could be emitted from many volcanoes present on the early Earth at that time. The necessary energy source would be the volcanic lightning accompanying magmatic eruptions. A little extra water vapor in the „volcanic apparatus" makes a big difference, which brings us to Navarro-González's research. In the same way water vapor deflects amino acids from sparks before they react and form other compounds, volcanic ash and gas clouds quickly remove organic molecules from the reaction zone. Finally, we conclude the Miller-Urey story on the conceptual importance of selecting relevant parameters by pointing out that new modifications to the volcanic apparatus could be a good simulation indicator of life conditions on early Mars and Titan. At the same time, current developments in instruments and technologies could provide insights into amino acids beneath the surface of the Red Planet (Petrescu et al., 2018).

5.2 About uncertainty: estimating fulgurite amounts

We have noted that the advantage of the Drake's equation is that it highlights the source of uncertainty in calculating probabilities in evolutionary problems. However, there are other types of probabilities that we can use in the life sciences that can even more successfully locate the sources of uncertainty and highlight the relevance of some parameters. To consider probabilities as an alternative to the Drake's equation, we will first introduce a calculation that goes back to early Earth and is not a complex average rate subject to hanging uncertainties.

In the part of our paper devoted to volcanic lightning as a possible contribution to the origin of chemical evolution, we mentioned the importance of the elements nitrogen and phosphorus. In order to enter into prebiotic synthesis, phosphorus must be reduced. Sources of phosphorus on the early Earth could be found in the form of the natural mineral *apatite* but also the form of the unusual meteorite mineral *schreibersite*. One of the arguments in favor of the influence of lightning on the formation of reactive phosphorus is the mineral *fulgurite*, also called „petrified lightning". Inside *fulgurite*, a hollow glass tube formed in quartz sand, we find significant amounts of *schreibersite*.

To determine whether lightning affected *schreibersite* formation, scientists tried to estimate the amount of phosphorus produced by lightning strikes from 4.5 Gyr, when Earth came into existence, to 3.5 Gyr when the earliest fossils evidence of the living world was found (Hess et al., 2021). To accomplish this unusual task, geologists had to estimate three things: the number of *fulgurites* formed each year, the amount of phosphorus in the rocks of the early Earth, and how much of that phosphorus was rendered usable by lightning strikes. *Fulgurites* form when lightning hits the ground. Therefore, the first step was to estimate the number of lightning strikes. This number can be determined by estimating the amount of carbon dioxide in the early Earth's atmosphere and the number of lightning strikes for different amounts of carbon dioxide. Estimating the amount of CO_2 in the atmosphere is a reliable indicator for estimating global temperature, a critical factor in estimating the frequency of thunderstorms (Hess et al., 1–3).

Using these variables, geologic models show that one hundred million to one billion lightning bolts struck the early Earth each year, with each bolt forming a *fulgurite*. In the first Gyr of Earth's history, up to a quintillion (1 ... 18 zero) *fulgurites* formed (Ibid: 4). The Hawaiian islands and volcanoes most closely represent the conditions of life on the early Earth. Therefore, the basaltic rocks of Hawaii were used to determine the average phosphorus content in rocks that were prevalent in the early Earth. By combining these factors, it was calculated that lightning strikes, on the annual level, produced more than ten tons of phosphorus that could be used for organic reactions. This means that lightning produced about as much phosphorus as meteorites, i.e., it produced all the phosphorus necessary for the origin of life on Earth (*Ibid*: 6).

Further research guidelines

We have presented specific examples where selecting relevant parameters can steer the research in a completely new direction and make significant differences in terms of results. We have also shown that determining the source of uncertainty in some evolutionary problems can be an appropriate criterion for selecting the type of probability that we will apply. To conclude our paper, we will present several criticisms of the applicability of the Drake equation to evolutionary problems and probabilities of cosmological proportions, and suggest some other potential candidates that would be more appropriate for the temporal context of Gyr magnitude.

One of the more interesting critiques of the Drake equation is Fermi's famous paradox[10] that asks, „How many piano tuners are there in Chicago?" (Prantzos, 2013). The parameters included in the average rate callculations are: five million people live in Chicago, on average two people live in a household, one in twenty households owns a piano, pianos need to be tuned once a year, it takes the piano tuner two hours to travel to a household and tune the piano, with each piano tuner working eight hours per day, five days per week, and 50 weeks per year. Based on these parameters and the estimated rate, there are 125 thousand piano tunings in Chicago per year or one thousand piano tunings per tuner per year. So there are 125 piano tuners in Chicago (Prantzos, 2013: 246–248).

---

10 I want to thank Professor Milan Ćirković of the Astronomical Observatory in Belgrade, who pointed me to this example and gave helpful comments on problems with the Drake equation during the *Sciences of the Origin: The Challenges of Selection Effects and Biases conference* (June 3–5, 2021).

The Drake equation is a modification of the Fermi problem in which we use the multiplication of parameters to determine the average rate of communicative civilizations in the Milky Way. The money-shoot question here is: if the abundance of civilizations can communicate within our galaxy, why have we not contacted them yet? This mesmerizing puzzle is called the *Fermi paradox*. The mere fact that the estimated number of existing civilizations capable of communication in the Milky Way is quite large and that the inhabitants of Earth, who are also capable of communication, have not yet made contact with their undersized purple neighbors (or whatever they look like) indicates that there is something wrong with this probability calculation. The issue is not the shortcomings of the formula itself but the suitability of its application to this kind of problem. Tuning a piano in Chicago is not the same as estimating the number of intelligent civilizations in the Milky Way. The application problem reduces to the lack of temporal structure and appreciation of the importance of evolutionary effects pointed out by Ćirković (2004).

In our case, things are a bit different. We cannot ask, „If the probability of lightning-triggered evolution is so great, why has evolution not occurred yet?“ Evolution has undoubtedly occurred; otherwise, we would not be so late in submitting this article to the *Belgrade Philosophical Annual*. However, the question here cannot be posted in the form of „if-then“, but „if not – then what?“. Schultheis (2020), Byrne and Johnson-Laird (2020), and Stalnaker (2021) have written about more detailed approaches to counterfactual probabilities as a possible digression of complicated statistical calculations into no less complex philosophical and sociological considerations.

The key philosophical issues of applying Drake's equation to the calculation of probability in evolutionary domains are based on the sources of uncertainty and the methodology for selecting appropriate parameters. In the Miller-Urey experiment, we showed that a small error in parameter selection could make a huge difference in the results of the phenomenon we are examining. In contrast, in an anecdotal example of a piano tuner, we showed that the average probability rate could not be applied with the same enthusiasm to the Chicago population and the vast number of habitable planets in the universe. The choice of possible types of probability calculations depends on many factors, which become especially important when dealing with data-rich contexts, cosmological timescales, and sources of uncertainty that lead to significant discrepancies in results. On the other hand, Bayesian probability theory incorporates sources of uncertainty much better into its formulas and considers research information gaps and blind spots much more carefully (Scharf and Cronin, 2016; Grimaldi and Marcy, 2018). Insights from the philosophy and sociology of science come to the fore, especially in this kind of research framework and „almost-metaphysical“ contexts. In any case, there is no shortage of questions in the life sciences, problems are still visibly unresolved and cascading, and it is up to philosophers to get into the game and try to contribute to new insights and methodological and conceptual pedantry.

References:

Benton, G. W. (1896). Louisville Filtration Experiments. In *Proceedings of the Indiana Academy of Science* (Vol. 6, pp. 63–65).

Black, R. A., & Blosser, M. C. (2016). A self-assembled aggregate composed of a fatty acid membrane and the building blocks of biological polymers provides a first step in the emergence of protocells. *Life*, *6*(3), 33.

Blokesch, M. (2016). Natural competence for transformation. *Current Biology*, *26*(21), R1126-R1130.

Browne, J. (1978). The Charles Darwin—Joseph Hooker correspondence: an analysis of manuscript resources and their use in biography. *Journal of the Society for the Bibliography of Natural History*, *8*(4), 351–366.

Byrne, R. M., & Johnson-Laird, P. N. (2020). If and or: Real and counterfactual possibilities in their truth and probability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(4), 760.

Ćirković, M. M. (2004). The temporal aspect of the Drake equation and SETI. *Astrobiology*, *4*(2), 225–231.

Dick, S. J. (2020). Consequences of Success in SETI: Lessons from the History of Science. In *Space, Time, and Aliens* (pp. 129–142). Springer, Cham.

Golberg, A. (2013). The impact of pulsed electric fields on cells and biomolecules: Comment on "Lightning-triggered electroporation and electrofusion as possible contributors to natural horizontal gene transfer" by Tadej Kotnik. *Physics of Life Reviews*, *10*(3), 382–383.

Grimaldi, C., & Marcy, G. W. (2018). Bayesian approach to SETI. *Proceedings of the National Academy of Sciences*, *115*(42), E9755-E9764.

Herzberg, C., Condie, K., & Korenaga, J. (2010). Thermal history of the Earth and its petrological expression. *Earth and Planetary Science Letters*, *292*(1–2), 79–88.

Hess, B. L., Piazolo, S., & Harvey, J. (2021). Lightning strikes as a major facilitator of prebiotic phosphorus reduction on early Earth. *Nature communications*, *12*(1), 1–8.

Kotnik, T. (2013). Lightning-triggered electroporation and electrofusion as possible contributors to natural horizontal gene transfer. *Physics of life reviews*, *10*(3), 351–370.

Kotnik, T. (2017). Lightning-triggered electroporation as a mechanism for horizontal gene transfer. *Handbook of electroporation. Springer, Cham*, 369–385.

Kumar, P., Nagarajan, A., & Uchil, P. D. (2019). Electroporation. *Cold Spring Harbor Protocols*, *2019*(7), pdb-top096271.

Lee, I. P. A., Eldakar, O. T., Gogarten, J. P., & Andam, C. P. (2021). Bacterial cooperation through horizontal gene transfer. *Trends in Ecology & Evolution*.

McCollom, T. M. (2013). Miller-Urey and beyond: what have we learned about prebiotic organic synthesis reactions in the past 60 years?. *Annual Review of Earth and Planetary Sciences*, *41*, 207–229.

Morigaki, K., & Walde, P. (2007). Fatty acid vesicles. *Current Opinion in Colloid & Interface Science*, *12*(2), 75–80.

Nag, A., Murphy, M. J., Schulz, W., & Cummins, K. L. (2015). Lightning locating systems: Insights on characteristics and validation techniques. *Earth and Space Science*, *2*(4), 65–93.

Navarro-González, R., & Segura, A. (2004). The possible role of volcanic lightning in chemical evolution. In *Origins* (pp. 139–152). Springer, Dordrecht.

Navarro-González, R., Molina, M. J., & Molina, L. T. (1998). Nitrogen fixation by volcanic lightning in the early Earth. *Geophysical Research Letters*, *25*(16), 3123–3126.

Obeso, J. R. (1997). Costs of reproduction in Ilex aquifolium: effects at tree, branch and leaf levels. *Journal of Ecology*, 159–166.

Papale, F. (2021). Evolution by means of natural selection without reproduction: revamping Lewontin's account. *Synthese*, *198*(11), 10429–10455.

Park, I. K., & Kang, D. H. (2013). Effect of electropermeabilization by ohmic heating for inactivation of Escherichia coli O157: H7, Salmonella enterica serovar Typhimurium, and Listeria monocytogenes in buffered peptone water and apple juice. *Applied and environmental microbiology*, *79*(23), 7122–7129.

Parker, E. T., Cleaves, J. H., Burton, A. S., Glavin, D. P., Dworkin, J. P., Zhou, M., ... & Fernández, F. M. (2014). Conducting miller-urey experiments. *JoVE (Journal of Visualized Experiments)*, (83), e51039.

Pasek, M. A. (2008). Rethinking early Earth phosphorus geochemistry. *Proceedings of the National Academy of Sciences*, *105*(3), 853–858.

Pasek, M., & Block, K. (2009). Lightning-induced reduction of phosphorus oxidation state. *Nature Geoscience*, *2*(8), 553–556.

Petrescu, R. V., Aversa, R., Apicella, A., & Petrescu, F. I. (2018). Nasa selects concepts for a new mission to titan, the moon of saturn. *Journal of Aircraft and Spacecraft Technology*, *2*(1), 40–52.

Prantzos, N. (2013). A joint analysis of the Drake equation and the Fermi paradox. *International Journal of Astrobiology*, *12*(3), 246–253.

Rauchfuss, H. (2008). *Chemical evolution and the origin of life*. Springer Science & Business Media.

Ruiz-Mirazo, K., Briones, C., & de la Escosura, A. (2017). Chemical roots of biological evolution: the origins of life as a process of development of autonomous functional systems. *Open Biology*, *7*(4), 170050.

Scharf, C., & Cronin, L. (2016). Quantifying the origins of life on a planetary scale. *Proceedings of the National Academy of Sciences*, *113*(29), 8127–8132.

Schultheis, G. (2020). Counterfactual probability. *Manuscript, University of Chicago*.

Sober, E. (2006). *Filozofija biologije*. Beograd: Plato.

Stalnaker, R. (2021). Counterfactuals and Probability. *Conditionals, Paradox, and Probability: Themes from the Philosophy of Dorothy Edgington*, 100.

Tiessen, H. (2008). Phosphorus in the global environment. In *The ecophysiology of plant-phosphorus interactions* (pp. 1–7). Springer, Dordrecht.

Van Gaever, S., Moodley, L., Pasotti, F., Houtekamer, M., Middelburg, J. J., Danovaro, R., & Vanreusel, A. (2009). Trophic specialisation of metazoan meiofauna at the Håkon Mosby Mud Volcano: fatty acid biomarker isotope evidence. *Marine Biology*, *156*(6), 1289–1296.

Vitousek, P. M., Hättenschwiler, S., Olander, L., & Allison, S. (2002). Nitrogen and nature. *AMBIO: A Journal of the Human Environment*, *31*(2), 97–101.

Walton, C. R., Shorttle, O., Jenner, F. E., Williams, H. M., Golden, J., Morrison, S. M., ... & Pasek, M. (2021). Phosphorus mineral evolution and prebiotic chemistry: From minerals to microbes. *Earth-Science Reviews*, *221*, 103806.

Weaver, J. C. (2013). Estimating the contribution of lightning to microbial evolution: guidance from the Drake equation. Comment on "Lightning-triggered electroporation and electrofusion as possible contributors to natural horizontal gene transfer" by Tadej Kotnik. *Physics of Life Reviews*, *10*(3), 373–376.

Weaver, J. C., & Chizmadzhev, Y. (2018). Electroporation. In *Biological and medical aspects of electromagnetic fields* (pp. 243–284). CRC Press.

Benedikt Paul Göcke
Ruhr-Universität Bochum
Benedikt.Goecke@rub.de

Original Scientific Paper
UDC: 2-2:162

THE ORIGIN OF ORIGINS A METAPHYSICAL ARGUMENT FOR THE EXISTENCE OF GOD IN THE TRADITION OF *DE ENTE ET ESSENTIA*

Creationem esse non tantum fides tenet, sed etiam ratio demonstrat.

(Not only does faith maintain that creation exists, but reason also demonstrates it.)

(Thomas Aquinas)

Abstract: *In current theology the possibility of conclusive arguments for the existence of God is largely rejected by reference to Hume or Kant. Purportedly post-metaphysical surrogates are put in place of a metaphysically founded theology, where either the existence of God may be believed in only as a rational possibility, or else a radical constructivism about the existence of God is fallen into. Nevertheless, in the following, a conclusive metaphysical argument for the existence of God in the tradition of scholastic metaphysics is formulated. It is shown that theological talk of creatio ex nihilo is only the other side of this metaphysical argument: Whoever accepts creatio ex nihilo cannot consistently deny the soundness of the argument.*[1]

Keywords: *The existence of God, de ente et essentia, creatio ex nihilo, metaphysically founded theology*

1. The Essences of Things are Abstract Entities

The metaphysical argument for the existence of God has four parts. The first part of the argument may be formulated as follows:

1) I can understand what kind of thing, a thing would be if it existed.
2) If I can understand what kind of thing, a thing would be if it existed then, regardless of whether it existed, I can understand what the essence of such a thing is.

1 I have already argued in Göcke (2021a) for the thesis that sound arguments for the existence of God are possible. I have already analyzed the general structure of philosophical arguments for the existence of God in Göcke (2020a) and Göcke (2013). That theology is necessarily a metaphysical discipline, since its primary object – God – is a metaphysical entity, I have shown in Göcke (2019) and Göcke (2021b). The present article may be understood as a continuation of Göcke (2022).

3) I can understand the essence of such a thing, whether or not it exists, only if there is the essence of such a thing, whether or not it exists.
4) If there is the essence of such a thing, regardless of whether it exists, then if it is possible for two different cognitive subjects to understand what the essence of such a thing is, then the essence of a thing is not a subjective construct.
5) It is possible for two different cognitive subjects to understand what the essence of a thing is.
6) The essence of a thing is not a subjective construct.
7) If the essence of a thing is not a subjective construct, then it is either an intersubjective construct or an abstract entity.
8) The essence of a thing is either an intersubjective construct or an abstract entity.
9) If the essence of a thing is an intersubjective construct, then if there are no intersubjective constructs, then there are no essences of things.
10) It is not the case that: if there are no intersubjective constructs, then there are no essences of things.
11) It is not the case that: the essence of a thing is an intersubjective construct.

So:

12) The essence of a thing is an abstract entity.

In propositional logic, the argument has the following form:

1)	A	Premise 1
2)	$A \supset B$	Premise 2
3)	$B \supset C$	Premise 3
4)	$C \supset (D \supset E)$	Premise 4
5)	D	Premise 5
6)	E	Premise 6, follows from premisses 1 to 5.
7)	$E \supset (F \vee G)$	Premise 7
8)	$F \vee G$	Premise 8, follows by MP ((6)/(7))
9)	$F \supset (H \supset I)$	Premise 9
10)	$\neg (H \supset I)$	Premise 10
11)	$\neg F$	Premise 11.

Therefore:

12)	G	Follows from premises 1 to 11.

Because it cannot be the case that the premises (1) – (11) are true, and the conclusion (12) is false, the argument is deductively valid. Furthermore, if the premises are true, then the argument is a deductively sound argument, which implies the truth of the conclusion.

Premise (1) is true because its negation states that it is not the case that I can understand what kind of thing, a thing would be if it existed. So the negation of (1) not only implies that I cannot understand what a Tyrannosaurus Rex or a unicorn would be, respectively, if a Tyrannosaurus Rex or a unicorn existed, but also that I cannot understand what a Higgs boson, a horse, or an electron is, as it presupposes that I can understand what kind of thing a Higgs boson, a horse, or an electron, if any such thing existed, would be. But since I can understand what a Tyrannosaurus Rex, a unicorn, a Higgs boson, a horse or an electron is, regardless of their existence, the negation of premise (1) must be rejected.

Premise (2) is conceptually true, because the concept of essence denotes only the kind to which a thing would belong, if it existed. The wholeness of a thing is its essence: As Thomas Aquinas says in *de ente et essentia* [14]: „Ex his enim que dicta sunt patet quod essentia est illud quod per diffinitionem rei significatur.“[2]

Premise (3) is true because it is an ontological implication of premises (1) – (2): It is only possible that I can understand the essence of a thing, whether or not it exists, if the essence of a thing, regardless of whether it exists, exists [$\neg\Diamond(K_s e \wedge \neg\exists x\ (x{=}e))$]. If the essence of a thing that does not exist did not exist independently of the existence of that thing, then the existence of the essence of a thing would necessarily be tied to the existence of a thing with that essence. In that case, I could only understand the essences of things that exist. But, because I can not only understand the essences of things that no

2 See: Kerr (2018: 40): "Essence is thus the principle of knowability of a thing insofar as it permits us to recognise the thing as one type of thing rather than another." This understanding of the essence of a thing goes back to the Aristotelian concept of *to ti en einai*, See: also *Metaphysics* Z, 4. In the Thomistic work, especially clearly in *de ente et essentia*, the concept of essence can be understood as follows: "Essence, then, is a principle of a finite being such that it is a necessary though not sufficient condition for the existence of such a being. Essence, then, is that through which and in which a thing has *esse*. Aquinas accordingly holds that beings are things that have essences such that beings are the type of things they are on account of the essences that they have" (Kerr (2015: 37). See also Kerr (2018: 38): „[E]ssence for Aquinas is the principle by means of which a concrete thing is the type of thing that it is and no other. It follows, then, that the essence of a thing is signified by its definition indicating what (*quid*) the thing in question is. And thus essence has commonly been taken to be synonymous with quiddity." John Locke understood the essence of a thing as follows: „[It is] the very being of any thing, whereby it is, what it is." (Locke 1975, III, III, 15). That I can understand what the essence of a thing is if it existed does not imply that I have complete knowledge of that essence *per se*, but that I can understand essential features of that essence. The fact that I can understand the essence of a thing also does not imply that errors are excluded, and that in the further course of scientific research I could not arrive at revised findings of the essence of a thing: *Nihil in intellectu nisi prius in sensu*. But it does not follow from this that it is wrong that I can understand what the essence of a thing is if it existed. Those who wish can read the argument in relation to electrons, which have a scientifically well-defined essence: We understand what an electron would be by its nature if it existed.

longer exist – I understand what a Tyrannosaurus Rex would be like if it existed – but I can also understand the essences of things that have never existed, I understand what a unicorn would be like if it existed. The negation of premise (3) is therefore false.[3]

Premise (4) is true: Although there is the essence of a thing independent of the existence of such a thing, it could be the case that the essence of such a thing only exists in the form of a subjective construct of my consciousness. If this were true then, because the contents of my consciousness are necessarily only accessible to me, there would be exactly one person who would have access to the essence of such a thing, understood as a subjective construct: *viz* me. So if it is possible that at least two different cognitive subjects can understand one and the same essence of a thing, if it is therefore possible that at least two cognitive subjects can understand what kind of thing something would be if it existed, then the essence of such a thing cannot be a subjective construct of my consciousness.[4]

Premise (5) is true: The negation of premise (5) says that it is not possible that there could be at least two cognitive subjects who could understand one and the same essence of a thing, if it existed. The negation of premise (5) thus implies that intersubjective communication about one and the same essence of a thing is impossible, since each participant in the discourse could only refer to his own subjective construct which, by definition, is distinct from the subjective constructs of the other participants in the discourse: It would not be a common topic of conversation. If it were not possible for at least two cognitive subjects to understand the essence of a thing, then we would not be able to understand what other people are trying to express when they say what the essence of a thing would be if it existed. Since such successful human communication is possible – a successful conversation, for example, about what a Tyrannosaurus Rex or a Higgs boson are according to their type is possible – premise (5) is true.

The truth of premise (6) follows from the truth of premises (1) – (5).

Premise (7) is true: It names the two plausible ways of understanding the specific form of the existence of essences of things, when they are not subjective constructs: On the one hand, they could be intersubjective constructs and, on the other hand, they could be abstract entities. There do not seem to be any further options.[5]

---

3 Our understanding of essences is therfore akin to our understanding of general concepts. See: Künne (2009: 47): „To possess the concept of (an) F is to be able to think of something as being, or as not being (an) F."

4 For further critique of constructivism, see: Boghossian (2013).

5 That that which is initially referred to as an abstract entity could be determined, in the Augustinian sense, in the course of its further ontological determination, as the idea of God, or part of the essence of God, is not excluded at this point, but cannot be called an option, because the existence of God has not yet been proven. The conclusiveness of the argument remains unaffected and the concept of the abstract entity could be

The truth of premise (8) follows validly by propositional calculus from the truth of premises (1) – (7).

Premise (9) is true: It merely explicates the logical implication that would result if it were true that the essences of things are intersubjective constructs: If an intersubjective construct in one form or another presupposes the existence of rational language users, then there could be no essences of things if there were no rational language-users.

Premise (10) is true: Although there are social constructs that would not exist without the existence of humans, it is not the case that if humans did not exist, there would be no essences of things.[6] First, convincing abductive arguments speak for the truth of the thesis that there existed things that had a certain essence long before humans existed: the sun of our planetary system, for example, i.e. that thing to which we refer when we speak of the sun of our planetary system. According to our best scientific theories the sun existed when humans did not exist, and would have existed had no humans developed, as was possible at that time.[7] So there are essences of things that are not intersubjective constructs. Second, performative philosophical arguments speak in favor of the truth of the thesis that there must be things that have a certain essence, independently of intersubjective constructional

---

replaced by the concept of an entity independent of human consciousness. The question of the ontological status of the essences of things is also known as the universal dispute and, in the case of the Augustinian option, amounts to a scholastic realism. See: Feser (2017: 102): "Scholastic realism affirms that universals exist only either in the things that instantiate them, or in intellects which entertain them. It agrees that there is no Platonic 'third realm' independent both of the material world and of all intellects. However, the Scholastic realist agrees with the Platonist that there must be some realm distinct both from the material world and from human and other finite intellects. In particular – and endorsing a thesis famously associated with Saint Augustine – it holds that universals, propositions, mathematical and logical truths, and necessities and possibilities exist in an *infinite*, *eternal*, *divine* intellect."

6 This is the thesis of metaphysical realism. See: Lowe (2008: 9): "Metaphysical realism is the view that most of the objects that populate the world exist independently of our thought and have their natures independently of how, if at all, we conceive of them." See also Alston (2002: 97–98): "[T]he kind of metaphysical realism being considered here [...] is opposed to the view that whatever there is, is constituted, at least in part, by our cognitive relations thereto, by the ways we conceptualize it or construe it, by the language we use to talk about it or the conceptual scheme(s) we use to think of it." See also Miller (2002: 13): [Metaphysical realism is the view] that the reality of the external world of concrete objects, and the truth about such reality, are what they are independently of our cognition of them. They are what they are independently of human beliefs, conceptualizations, descriptions, sentences, perceptions, conventions, languages, and so on." See Göcke (2020c) for further arguments for why theology is committed to metaphysical realism.

7 For example Nolt (2004: 71–72) argues for this thesis: „(1) The cosmos existed and had structure before we existed and (2) During some of this time, it was possible that we would never exist. From these premises it follows that (3) The cosmos has structure that would have existed even if we never had. Therefore (4) The cosmos has structure that is independent of our cognition – i.e., intrinsic structure."

achievements. Even the intersubjective constructivist, if he wishes to eschew solipsism, must presuppose that there exist, independently of him, things with certain essences which must logically precede the processes of intersubjective construction, *viz*: rational language users. Anyone who denies this and wishes to avoid solipsism must argue that rational language users have like Münchhausen, constructed themselves into being as a linguistic community, out of nothing, before they could agree on intersubjective constructs, which is simply absurd, because it is contradictory.[8]

The truth of premise (11) follows soundly from the truth of premises (1) – (10).

The truth of the conclusion (12) in turn follows validly by propositional calculus from the truth of the premises (1) – (11), by which is proven: The essence of a thing is an abstract entity.

2. The Cause of the Existence of a Thing is an already existing Thing

The second part of the argument assumes the conclusion of the first part of the argument as true, and may be formulated as an argument in propositional logic, as follows:

1) The essence of a thing is an abstract entity.
2) If the essence of a thing is an abstract entity, then the existence of a thing is something added to its essence.
3) If the existence of a thing is something added to its essence, then there is a cause for the essence of an existing thing having existence.
4) If there is a cause for the essence of an existing thing having existence, then the cause of the existence of that thing is either the essence of that thing itself, or the essence of another thing, or a pre-existing thing.
5) If the cause of the existence of this thing is the essence of this thing itself, or the essence of another thing, then it is possible that abstract entities are causally efficacious.
6) It is not possible that abstract entities are causally efficacious.

So:

7) The cause of the existence of a thing is a pre-existing thing.

In propositional calculus, the argument has the following form:

8 As Cramer (2010: 27) formulates it: „For that which has experiences cannot be an experience that is had. The subject that has experiences, perception and thinking, that which perceives and thinks, is an existence (*Dasein*) in the sense in which pre-Kantian metaphysics meant an existence (*Dasein*), or existence (*Existenz*), when it assumed that something existed in the cosmological proof.“

1)	G	Premise 1/logical consequence of Part 1.
2)	$G \supset J$	Premise 2
3)	$J \supset K$	Premise 3
4)	$K \supset (L \vee M)$	Premise 4
5)	$L \supset N$	Premise 5
6)	$\neg N$	Premise 6

Therefore:

7)	M	Conclusion, follows from premisses 1 to 6

The argument is deductively valid: The negation of the conclusion (7) leads to a logical contradiction. So, if the premises are true then, necessarily, the conclusion is also true.

The truth of premise (1) was justified in the first part of the argument.

Premise (2) is true: If the essence of a thing is an abstract entity that exists regardless of whether there is such a thing, then, if there is an ontological difference between the existence of a thing with a certain essence and the existence of the essence of that thing, the existence of a thing must be something that is ontologically added to its essence. There is a difference between the existence of a thing and the existence of its essence: For example, only existing things can be causally effective, and may be grasped with the five senses. Only an existing Tyrannosaurus Rex, not its essence, can eat me or be watched by me. Only an existing Higgs boson, but not its essence, can in principle be proven experimentally. Hence it follows that the existence of a thing must be something ontologically added to its essence. As Thomas Aquinas puts it in *de ente et essentia* [94ff]:

Quicquid enim non est de intellectu essentie vel quiditatis, hoc est adveniens extra faciens compositionem cum essentia, quia nulla essentia sine hiis que sunt partes essentie intelligi potest. Omnis autem essentia vel quiditas potest intelligi sine hoc quod aliquid intelligatur de esse suo: possum enim intelligere quid est homo vel fenix et tamen ignorare an esse habeat in rerum natura, ergo patet quod esse est aliud ab essentia vel quiditas sit ipsum suum esse.[9]

Premise (3) is true if the Principle of Sufficient Reason is assumed, since premise (3) is an application of the Principle of Sufficient Reason.[10]

The Principle of Sufficient Reason has an ontological and an epistemological component. The ontological component says that there is

9 See Feser (2017: 118): "[T]he existence of the creatures that do exist must be really distinct from their essences, otherwise one *could* know of their existence merely from knowing their essences." Although the existence of a thing is something added to its essence, and in this sense there is an ontological distinction between the existence of a thing and the existence of its essence, the existence of an existing thing in the act of its existence is nothing separable from its essence in the sense of: It is not a preexisting thing to which an essence is assigned.

10 For further analysis of the Principle of Sufficient Reason, see: Pruss (2011).

a reason for everything that happens. The epistemological component says that whoever can understand that reason is able to understand why what is happening is happening. With regard to the existence of a thing, the Principle of Sufficient Reason in its entirety states that if a thing exists, there is a cause for its existence that enables those who can understand those causes to understand why the existing thing exists. Anyone who denies the Principle of Sufficient Reason claims that there can exist things whose existence has no cause. The intelligibility of our understanding of reality excludes the existence of things whose existence has no cause. The Principle of Sufficient Reason is a basic axiom of our rational understanding of reality and constitutive of the scientific development of reality. It follows that the denial of the Principle of Sufficient Reason goes hand in hand with denial of the intelligibility of reality and the possibility of scientific development. Premise (3) can only be negated by those who are willing to give up the intelligibility of reality and the possibility of its scientific development.[11]

---

11 See: Hermanni (2017: 295): „The principle of sufficient reason seems to be a fundamental principle of our use of reason, a principle which assumes the continuous recognizability of the real, and is therefore the 'basis of all science'. But does it also have objective validity? Since it has proven itself in all previous cases, one may assume so with some justification." See: also Göcke (2020b) for an analysis of the Principle of Sufficient Reason in the work of Karl Christian Friedrich Krause and Arthur Schopenhauer: Krause argued that the Principle of Sufficient Reason itself needs a ground, and this ground can only be God himself. See: Krause (1869: 259): „Now, however, the Principle of Sufficient Reason and the Principle itself are definitely finite [...], consequently, even with regard to the Principle of Sufficient Reason, this principle must also be applied to itself: It must be asked about the reason for the Reason, about the why of the why, about the way through the through (*dem Durch des Durch*)." See: also Krause (1869: 300): „As soon as the finite spirit got there, the thought: essence or God (*Wesen oder Gott*) would become recognized as the fundamental truth, and then the thought: reason, and likewise the thought of the reason for reason; In this way the general validity of the Principle of Sufficient Reason would also become recognizable to it. Because essence (*Wesen*) is thought as being everything by and in itself (*als alles an und in sich seiend*), it is just thought of as the ground of everything. Consequently, the Principle of Sufficient Reason applies to everything finite, and it is therefore the basic thought: being or God (*Wesen oder Gott*). Also, at the same time, [it is] the ground of the authority of the general applicability of the Principle of Sufficient Reason to everything finite, according to any essence (*Wesenheit*), in any respect. Cramer argues in a similar way (2017: 53): „The fact that one has to go over from reason to reason (*von Grund zu Grund*) is to say the Principle of Sufficient Reason is in itself already related to the reason with no reason (*dem grundlosen Grund*). The Principle of Sufficient Reason is in the reason with no reason from the beginning. The cosmological proof is therefore not conclusive. It does not start from conditioned existence and first concludes unconditional existence. But that which makes it necessary to relate contingent existence to a ground of its existence is the groundless." If this is accepted, if God is therefore the *ratio essendi* of all things, and so also of all principles of knowledge, then every *ratio cognoscendi* of the existence of God, which is based on the Principle of Sufficient Reason presupposes the existence of God and is, *sensu strictu*, circular. Krause accepted this and saw the function of proofs of God as didactic and anamnetic. See Göcke (2012) for a further analysis of this position, which

Premise (4) is true because it only specifies what the cause of the existence of a thing could be: either the essence of this thing itself, the essence of another thing, or a pre-existing thing. As Thomas Aquinas says in *de ente et essentia* [127]: "Omne autem quod convenit alicui vel est causatum ex principiis naturae sue, sicut risibile in homine; vel advenit ab aliquo principio extrinseco, sicut lumen in aere ex influentia solis.

Premise (5) is true: Because the essences of things are abstract entities, abstract entities must be able to be causally efficacious, if there is any question of their being causes of the existence of a thing.

Premise (6) is true: It is ruled out that abstract entities are causally efficacious. Abstract entities, like the number 3, cannot by themselves have any causal influence on reality. Since the essences of things are abstract entities, it follows that no essence of a thing, *qua* essence, can add existence to a thing by itself. A thing would have to already exist in order to add existence to its essence, which is a contradiction and therefore ruled out as a possibility. As Thomas Aquinas put it in *de ente et essentia* [131f]:

> Non autem potest esse quod ipsum esse sit causatum ab ipsa forma vel quiditate rei, dico sicut a causa efficiente, quia sic aliqua res esset sui ipsius causa et aliqua res se ipsam in esse produceret: quod est impossibile. Ergo oportet quod omnis talis res cuius esse est aliud quam natura sua habeat esse ab alio.

Because the truth of the conclusion is implied by the truth of premises (1) – (6), and the truth of the premises has been demonstrated, it follows that it is true that the cause of the existence of a thing is a pre-existing thing.

3. There is a Thing whose Existence is Identical to its Essence

The third part of the argument assumes the truth of the conclusion of the second part, and may be formulated as follows:

1) The cause of the existence of a thing is a pre-existing thing.
2) If the cause of the existence of that thing is a pre-existing thing, then that cause is either a pre-existing thing, the existence of the essence of which is itself an additional thing, or a pre-existing thing, the essence of which is identical to its existence.
3) If the cause is a pre-existing thing, the existence of which is itself something additional to its essence, then there is either an infinite hierarchical series of causes of existence, or this cause is a pre-existing thing, the essence of which is identical to its existence.

---

will not be pursued further in the following, since it does not affect the soundess of the argument. From a Krausist point of view, it only leads to a hermeneutical re-regulation of the systematic position of the argument in the overall system of philosophy.

4) It is not the case that there is an infinite hierarchically ordered series of causes of existence.
5) The cause of the existence of a thing is a pre-existing thing, the essence of which is identical to its existence.
6) If the cause of a thing's existence is a pre-existing thing whose essence is identical to its existence, then there is a pre-existing thing whose existence is identical to its essence.

So:

7) There is a pre-existing thing whose existence is identical to its essence.

The argument has the following form in propositional logic:

1)	M	Premise 1, conclusion of the second part
2)	$M \supset (O \vee P)$	Premise 2
3)	$O \supset (Q \vee P)$	Premise 3
4)	$\neg Q$	Premise 4
5)	P	Premise 5
6)	$P \supset T$	Premise 6

Also:

7)	T	Conclusion, follows from premisses 1 to 6.

As this argument is also valid from the point of view of propositional calculus, the truth of the premises is also decisive here.

Premise (1) is true, as the second part of the argument has shown.

Premise (2) is also true: Logically there are only two possibilities: Either the existence of a pre-existing thing is something additional to its essence, or it is not the case that the existence of a pre-existing thing is something additional to its essence.[12]

If the cause of the existence of a thing is a pre-existing thing, then that thing is either one such that its existence is ontologically something additional to its essence, or it is a pre-existing thing whose existence does not add to its essence, which means that its essence is identical with its existence: If there is an existing thing the existence of which is nothing additional to its essence, then this thing must be such that its existence is identical with its essence. It would have to exist from out of itself and thus be pure being (pure existence) in the full sense of the word: *actus purus*, *esse ipse subsistens* and *esse tantum*.[13]

12 It comes into question why a thing exists in the here and now. See Feser 2017: 26): "[W] hat makes it true that the coffee exists *here* and *now*, and at any particular moment that it exists? What *keeps* it in existence?"

13 See Feser (2017: 119): "If existence were just part of what it is, then it would not need something else to cause it, and there would not be anything in it that could give it the potential to go out of existence."

Premise (3) is true: If the cause is an already existing thing, then there are only the two possibilities mentioned in premise (2): If the cause is a thing whose existence is something additional to its essence, then there is either for any cause whose existence is not identical to its essence, another pre-existing cause that causes its existence, or there is a cause whose existence is identical to its essence. If we hypothetically rule out the case that there is a cause whose existence is identical to its essence, the following case arises: If for every pre-existing thing whose existence is something to be added to its essence, there is another pre-existing thing that causes the existence of the thing that is dependent on it, then there is an infinite hierarchical series of causes of existence. The series of causes is (countably) infinite, since for every existing thing there is an existing thing that causes its existence. The series of causes is ordered hierarchically, since the existence of a thing *x* depends on the existence of a thing *y* which causes the existence of *x*, the existence of which in turn depends on a thing *z* which causes the existence of *y* and already exists, etc. So if *z* did not exist, then *y* would not be able to exist and therefore *x* would not be able to exist. The only alternative to this infinite hierarchically ordered series of causes of existence is given by the existence of a thing whose existence is identical to its essence, which therefore does not need a cause for its existence, since the question cannot be posed.[14]

Premise (4) is true: It is not the case that there can be an infinite hierarchically ordered series of causes of existence. Therefore there is no infinite hierarchically ordered series of causes of existence. First: An infinite hierarchically ordered series of causes of existence is not possible, because for every existing thing chosen arbitrarily in this series it would be true that it is not the case that it can cause the existence of a thing hierarchically subordinate to it, since every one of those existing things would have to rely for their existence on an already existing thing superior to it. It would therefore apply to all existing things of the infinite hierarchically ordered series of causes of existence that they cannot by themselves cause the existence of the things that are hierarchically subordinate to them, since they are dependent on receiving their existence from an already existing thing above them. The question now is: Can there be such an infinite hierarchically ordered series if every thing in this series receives its existence from another thing? i.e: There is no existing thing that can cause the existence of a thing without it already having received

14 See also Feser (2017. 23): "[T]he idea of a hierarchical series is best introduced by thinking in terms of a sequence whose members exist all together at a single moment of time, such as the cup which is held up by the desk which is help up by the floor. So, when it is said that such a series must have a first member, the claim is *not* that the series has to be traced back to some beginning point in the past (as the Big Bang, say). The idea is rather this. Since the desk, the floor, and the foundation have no power of their own to hold the cup aloft, the series could not exist in the first place unless there were something that *did* have the power to hold up these intermediaries, and the cup with them, *without* having to be held up *itself*."

it from another thing. This cannot be, since there is nothing that could first and foremost lend existence to such a series: Where is the existence of things that constitutes this series supposed to come from, if there is nothing that can give existence to a thing from itself? For logical reasons, there cannot be an infinite hierarchically ordered series of causes of existence, since it simply cannot *have* any existence. On the one hand, it would be as paradoxical as the assumption that there could be an infinite row of mirrors that are set up in such a way that each mirror reflects a beam of light exactly to a downstream mirror, but that there would be no light source that even emits a beam of light and can send it to a mirror. Or it would be as paradoxical as, on the other hand, the assumption that an infinite number of people could play „silent mail“ so that every person passes on the word they have heard to the next person, even though there was never a person who was even given a word in the round.[15]

The truth of premise (5) follows conclusively from the truth of premises (1) – (4).

Premise (6) is true because it expresses only an ontological implication of premise (6): Something whose existence is identical to its essence can only then be the cause of the existence of a thing whose existence is something additional to its essence, if it exists.

The logical conclusion (7) follows: There is an already existing thing, the existence of which is identical to its essence.

4. God Exists

The fourth and last part of the argument assumes the conclusions of the previous arguments:

1) There is a pre-existing thing whose existence is identical to its essence.
2) If there is a pre-existing thing whose existence is identical to its essence, then there is exactly one thing whose existence is identical to its essence.
3) If there is exactly one thing whose existence is identical to its essence, then God exists.

15 See also Hermanni (2017: 293): „Varying a thought experiment by Hume, suppose that books reproduce like organisms, so that each book is conceived by another, and that the series of books therefore goes back to infinity. Also, suppose for a moment that the existence of each book is adequately explained by reference to the procreative activity of another. Although in this case one would have a sufficient explanation for the existence of each individual book, this would in no way explain the existence of the book series. Because since the explanation of an individual book would always refer to the existence of another book, the sum of the individual explanations would leave the question open, why there are books at all and why, of all things, these strange natural books.“

So:

4) God exists.

The argument has the following form in propositional calculus:

1)	T	Conclusion of part 3, premise 1
2)	$T \supset R$	Premise 2
3)	$R \supset S$	Premise 3

Therefore:

4)	S	Conclusion, follows from premises 1 to 3.

The argument is deductively valid.

The truth of premise (1) has already been shown.

Premise (2) is true: It says that there is exactly one thing whose essence is identical to its existence. While in the case of things that are not identical to their essence, there can be different things with the same essence, there can only be one thing whose essence is identical to its existence. This can be shown by *reductio ad absurdum* of the assumption that there can be more than one thing whose existence is identical to its essence. Let us assume that a and b are each something whose essence is identical to its existence: a is *esse tantum* and b is *esse tantum*, but a is not identical to b. Then, if a and b are not identical, there must be a property F that a has but not b (or vice versa). Let us assume that the reason a and b are different is that a has the property F, but b does not (or vice versa). In this case a is different from b, but at the same time no longer such a thing whose existence is identical with its essence, since in this case it would be a thing whose existence is identical with its essence and would also be F. But this contradicts the assumption that a and b are both *esse tantum*. Hence there can only be exactly one thing whose existence is identical to its essence. So if there is one thing whose essence is identical to its existence, then there is exactly one such a thing.[16]

16 Thomas Aquinas formulates the argument in *de ente et essentia* [100ff] as follows: „[E]t hec res non potest esse nisi una et prima. quia impossibile est ut fiat plurificatio alicuius nisi per additionem alicuius differentie, sicut multiplicatur natura generis in species; vel per hoc quod forma recipitur in diversis materiis, sicut multiplicatur natura speciei in diversis individuis; vel per hoc quod unum est absolutum et aliud in aliquo receptum, sicut si esset quidam calor separatus, esset alius a calore non separato ex ipsa sua separatione. Si autem ponatur aliqua res, que sit esse tantum ita ut ipsum esse sit subsistens, hoc esse non recipiet additionem differentie, quia iam non esset esse tantum, sed esse et preter hoc forma aliqua; et multo minus reciperet additionem materie, quia iam esset esse non subsistens sed materiale. Unde relinquitur quod talis res que sit suum esse non potest esse nisi una; unde oportet quod in qualibet alia re preter eam aliud sit esse suum et aliud quiditas vel natura seu forma sua.“ For further arguments that there can only be exactly one thing whose essence is identical with its essence, see: Baldner/Carroll (1997: 66ff). See: also Feser (2017: 121): "[F]or there to be more than one thing which is *that which just is existence itself*, there would have to be something that made it

Premise (3) may be viewed as a stipulative definition, or a semantic act of baptism, because it only specifies that that whose existence is identical to its essence is called „God". Although it is of a definitional character, there is some evidence in favor of calling the *esse tantum*, the *actus purus* and the *esse ipse subsistens* "God": Traditional theology holds that all things that do not exist in and of themselves, have their being from God. This is traditionally *all* things except God. Since that whose existence is identical with its essence is that from which everything has its being, the name „God" is appropriate.

Because of the truth of the premises (1) – (3), the conclusion follows soundly by propositional logic (4): God exists.

The four-part argument outlined is a sound metaphysical argument for the existence of God. In this sense it is a successful proof of God. Since the argument is deductively valid in toto, whoever wishes to *discredit* this argument must show that the truth of the premises does not follow from their justification. It does not follow from this that the premises are false, but it would be shown that the justification given does not provide sufficient grounds for assuming the truth of the conclusion. Whoever wishes to *refute* this argument, has to show that at least one premise is false. Sufficient reasons for the falsity of at least one premise must therefore be formulated *expressis verbis*.[17]

As demonstrated, the negation of the premises at least *prima* and *secunda facie* has absurd consequences, such as: the fact that we cannot understand

---

the case that *this* instance of *that which just is existence itself* differed from that *instance*. And each such instance would, then, not really be *that which just is existence itself* after all, but rather that *that which just is existence itself* PLUS whatever the differentiating feature is. So, there really is no sense to be made of there being more than one of something which *just is* existence itself. And in that case there is no way to make sense of there being more than one of something whose essence and existence are not really distinct."

17 So it is not enough to counter a proof of God with some sweeping judgments, as Wendel (2020: 113) does: "For theoretical reason, knowledge of God is impossible, as it gets entangled in transcendental illusion and speculative dogmatism when it seeks to prove the existence of God, because knowledge of God is beyond the realm of possible experience." Why should that be true? Perhaps there are arguments for the existence of God to which Wendel's dictum applies, but as far as the argument outlined in this article is concerned, it is not true that it is entangled in "transcendental illusion" or "speculative dogmatism". That the existence of God cannot be proven, because God is not an object of possible experience and knowledge can only be obtained from objects of experience, is a mere dictum of Kantian philosophy that has long been rejected in analytical philosophy, and which Kant himself presupposed but did not justify. As Rhonheimer (2020 59) aptly put it: "Kant's Kritk der Metaphysik – his 'Critique of Pure Reason' – is a fascinating, ingeniously constructed attempt to get out of the dead ends of what he called the 'dogmatic' rationalism of German school philosophy and of the skeptical empiricism that flourished in the British Isles, but at the same time a gigantic mistake, because it completely misses its goal of a fundamental critique of metaphysics. Anyone who practices metaphysics in the classical tradition today need not feel affected in any way by Kant's critique of metaphysics, because it simply does not apply to what metaphysics was in the classical tradition and still is today."

what we are talking about when we talk about what kind of thing something would be if it existed; or that there is no essence of things independent of human consciousness; or that contingent things just exist; or that man has constructed himself into being; or that there is an infinite hierarchically ordered causal chain of causes of existence, etc. Therefore: The price for the refutation of the outlined argument for the existence of God, understood as *esse tantum*, seems simply too high for anyone to seriously wish to pay, anyone who would like to proceed on the assumption of the intelligibility of reality.[18]

5. *Creatio ex nihilo* and the Metaphysical Argument for the Existence of God

It is part of the Christian faith to understand God as the free Creator of the world *ex nihilo*.[19]

Before the divine act of creation, therefore, from a Christian point of view, there was nothing but God. The Christian faith thus implies that God did not create (ie *form*) the world from already existing things in the manner of the Platonic Demiurge.[20]

From a Christian point of view, the creation of the world is also not a one-off event in the past: The doctrine of the divine creation of the world *ex nihilo* rather leads to the thesis of *creatio continua*: Because existence is essential to no thing except God, i.e. Nothing but God could exist if God did not give them existence, it follows that the divine creation of the world continues *ex nihilo* as long as there exist things whose existence is not identical with their essence.[21]

18 As Feser (2017: 15) expresses it: "The real debate is not between atheism and theism. The real debate is between theists of different stripes – Jews, Christians, Muslims, Hindus, purely philosophical theists, and so forth – and begins where natural theology leaves off." See also Göcke (2019) and Göcke (2020a).

19 As it is formulated in the first chapter of *Dei Filius*: „Hic solus verus Deus bonitate sua et omnipotenti virtute non ad augendam suam beatitudinem, nec ad acquirendam, sed ad manifestandam perfectionem suam per bona, quae creaturis impertitur, liberrimo consilio simul ab initio temporis utramque de nihilo condidit creaturam, spiritualem et corporalem, angelicam videlicet et mundanam, ac deinde humanam quasi communem ex spiritu et corpore constitutam."

20 See Murray (2002: 94): "For the Christian, there is a world that exists and is what it is apart from all human conceptual commitments, because this world is created by an act of God. Thus there is a world that is the way it is in part because of the divine creative intentions that the world contain such-and-such kinds and such-and-such substances, simple and composite."

21 See also (Baldner/Carroll 1997: 48) on the Thomist foundations of this thought: "Aquinas spoke about the being of the creature as though it were something quite accidental to the creature, something that must be entirely caused by God. Of its own nature – that is, left completely to itself – the creature is non-being rather than being, and it must

The Christian belief in *creatio ex nihilo* and *creatio continuans*, i.e. the belief that God created the world out of nothing and that it is continuously in existence, because no thing except God is essential to existence, is only the other side of the metaphysical argument outlined above for the existence of God. The metaphysical argument developed in the tradition of *de ente et essentia* not only formulates in purely philosophical means what *creatio ex nihilo* means; that things whose essence is not identical with their existence are given existence, but also demonstrates solely through human reason that there is God, understood as *actus purus*, *esse tantum* or *esse ipse*, also that God created the world from nothing and keeps it in being at every moment of its existence. To believe that God created the world from nothing and keeps it in being, every second of its being, is therefore to know that God exists. Christian theology, which continues by proceeding from *creatio ex nihilo* and *creatio continuans*, cannot therefore deny the conclusiveness of the Thomistic argument for the existence of God without abandoning itself as a philosophically founded reflection of Christian faith.

Appendix: The Metaphysical Argument for the Existence of God

The metaphysical argument for the existence of God in the tradition of *de ente et essentia* reads, in its entirety, as follows:

a) I can understand what kind of thing a thing would be if it existed.
b) If I can understand what kind of thing a thing would be if it existed, then regardless of whether it existed, I can understand what the essence of such a thing is.
c) I can only understand the essence of such a thing, whether or not it exists, if there is the essence of such a thing, whether or not it exists.
d) If there is the essence of such a thing, regardless of whether it exists, then if it is possible for two different cognitive subjects to understand what the essence of such a thing is, then the essence of a thing is not a subjective construct.
e) It is possible for two different cognitive subjects to understand what the essence of a thing is.
f) The essence of a thing is not a subjective construct.
g) If the essence of a thing is not a subjective construct, then it is either an intersubjective construct or an abstract entity.

---

be caused by God continuously lest it return to the non-being which it properly is. It is true to say that the creature is literally nothing without the creative causality of God. Nevertheless, we must remember that the being of creatures, far from being an accident, is the ultimate perfection or actuality of the creature (In 1 Sent. 8.1.3) […] In giving being to the creature, God does not merely make the creature an extension of Himself; rather He gives the creature an inherent stability in being, i.e., a tendency to exist."

h) The essence of a thing is either an intersubjective construct or an abstract entity.
i) If the essence of a thing is an intersubjective construct, then if there are no intersubjective constructs, then there are no essences of things.
j) It is not the case: if there are no intersubjective constructs, then there are no essences of things.
k) It is not the case that: The essence of a thing is an intersubjective construct.
l) The essence of a thing is an abstract entity.
m) If the essence of a thing is an abstract entity, then the existence of a thing is something added to its essence.
n) If the existence of a thing is something additive to its essence, then there is a cause for the essence of an existing thing to add existence.
o) If there is a cause for the essence of an existing thing to have existence, then the cause of the existence of that thing is either the essence of that thing itself or the essence of another thing or a pre-existing thing.
p) If the cause of the existence of this thing is the essence of that thing itself or the essence of another thing, then it is possible that abstract entities are causally effective.
q) It is not possible for abstract entities to be causally effective.
r) The cause of the existence of this thing is a pre-existing thing.
s) If the cause of the existence of this thing is a pre-existing thing, then that cause is either a pre-existing thing, the existence of which is something additional to its essence, or a pre-existing thing, the essence of which is identical to its existence.
t) If the cause is a pre-existing thing, the existence of which is itself something additional to its essence, then there is either an infinitely hierarchical series of causes of existence, or this cause is a pre-existing thing, the essence of which is identical to its existence.
u) It is not the case that there is an infinite hierarchically ordered series of causes of existence.
v) The cause of a thing's existence is a pre-existing thing, the essence of which is identical to its existence.
w) If a thing's cause of existence is a pre-existing thing whose essence is identical to its existence, then there is a pre-existing thing whose existence is identical to its essence.
x) There is a pre-existing thing whose existence is identical to its essence.

y) If there is one pre-existing thing whose existence is identical to its essence, then there is exactly one thing whose existence is identical to its essence.
z) If there is just one thing whose existence is identical to its essence, then God exists.

So:

aa) God exists.

The argument has the following form in propositional logic:

a) A
b) $A \supset B$
c) $B \supset C$
d) $C \supset (D \supset E)$
e) D
f) E
g) $E \supset (F \vee G)$
h) $F \vee G$
i) $F \supset (H \supset I)$
j) $\neg (H \supset I)$
k) $\neg F$
l) G
m) $G \supset J$
n) $J \supset K$
o) $K \supset (L \vee M)$
p) $L \supset N$
q) $\neg N$
r) M
s) $M \supset (O \vee P)$
t) $O \supset (Q \vee P)$
u) $\neg Q$
v) P
w) $P \supset T$
x) T
y) $T \supset R$
z) $R \supset S$

Therefore:

aa) S

Bibliographie

Alston, William. 2002. "What Metaphysical Realism is not." In: William Alston (ed.): Realism and Antirealism. Cornell: Cornell University Press. 97–115.

Aquin, Thomas von. 2018. *De Ente et Essentia*. Reclam.

Baldner, Steven E., William E. Carroll. 1997. *Aquinas on Creation*. Toronto: PIMS

Boghossian, Paul. 2013. Angst vor der Wahrheit. Ein Plädoyer gegen Relativismus und Konstruktivismus. Frankfurt am Main: Suhrkamp.

Carnap, Rudolf. 1979. *Der logische Aufbau der Welt*, Frankfurt am Main: Ullstein.

Craig, William Lane. 2018. God Above All. Divine Aseity and the Challenge of Platonism. Oxford: Oxford University Press.

Cramer, Wolfgang. 2010. Gottesbeweise und ihre Kritik. Frankfurt am Main: Klostermann.

Feser, Edward. 2017. Five Proofs of the Existence of God. San Francisco: Ignatius Press.

Göcke, Benedikt Paul. 2012. *Alles in Gott? Zur Aktualität des Panentheismus Karl Christian Friedrich Krauses*. Friedrich Pustet.

Göcke, Benedikt Paul. 2013. „An Analytic Theologian's Stance on the Existence of God." In: *European Journal for Philosophy of Religion*. Vol 5(2). 129–146.

Göcke, Benedikt Paul. 2018. „Der Anspruch christlichen Glaubens auf objektive Wahrheit." In: Michael Seewald (ed.): *Glaube ohne Wahrheit? Theologie und Kirche vor den Anfragen des Relativismus*. Freiburg: Herder. 56–73.

Göcke, Benedikt Paul. 2019. „Notiz über die Unmöglichkeit einer Theologie ohne Metaphysik." In: Benedikt Paul Göcke, Christian Pelz (ed.): *Die Wissenschaftlichkeit der Theologie. Band 3: Theologie und Metaphysik* (Studien zur Theologie, Ethik und Philosophie). Münster: Aschendorff *Theologie und Metaphysik*. 423–425.

Göcke, Benedikt Paul. 2019a. „Gottesbeweise, Offenbarung und propositionaler Gehalt. Über den Glauben nachdenken mit Dei Filius." In: Julia Knop, Michael Seewald (ed.): *Das erste vatikanische Konzil. Eine Zwischenbilanz 150 Jahre danach*. Darmstadt: Wissenschaftliche Buchgesellschaft. 117–135.

Göcke, Benedikt Paul. 2020a. „Der naturwissenschaftliche Beitrag zur Frage der Existenz Gottes. Allgemeine Grenzen und das Beispiel der Anfangssingularität des Universums." In: *Frankfurter Zeitschrift für Islamisch-Theologische Studien*. Vol 5. 167–182.

Göcke, Benedikt Paul. 2020b. „Karl Christian Friedrich Krauses Einfluss auf Arthur Schopenhauers ‚Die Welt als Wille und Vorstellung'" In: *Archiv für Geschichte der Philosophie*. Online First.

Göcke, Benedikt Paul. 2020c. „Theologie als regina scientiae". In: Benedikt Göcke, Thomas Schärtl (ed.): *Freiheit ohne Wirklichkeit? Anfragen an eine Denkform*. Münster: Aschendorff. 375–405.

Göcke, Benedikt Paul. 2021a. „Realistische Metaphysik und Gottesbeweise. Reflektionen über den Kern der Debatte zwischen „analytischer" und „kontinentaler" Theologie." In: Höhn, Hans-Joachim/ Wendel, Saskia/ Reimann, Gregor/ Tappen, Julian (ed.): *Wir müssen reden. Analytische und Kontinentale Theologie im Dialog (Quaestiones Disputatae)*. Freiburg i. Br.: Herder. 18–44.

Göcke, Benedikt Paul. 2021b. „Die leere Bibliothek. Über den Begriff der Metaphysik in der Theologie" In: *Herder Korrespondenz*. 2021/3. 25–27

Göcke, Benedikt Paul. 2022. Göcke, breul ZTP

Grössl, Johannes, Matthias Remenyi. 2021. „Glückliches Scheitern. Über Grenze und Größe der Gottesbeweise." In: Concilium (2021/1). 22–36.

Hermanni, Friedrich. 2017. "Warum ist überhaupt etwas? Überlegungen zum kosmologischen und ontologischen Argument." In: Thomas Buchheim, Friedrich Hermanni, Axel Hutter, Christoph Schwöbel (ed.): Gottesbeweise als Herausfgorderungen für die moderne Vernunft. Tübingen: Mohr Siebeck. 289–315.

Kerr, Gaven. 2015. Aquinas's Way to God. The Proof in *De Ente et Essentia*. Oxford: Oxford University Press.

Krause, Karl Christian Friedrich. 1869. *Der zur Gewissheit der Gotteserkenntnis als des höchsten Wissenschaftsprinzips emporleitende Theil der Philosophie*. Prag: F. Tempsky.

Künne, Wolfgang. 2009. Conceptions of Truth. Oxford: Oxford University Press.

Locke, John. 1975 An Essay Concerning Human Understanding, ed. P. H. Nidditch, Oxford: Oxford University Press.

Lowe, E. Jonathan. 2008. "Essentialism, Metaphysical Realism, and the Errors of Conceptualism." In: Philosopohia Scientificae. Vol. 12 (1). 9–33.

Miller, Caleb. 2002. "Realism, Antirealism, and Common Sense." In: William Alston (ed.): Realism and Antirealism. Cornell: Cornell University Press. 13–25.

Murray, Michael. 2002. "The God's I Point of View." In: William Alston (ed.): Realism and Antirealism. Cornell: Cornell University Press. 79–96.

Nolt, John. 2004. "An Argument for Metaphysical Realism." In: Journal for General Philosophy of Science. Vol. 35. 71–90.

Pruss, Alexander. 2011. The Principle of Sufficient Reason: A Reassessment. Cambridge: Cambridge UP.

Rhonheimer, Martin. 2020. „Die Frage nach dem Ursprung als die Frage nach Gott." In Christoph Böhr (ed.): Metaphysik. Von einem unabweislichen Bedürfnis der menschlichen Vernunft. Dordrecht: Springer. 57–80.

Wendel, Saskia. 2020. In Freiheit glauben. Grundzüge eines libertarischen Verständnisses von Glauben und Offenbarung. Regensburg.

Bogdana Stamenković
Institute for Philosophy
Faculty of Philosophy University of Belgrade
E-mail: bogdanastamenkovic@yahoo.com

Original Scientific Paper
UDC: 113/119 Кант И.
575.822 Кант И.

NATURAL HISTORY AND VARIABILITY OF ORGANIZED BEINGS IN KANT'S PHILOSOPHY

Abstract: *This paper aims to examine Kant's views on evolution of organized beings and to show that Kant's antievolutionary conclusions stem from his study of natural history and variability of organisms. Accordingly, I discuss Kant's study of natural history and consider whether his conclusion about impossibility of knowledge about such history expands on the research of history of organized beings. Moving forward, I examine the notion of variability in Kant's philosophy, and show that his theory of organized beings relies on the preformationist conception of variability that provides limited insight into the history of organisms. I explain that Kant's endorsement of preformationism is conditioned by a lack of knowledge about the mechanism that successfully explains adaptation and transmutation of organisms leading towards the creation of new species. Finally, I sumarize the following reasons for Kant's rejection of the hypothesis of evolution: lack of cognitive ability to discover all changes of natural phenomena in different periods of time and adoption of preformationist conception of variability of organized beings. I finish off with a discussion about mechanical inexplicability of organisms and find a third reason Kant believes that the idea of evolution is only "a daring adventure on the part of reason".*

Key words: *Kant; evolution; organized beings; natural history; variability;*

In *Critique of Judgement* ([1790] 1872) Kant makes a notable statement that the hypothesis about the evolution of organisms is nothing more than "a daring adventure on the part of reason" that occurred even to the most acute scientist of the time (Kant 1872: 301).[1] As known, theory of evolution by natural selection was not formulated in Kant's time – that task was undertaken by Darwin (1890–1882) who published *The Origin of Species* in 1859. However, some scholars entertained the idea about variability of

1 All references to Kant's *Critique of Judgement* are to German edition of *Kritik der Urtheilskraft* ([1790] 1872), whilst the translation is provided by J. C. Meredith in English edition of Oxford University Press (2007c). The references to the Kant's essays used in this paper are to English editions provided by Cambridge University Press (*Metaphysical Foundations of Natural Science*, 2004; *Of the different races of human beings*, 2007a; *On the use of teleological principles in philosophy*, 2007b; *Physical Geography*, 2012).

species before Darwin,[2] and apparently Kant was one of such scholars. Kant questions the variability of species and means of their adaptation to distinct environment. However, his theory (compared to Darwin's) is often characterised as antievolutionary and has become an attractive theme for many contemporary authors.[3] In this paper, I will try to contribute to the contemporary discussion of Kant's theory of organized beings. My intention is to show how Kant's antievolutionary theory of organisms stems from his study of natural history and preformationist conception of variability of organisms. For the sake of clarity, the paper is divided into several sections. In the first part, I provide a brief overview of dominant philosophical and scientific ideas of Kant's time; these ideas undoubtedly influenced Kant's theory of nature, history, and variability of organized beings. The second section presents a discussion of Kant's study of natural history. As known, Kant believes knowledge about such history is impossible. No one can uncover *all* changes in nature as they occur in different periods of time. Does this conclusion imply the impossibility of knowledge about the history of organized beings? I believe the answer to this question depends on the interpretation of *variation* of organisms, and in third section I show that Kant adopts the preformationist conception of variation. This conception is also known as *preformationism of natural predispositions* (Cohen, 2020) and presupposes that all potential variations of an individual are already contained within it. Thus, the variability of organisms is reduced to the manifestation of certain predispositions. If my analysis is successful, it shows that Kant's preformationism enables limited knowledge about the history of organized beings. However, one should note that this preformationism neutralizes the need to discover the mechanism that explains successful adaptation and transmutation of organisms. Hence, the idea about the evolution of natural species remains a daring, but unscientific hypothesis about organized beings. As known, Darwin revisits this question and uncovers the mechanism – natural selection – that successfully explains the adaptation and evolution of natural species, and transforms Kant's daring hypothesis into scientific theory.

Kant's endorsement of preformationism of natural predispositions ties his general study of natural history to the problem of evolution of organisms and unveils several reasons Kant dismisses the idea of evolution. The first reason is the absence of adequate cognitive abilities to discover and understand *all* changes in organized beings as they happen in different instances of time. As

---

2 Considering the process of selective breeding, Comte de Buffon (1707–1788) and Erasmus Darwin (1731–1802) believed that intentionally caused changes in the species of domestic animals confirms the assumption that equally significant changes can occur in wild animals after a long period of time (Butler 1879); Lamarck's (1744–1829) *Zoological Philosophy* (*Philosophie Zoologie*) offered an intriguing view that forces of nature, and not an intelligent creator, conditioned the adaptation of plants and animals which, after a long period of time, could cause the emergence of new species (Lamarck [1809] 1963).

3 For example, Kolb (1992); Richards (2000); Weber, Varela (2002); Morris (2011); Fisher (2014); Cohen (2020).

noted above, we find the second reason in the preformationism of natural predispositions. Finally, the third reason presents the unique organization of living beings that, as Kant believes, cannot be explained by mechanical laws. I turn to this issue in the last part of the paper.

1. Philosophical and scientific ideas of Kant's time: *vitalism* vs. *reductionism*, and *preformationism* vs. *theory of epigenesis*

We can distinguish several dominant philosophical and scientific ideas of Kant's time. On the one hand, this period is marked by the development of vitalism and reductionism; on the other, there were two influential theories about embryological development – preformationism and theory of epigenesis. Undoubtedly, all these theories influenced Kant's thought about nature, history, and organized beings.

Vitalism states that the cause of the unity of organized beings can be found in their specific structure and internal forces (Kolb 1992: 11). Blumenbach's theory of *Bildungstrieb*, the vital power or formative impulse, presents a good example of the development of vitalist ideas. Namely, Blumenbach (1752–1840) believes that the *Bildungstrieb* is the cause of reproduction, maintenance, and restoration of injured parts of an organism. Further, *Bildungstrieb* enables the realization of various stages of individual development (Blumenbach 1825: 49–51). In other words, Blumenbach assigns a constitutive role to the *Bildungstrieb* in the creation and development of organisms (Richards 2000: 21). Opposing vitalism, reductionists try to explain organized beings through the same principles and laws that govern the behaviour of inorganic entities (Roll-Hensen 1976: 62).

As Richards notes, we find the original idea of preformationisam in Swammerdam's works (Richards, 2000).[4] Swammerdam, namely, believes that the female seed contains a miniature, complete form of an adult individual of certain species. That being the case, embryological development consists in gradual growth or enlargement of an organism (Richards 2000: 13). Whilst preformationism insisted on the existence of an initially formed material unit within parental seed, the theory of epigenesis advocates a different idea: in the beginning of development, an individual is nothing but an unformed mass that gradually becomes articulated; its final form reveals the organism that belongs to certain species (Richards 2000: 13). Accordingly, the term "evolution" refers to the gradual formation and development of an individual species.[5]

4 Jan Swammerdam (1636–1680) was Dutch entomologist. His work showed that various phases during the life of an insect—egg, larva, pupa, and adult—are different forms of the same animal.

5 However, I use the term "evolution" in a Darwinian sense. In other cases, I will rely on the term "embryonic development of the individual."

Kant prefers the theory of epigenesis because it "regards nature as at least itself productive in respect of the continuation of the process, and not as merely unfolding something" (Kant 1872: 306–307). At the same time, he rejects vitalist view that certain fundamental forces – such as Blumenbach's *Bildungstrieb* – cause the emergence and development of living beings (Ginsborg 2006: 456).[6] Although Kant is a defender of Newton's reductionist theory, he admits that mechanical laws cannot explain the behaviour of organized beings because they significantly differ from physical, inorganic matter. Whilst physical matter is lifeless, organized beings represent matter "in which everything is mutually related to each other as end and means, which can only be thought as a *system of final causes*" (Kant 2007b: 214). Since that is the case, Kant agrees with Blumenbach that life could have not "sprung from the nature of what is void of life" (Kant 1872: 307). Apparently, Kant believes there are notable differences between physical matter and organized beings:

1) Physical matter is lifeless; organized beings present formed, *alive* matter;
2) Whilst physical entities represent systems of classical causation, organized beings are systems of *final causality* whose parts are related as means and ends.[7]

These differences lead Kant towards conclusion that life, i.e., organized beings, could not have arisen from inanimate matter. In fact, Kant argues that we cannot know how distinct types of organisms came into being: "This problem lies entirely beyond the limits of all physics possible to human beings, within which I believed that I had to hold myself" (Kant 2007b: 214). Further, differences between physical entities and organized beings affect their possible explanations. Whilst the behaviour of physical entities can be explained by mechanical laws, the explanation of organized beings must be found in teleology (Kant 2007b: 214).

I believe there are two possible interpretations of Kant's conclusion that we cannot uncover the origin of distinct types of organisms. First, "the origin of different types of organisms" can refer to the moment of actual creation, the beginning of the existence from non-existence. Given that, we can say that Kant denies the possibility of knowledge about the origin of *life* because any explanation of such an origin must begin with an already organized matter; the archaeologist of life "is obliged eventually to attribute to this universal mother an organization suitably constituted with a view to all these forms of life" (Kant 1872: 300–301). Second, it seems that "the origin of different types

---

6 Richards discusses the relation between Blumenbach's and Kant's view in his paper *Kant and Blumenbach on the Bildungstrieb: A Historical Misunderstanding* (2000).

7 Classical causation refers to the causal relation in which the cause always precedes its consequence.

of organisms" can be understood in Darwinian sense – as an emergence of new species (or types) through the evolutionary process. Thus, we can state that Kant denies the possibility of knowing the *evolution* of organisms, which remains nothing but "a daring adventure on the part of reason" (Kant 1872: 301). In the following sections, I discuss reasons that led Kant towards this conclusion. As I show, these reasons are scattered throughout his various essays. Nevertheless, they provide us with a coherent theory about the variability of organized beings.

2. Kant's natural (biological) science: nature as *physical system*

As some scholars rightly emphasize, Kant's discussion about the idea of evolution is intimately related to his study of *the physical description of the earth* (*Naturbeschreiben*) and *natural history* (*Naturgeschichte*).[8] In the explanation of the physical description of the earth, i.e., physical geography, Kant writes:

> We owe our **knowledge of nature** to **physical geography**, that is, to a **description of the earth**. In the strictest sense, there are no **experiences**, only **perceptions**, which, taken together, constitute **experience**... The physical description of the earth is thus the **first** part of knowledge of the world... Consequently, it is necessary to learn the physical description of the earth as a knowledge that can be completed and corrected with the help of experience (Kant 2012: 445–446).

The knowledge of physical geography is defined by our experience. Since our experience is confined to the present state of affairs (Kant 2012: 445), the knowledge about physical geography is also limited to the present state of natural phenomena.

Kant states that the history of nature includes "different geographies", i.e., different states of nature throughout geological time. Only if one were to describe "all events of the whole nature, as it has been through all time, then and only then would one write a real so-called natural history" (Kant 2012: 449–450). In other words

> Yet *natural history* would only consist in tracing back, as far as the analogy permits, the connection between certain present-day conditions of the things in nature and their causes in earlier times according to laws of efficient causality, which we do not make up but derive from the powers of nature as it presents itself to us now. Such

8 For example, Fisher (2007); Morris (2011); Cohen (2020). At the same time, some scholars state the value of Kant's study on physical geography and natural history is unveiled in discussions of Kant's views on history, racism, as well as the general interpretation of the *Critique of Judgment* (Louden, (2011) Clewis (2018)).

> would be a natural history that is not only possible but that also has been attempted often enough, e.g., in the theories of the earth (Kant 2007b: 197).

However, the discovery of natural history does not seem to be possible for the human intellect which, due to the absence of testimony, tends to speculate: "But there is the problem that it has to be guessed, more through experiments than by accurate testimony... But we cannot guarantee the accuracy of our information, even since the invention of writing" (Kant 2012: 450). Kant believes that the human mind cannot "glance" into the past and discover the causes of distinct states of nature without speculation. That being the case, he concludes we cannot have adequate and complete knowledge of natural history.

Kant's conclusion on the impossibility of knowledge about natural history is a consequence of certain epistemological assumptions of his philosophical system. The first and apparent assumption is the one about the *limits of possible knowledge* imposed on us by our cognitive apparatus.[9] However, the second assumption describes the knowledge as *system:*

> Moreover, we need to become acquainted with the objects of our experience **as a whole**. Thereby our knowledge is not an **aggregation** but a **system**; for in a system the **whole** is prior to the parts, while in an aggregation the **parts** have priority (Kant 2012: 446).

I noted that Kant insists that knowledge about natural history includes information about any changes nature has undergone throughout different instances of time (Kant 2012: 450). Put differently, Kant believes we must learn the natural history as a *whole*. Such knowledge is not possible if we consider nature as *logical system* in which natural phenomena are not studied as wholes, but through relations of similarity and difference of their parts – size, colour, number etc. Consequently, our knowledge becomes an aggregate of information about nature. However, the study of nature as *physical system* implies that we consider nature and its phenomena as wholes with mutual *geographical* and *historical* relations. Whilst the former type of relation is found between entities existing at the *same* time, the latter occurs amongst entities that occupy certain places in a single *causal chain* (Fisher 2007: 105–106).

We can say that knowledge of natural history requires the fulfilment of two epistemological-methodological conditions: 1) developing adequate cognitive abilities to discover *all* historical states of nature as they happen in different geological times and 2) studying nature as a *whole*, i.e., the physical system in which phenomena are connected through geographical and historical relations. The development of certain cognitive abilities seems to be mandatory for creating the possibility of knowing natural history. Without an adequate cognitive apparatus, such knowledge is not possible.

9 Kant's study about the limits of possible knowledge is well-known, which is why I believe there is no need for more detailed explanation of this study.

The fact that no human intellect can uncover and record all states of nature, as they happened in different instances of time, indicates the possession of an inadequate apparatus for discovering natural history. Apparently, the same conclusion applies to the study of the history of *organized beings*:

> If, for example, one were to consider how the various breeds of dogs descended from one line, and what changes have befallen them through all time as a result of differences in country, climate, reproduction, etc., then this would constitute a *natural history* of dogs. Such a history could be compiled for every single part of nature, for instance, on plants and so forth. But there is the problem that it has to be guessed, more through experiments than by accurate testimony. For natural history is not one whit shorter than the world itself. But we cannot guarantee the accuracy of our information (Kant 2012: 450, emphasis added).

The history of organized beings, as well as the history of nature, reflects the causal chain of their changes throughout time. Thus, if we cannot know the history of nature, then we cannot uncover the history of organized beings. In the next section, I argue this is not necessarily the case because the ability to know the history of organized beings depends on the interpretation of their *variability*.

3. Variation of organisms: compatibility of preformationism and epigenesis

Kant defines *variation* as hereditary properties of phyletic origin that "agree with their point of origination" (Kant 2007a: 85). In other words, variation is understood as hereditary *change* that indicates the origin of an organism. Further, in *Metaphysical Foundations of Natural Science* ([1786] 2004), Kant refers to the *Second Law of Mechanics*: every change in matter has some external cause (Kant 2004: 82). I mentioned that Kant believes organisms differ from lifeless physical matter. However, it should be noted that organized beings represent formed, alive *matter*, raising the question whether the Second Law of Mechanics applies to organized beings as well. The positive answer to this question indicates that every change – variation – of organized being has some external cause that affects the formation of its natural history. Accordingly, the knowledge about such cause would imply the knowledge about a specific part of individual's history.

Some textual evidence confirms such conclusion. For example, in his essay *On the Different Races of Human Beings* ([1775] 2007a), Kant writes that

> The condition of the soil (humidity or aridity), likewise that of nutrition, gradually introduce a hereditary difference or sort among animals of one and the same phylum and race, chiefly with respect

> to size, proportion of the limbs (heavy or thin), as well as natural disposition, which, while resulting in half-breeds in mixing with foreign ones, disappears over the course of few generations on other soil and with different nutrition (even without a change of climate) (Kant 2007a: 86).

Few pages later, he notes that *air* and *sun* "appear to be those causes which most deeply influence the generative power and produce an enduring development of the germs and predispositions" (Kant 2007a: 90). Finally, Kant suggests that individual physical characteristics can reveal changes of an organism:

> But the most important point, namely the *derivation of the Americans* as an incompletely adapted race, a people that long resided in the northernmost region, is quite well confirmed through the suppressed hair growth on all parts of the body except the head, through the reddish rust iron color in the colder and the darker copper color in the hotter regions of this part of the world. For the red-brown color appears (as an effect of aerial acid) to be as suitable to the cold climate as the olive-brown color (as an effect of the alkaline-bilious nature of the fluids) to the hot region (Kant 2007a: 92).

Apparently, Kant believes that external conditions of certain environment – soil, humidity or aridity, air and sun (climate), as well as nutrition – *can cause changes in physical characteristics and reproductive ability of an organism*. In other words, external conditions affect the formation of natural history of organized beings, raising once again the question whether the Second Law of Mechanics can be applied to the behaviour of organized beings. And the answer to such question is essential, particularly if we consider Kant's theory of natural history. As explained above, natural history reflects the causal chain of changes in organized beings, and Kant believes the discovery of natural history exceeds the limits of our possible knowledge. However, it seems that we *can* discover natural history *if* we can successfully apply the Second Law of Mechanics to the behaviour of organisms. Although this appears to be the case, one should not overlook that, in his *Critique of Judgement*, Kant criticise reductionist attempts to explain organized beings by referring to mechanical laws. In an essay *On the Use of Teleological Principles in Philosophy* ([1788] 2007b), he emphasizes that we can judge organisms only teleologically, not physical-mechanically, "at least as far as *human* reason is concerned" (Kant 2007b: 214). Yet, Kant seemingly adopts the opposite stance, and argues that external conditions can explain the changes of organized beings. Does Kant make the opposite statements about mechanical explicability of organisms?[10]

10 It is worth noticing that a positive answer to this question would bring Kant's naturalistic theory closer to Darwin's theory of evolution. As known, Darwin argues that new hereditary characteristics arise due to the work of natural selection; the gradual

As indicated in the previous section, the answer to these questions depends on our understanding of the concept of variation of organized beings. I noted that Kant's work is marked by dominant ideas of his time – preformationism and theory of epigenesis. Although Kant prefers the theory of epigenesis, he does not completely dismiss preformationism. Kant writes:

> In birds of the same kind which yet are supposed to live in different climates there lie germs for the unfolding of a new layer of feathers if they live in a cold climate... Since in a cold country the wheat kernel must be more protected against the humid cold than in a dry or warm climate, there lies in it a previously determined capacity or a natural predisposition to gradually produce a thicker skin... Chance or the universal mechanical laws could not produce such agreements. Therefore we must consider such occasional unfoldings as *preformed* (Kant 2007a: 89–90).

Kant apparently adopts the preformationist assumption that *all* potential changes of organism are initially contained in the germ of the original parental pair of a particular species. In other words:

> What is supposed to propagate itself must have laid *previously* in the generative power as antecedently determined to an occasional unfolding in accordance with the circumstances in which the creature can find itself and in which it is supposed to persistently preserve itself (Kant 2007a: 90, emphasis added).

Even though change represents the transition of an individual from one state to another, Kant believes that both states are already present in the same organism. Thus, a change – or variation – of an organism is the appearance of a certain property already contained within it. External factors such as climate, soil and nutrition do not influence the *development of the new*, *but the manifestation of an existing property*. Due to this – *preformationist* – conception of variation, we cannot employ the Second Law of Mechanics (and mechanical explanations) to the organized beings.[11] However, this does not imply the impossibility of knowledge about their natural history. Kant recognizes that birds migrating to colder climates develop an additional layer of feathers that protects them from freezing at low temperatures. In cold regions, a grain of wheat forms harder ligule to protect itself from humidity

---

accumulation of small changes eventually leads towards the emergence of new species of offspring that significantly differ from their parental species. In other words, the gradual accumulation of changes causes the emergence of a new Kantian strain that leads to the formation of a new species. Further, Darwin (like Kant) argues that external conditions have a significant impact on the organization of organisms and their reproductive organs (Darwin 2009: 31–32).

11 Mechanical conception of change would imply that from the existing state A (cause) arises a new, previously non-existent state B (consequence).

and cold. And, Americans inhabiting the northern parts of the continent have ”reddish rust iron” skin colour. These examples indicate that 1) the manifestation of certain predisposition represents a *variation* of organized beings, and that 2) the appearance of any predisposition is caused by *external factors of certain geographical region* (Kant 2007b: 208). In other words, the uniqueness of each geographical region enables us to judge organized beings historically as organisms with distinctive origin (Morris 2011: 178). That being the case, investigation of external factors that caused the manifestation of particular predisposition *can* provide insight into the natural history of an individual or the whole species.

Kant’s view on natural history of organized beings reveals compatibilism of theory of epigenesis and preformationism. Whilst the first studies nature as producing entity in which an organism is always a product of another organic being, the second explains the productive capacity of living beings by means of their internal purposeful predispositions (Cohen 2020: 126–127). Why does Kant adopt preformationism? Cohen states that Kant seeks naturalistic *mechanism* to explain *how* organized beings successfully adapt to distinct natural environment. In absence of such mechanism, Kant turns to preformationism of natural predispositions (Cohen 2020: 134), and I concur. However, I believe Kant’s investigation of naturalistic mechanism remains unsuccessful because it completely relies on our *present* experience that does not show the transformation of species. In long annotation in third *Critique*, Kant writes that hypothesis about the evolution of organized beings

> [It] Never ceases to be *generatio univoca* in the widest acceptation of the word, as it only implies the generation of something organic from something else that is also organic, although, within the class of organic beings, differing specifically from it. It would be as if we supposed that certain water animals transformed themselves by degrees into marsh animals, and from these after some generations into land animals. In the judgement of plain reason there is nothing a priori self-contradictory in this. But experience offers no example of it. On the contrary, as far as experience goes, all generation known to us is generatio homonyma... and a generatio heteronyma is not met with anywhere within the range of our experience (Kant 1872: 301).[12]

Kant realizes the idea about the evolution of species raises important questions that our experience cannot answer: 1) *how* do organisms adapt to distinct environment and 2) *how* members of one species produce offspring that belong

12 In fact, many of Darwin’s contemporaries emphasized that our experience does not show transformation of one species into another. In his response, Darwin wrote well-known words: ”We see nothing of these slow changes in progress, until the hand of time has marked the lapse of ages, and then so imperfect is our view into long-past geological ages, that we see only that the forms of life are now different from what they formerly were“ (Darwin 2009: 66).

to another natural species? In other words, Kant recognizes that hypothesis of evolution of species raises the question about *mechanism* that would explain the adaptation and transformation of natural species. Since our experience does not reach the knowledge about such mechanism, this hypothesis remains "a daring adventure on the part of reason" (Kant 1872: 301).

In his work *The Origin of Species*, Darwin provides the answers to aforementioned questions. He recognizes that individuals of the same species possess distinct characteristics, i.e. variations, and defines the natural selection as preservation of advantageous and rejection of harmful variations through struggle for survival (Darwin 2009: 64–65). As known, the preservation of advantageous variations enables the successful adaptation of natural species to specific environment, and the accumulation of such variation eventually leads to the creation of new, descendant species. Certainly, there was no theory in Kant's time that provided similar answers to the problem of evolution. However, I should note that Kant dismisses the idea of evolution of organized beings due to the implications of his theory of natural history. As I explained, history encompasses all changes – variations – of natural phenomena in different instances of time. Knowledge of natural history includes information about the causes of these changes, and can be discovered by an intellect who 1) possess adequate cognitive abilities to discover all historical states of nature as they happen in different geological times, and 2) studies nature as a *whole*, i.e., the physical system in which phenomena are connected through geographical and historical relations. Further, the possibility of knowledge about natural history of organized beings depends on the concept of variability. Whilst change of physical entities signify their transition from one state of affairs to another, new and previously non-existing state, variability of organized beings represents the manifestation of already existing predispositions. These predispositions appear due to an influence of certain environmental factors. Discovery of such factors should provide us with an insight into the history of natural species. Thus, knowledge about history of organized beings is possible due to preformationism of natural predispositions. However, this preformationism also neutralizes the demand for naturalistic mechanism that would explain how individuals evolve and successfully adapt to different environments. As Kant writes:

> The human being was destined for all climates and for every soil; consequently, various germs and natural predispositions had to lie ready in him to be on occasion either unfolded or restrained, so that he would become suited to his place in the world and over the course of the generations would appear to be as it were native to and made for that place (Kant 2007a: 90).

Thus, preformationism of natural predispositions demotes the idea of evolution to a degree of daring, but unscientific hypothesis about organized beings.

As indicated in the beginning of this paper, Kant's preformationism ties his general study of natural history to the problem of evolution of organisms and reveals several reasons Kant dismisses the idea of evolution. However, another reason is found in *Critique of Judgement*, where Kant explores his assumption about *mechanical inexplicability* of organisms. In the next section I review this thesis and complete my analysis.

4. *Critique of Judgement*: Mechanical Inexplicability of Organized Beings

The thesis about mechanical inexplicability of organisms is tied to Kant's study about the limits of possible knowledge and his inquiry about the status of teleology in natural science. Being a defender of Newton's mechanics, Kant believes that nature represents a system with physical-mathematical relations constructed by subjective, but necessary apperception of space and time, along with the categories of pure reason. Yet, it seems that some natural phenomena are not subjected to the operation of mechanical laws, meaning they cannot be subsumed under *a priori* principles of pure reason (Weber, Varela 2002: 104). It seems that we find such phenomena in biology, a secondary science that studies organized beings. As noted before, Kant admits organisms differ from physical, inorganic matter. Whilst physical matter is lifeless, organisms represent organized matter, a system of *final causes* in which everything is "mutually related to each other as end and means... and since therefore their possibility only leaves the teleological... mode of explanation, at least as far as human reason is concerned" (Kant 2007b: 214). Thus, in order to explain biological phenomena, Kant introduces teleology to natural (biological) science.

Kant offers a detailed analysis of the status of teleology and so-called *purposiveness of nature* in his third *Critique*. Here, he observes that nature is governed by innumerable empirical laws that cannot be subsumed under *a priori* principles of pure reason (Kant 1872: 20). In other words, Kant notices that natural phenomena are not subjected to the operation of mechanical laws, leaving us with a *contingent* unity of nature itself and our experience of natural phenomena. Although Kant admits that biological phenomena cannot be explained by mechanical laws, he states that empirical laws of nature should be recognized as *necessary*: "Such a unity is one which must be necessarily presupposed and assumed, as otherwise we should not have a thoroughgoing connection of empirical cognition in a whole of experience" (Kant 1872: 20–22). That being the case, the power of judgment – a third cognitive faculty – formulates and adopts the principle of purposiveness of nature as an *a priori* principle stipulating that "what is for human insight contingent in the particular (empirical) laws of nature contains nevertheless unity of law in the synthesis of its manifold in an intrinsically possible experience" (Kant 1872: 20–22).

The principle of purposiveness of nature grants the unity of nature itself and our experience of natural phenomena. It is a principle that compensates for our cognitive inability to provide mechanical explanations of biological phenomena. However, the principle of purposiveness is merely a *regulative* principle that lies solely in reflective judgment, meaning that

> [For] We cannot ascribe to the products of nature anything like a reference of nature in them to ends, but we can only make use of this concept to reflect upon them in respect of the nexus of phenomena in nature — a nexus given according to empirical laws (Kant 1872: 18).

In other words, the principle of purposiveness is essentially an epistemological principle formulated and adopted by reflective judgment in order to grant us a coherent experience of nature governed by empirical and apparently necessary laws. Further, it enables us to associate distinct parts of our experience of the abundant natural phenomena. Yet, since it is a regulative principle of reflective judgement, it does not provide any objective claim about natural entities as purposes. As Zuckert rightfully states, purposiveness does not represent an ontological characteristic of natural entity, but an epistemic principle that governs the unity of representations or judgments (Zuckert 2007: 10).

The principle of purposiveness is closely related to Kant's thesis about the mechanical inexplicability of organisms. Generally speaking, this thesis states that behaviour of organized beings cannot be explained by physical-mechanical laws. In *Critique of Judgement*, Kant writes:

> So where the structure of a bird, for instance, the hollow formation of its bones, the position of its wings for producing motion and of its tail for steering, are cited, we are told that all this is in the highest degree contingent if we simply look to the nexus effectivus in nature, and do not call in aid a special kind of causality, namely, that of ends (nexus finalis). This means that nature, regarded as mere mechanism, could have fashioned itself in a thousand other different ways without lighting precisely on the unity based on a principle like this, and that, accordingly, it is only outside the conception of nature, and not in it, that we may hope to find some shadow of ground a priori for that unity (Kant 1872: 232).

Kant recognizes organized beings as unique phenomena. Their uniqueness is reflected through their structure that cannot be explained by mechanical laws. Mechanical mode of explanation treats organization and behaviour of organisms as *contingency*. Matter, of which organisms are composed, can organize itself in a thousand different ways. If allowed, this contingency threatens to undermine the mechanical conception of nature where *all* phenomena should be subsumed under the operation of necessary mechanical laws (Allen 2003: 377–378). Kant resolves this problem by introducing the so-called thesis of mechanical inexplicability of organisms (Ginsborg, 2001). In this respect, it is not possible to use the mechanical laws to explain the structure and behaviour of organisms:

> Indeed, so certain it is, that we may confidently assert that it is absurd for human beings even to entertain any thought of so doing or to hope that maybe another Newton may some day arise, to make intelligible to us even the genesis of but a blade of grass from natural laws that no design has ordered (Kant 1872: 278).

Since the structure and behaviour of organisms cannot be explained by mechanical laws, Kant relies on teleology and utilises the notion of *purpose*. His *Critique of Judgement* shows that purpose has *heuristic*, not constitutive role:

> We are right, however, in drawing upon teleological judging, at least problematically, with regard to the investigation of nature; but only with a view to bringing it under principles of observation and research by analogy to the causality that looks to ends, while not pretending to explain it by this means. Thus this is an activity of reflective, not of determining, judgement (Kant 1872: 232).

The concept of purpose should explain the organization and behaviour of organisms *as if* they were intentionally created by an intelligent designer (Kant 1872: 259). The conjunction "as if" has an important role because it allows us to compare organisms to human artefacts whilst simultaneously leading to a different conclusion: although they look *as if* they were created by an intelligent designer, organisms are actually the products of nature. The notion of purpose *does not have constitutive role in creating their organizational unity*, but it does play a role in our ability to *cognize* them in a respectful way (Richards 2002: 8–9).

Kant introduces teleology to natural science to provide explanation of mechanically inexplicable organized beings. Apparently, his theory leaves us with two worlds of nature: the world of physical entities determined by physical-mechanical laws, and the world of organized beings and teleology (Kolb 1992: 26). At the first glance, these worlds seem to oppose each other. However, this is not the case. As Kant emphasizes, teleology provides an additional mode of explanation that complements mechanical research which by itself is insufficient as a method of empirical research (Kant 1871: 259). Thus, Kant believes that Newtonian mechanics and teleology are methodologically compatible. Whilst our investigation should follow mechanical principles and laws, we need to acknowledge teleological explanation of organisms.

5. Concluding remarks

My aim in this paper was to analyse various parts of Kant's naturalistic philosophy, and explore his theory of organized beings. I attempted to show how Kant's theory of natural science and organized beings relies on his study of natural history. Kant's theory of organized being is antievolutionary; he dismisses the idea about evolution of natural species as daring, but unscientific hypothesis. Kant's reasons are presented in a form of three major thesis of his philosophical system. The first one is the thesis about the absence of adequate cognitive abilities which builds upon Kant's doctrine about the limits of possible knowledge. The second thesis is formulated as preformationism of natural predispositions, whilst the third establishes the mechanical inexplicability of organisms. As shown, the last thesis extends Kant's study of purposiveness of nature where organized beings are perceived as natural purposes, systems of final causes. Since nature shows extraordinary diversity and contingency, human intellect cannot explain it by means of *a priori* concepts and principles (Van de Vijver et al 2003: 106). Thus, purposive conception of nature appears to be necessary for Kant. The principle of purposiveness of nature neutralizes the threat of the contingency of natural system. As Allen notices, without such principle, each individual entity and nature itself would be perceived as disorganized, chaotic system (Allen 2003: 379).

We can say that Kant provides convincing epistemological reasons against the idea of evolution. At the same time, he poses a challenge for sympathizers of evolutionary hypothesis of his and later time. Evolutionist need to explain how organisms successfully adapt to different environment and emergence of new species. In other words, Kant challenges evolutionists to find a mechanism that provides answers to these questions. In XIX century, Darwin accepted Kantian challenge and provided necessary answers in his famous essay *The Origin of Species*. That, however, is a theme for another paper.

Acknowledgments: I thank two anonymous referees for carefully reading and commenting on previous versions of the manuscript. Their suggestions helped to greatly improve it.

Bibliography

Allen, Barry. 2003. The Abyss of Contingency: Purposiveness and Contingency in Darwin and Kant. *History of Philosophy Quarterly* 20 (4): 373–391.

Blumenbach, I. F. 1825. *A Manual of the Elements of Natural History*. London: C. Smith, Angel Court, Strand.

Butler, Samuel. 1879. *Evolution, Old and New; Or, the Theories of Buffon, Dr. Erasmus Darwin, and Lamarck, as Compared with that of Mr. Charles Darwin*. London: Hardwicke and Bogue.

Charles, Darwin. [1859] 2009. *The Origin of Species: By Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life* (6th ed., Cambridge Library Collection – Darwin, Evolution and Genetics). Cambridge: Cambridge University Press.

Clewis, Robert. 2018. Kant's Physical Geography and the Critical Philosophy. *Epoche: A Journal for the History of Philosophy* 22 (2): 411–427.

Cohen, Alix. 2020. Kant on Evolution: A Re-evaluation. *Kant and Animals*. ed. J.J. Callanan and L. Allais, 123–135. Oxford: Oxford University Press.

Fisher. Mark. 2007. Kant's Explanatory Natural History: Generation and Classification of Organisms in Kant's Natural Philosophy. *Understanding Purpose: Kant and the Philosophy of Biology*. ed. P. Huneman, 101–121. Rochester: University of Rochester Press.

Fisher, Mark. 2014. Metaphysics and Physiology in Kant's Attitude towards Theories of Preformation. *Kant's Theory of Biology*. ed. I. Goy and E. Watkins, 25–41. Berlin/Boston: Walter De Gruyter.

Ginsborg, Hannah. 2001. Kant on Understanding Organisms as Natural Purposes. In *Kant and the Sciences*, ed. E. Watkins, 231–258. New York: Oxford University Press.

Ginsborg, Hannah. 2006. Kant's Biological Teleology and its Philosophical Significance. *A Companion to Kant*. ed. G. Bird, 455–469. Oxford: Blackwell Publishing.

Kant, Immanuel. [1775] 2007a. Of the different races of human beings. *Anthropology, History, and Education*. ed. G. Zöller and R. B. Louden, 82–97. New York: Cambridge University Press.

Kant, Immanuel. [1786] 2004. *Metaphysical Foundations of Natural Science*. New York: Cambridge University Press.

Kant, Immanuel. [1788] 2007b. *On the use of teleological principles in philosophy*. ed. G. Zöller and R. B. Louden, 192–218. New York: Cambridge University Press.

Kant, Immanuel. [1790] 2007c. *Critique of Judgement*. New York: Oxford University Press

Kant, Immanuel. [1790] 1872. *Kritik der Urteilskraft*. Berlin: L. Heimann's Verlag.

Kant, Immanuel. [1802] 2012. Physical Geography. *Natural Science*, ed. P. Guyer and A. W. Wood, 434–679. Cambridge: Cambridge University Press.

Kolb, Daniel. 1992. Kant, Teleology, and Evolution. *Synthese* 91: 9–28.

Lamarck, J. B. [1809] 1963. *Zoological Philosophy: An Exposition With Regard to the Natural History of Animals*. New York and London: Hafner Publishing Company.

Louden, Robert. 2014. The Last Frontier: The Importance of Kant's *Geography*. *Environment and Planning D: Society and Space*, *32*(3), 450–465. https://doi.org/10.1068/d6813

Morris, David. 2011. The Place of the Organism in Kantian Philosophy: Geography, Teleology, and the Limits of Philosophy. *Reading Kant's Geography*. ed. S. Elden and E. Mendieta, 173–192. New York: State University of New York Press.

Roll-Hansen, Nils. 1976. Critical Teleology: Immanuel Kant and Claude Bernard on the Limitations of Experimental Biology. *Journal of the History of Biology* 9 (1): 59–91.

Richards, J. Robert. 2000. Kant and Blumenbach on the *Bildungstrieb*: A Historical Misunderstanding. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 31 (1): 11–32.

Richards, J. Robert. 2002. *The Romantic Conception of Life: Science and Philosophy in the Age of Goethe*. Chicago and London: The University of Chicago Press.

Van de Vijver, G., Van Speybroeck, L., Vandevyvere, W. 2003. Reflecting on complexity of biological systems: Kant and beyond? *Acta Biotheoretica* 51: 101–140.

Weber, A., Varela, F. J. 2002. Life after Kant: Natural purposes and the autopoietic foundations of biological individuality. *Phenomenology and the Cognitive Sciences* 1: 97–125.

Zuckert, Rachel. 2007. *Kant on Beauty and Biology: An Interpretation of the Critique of Judgement*. New York: Cambridge University Press.

ARTICLES

Filip D. Jevtić
Mathematical Institute
SASA, Belgrade
Slobodan Vujošević
Mathematical Institute
SASA, Belgrade

Original Scientific Paper
UDC: 2-2:161/164
2-2:510.643
510 Гедел К.

LOGIC OF THE ONTOLOGICAL ARGUMENT

Abstract: *In his ontological argument Gödel says nothing about its underlying logic. The argument is modal and at least of second-order and since S5 axiom is used so it is widely accepted that the logic of the argument is the S5 second-order modal logic. However, there is a step in the proof in which Gödel applies the necessitation rule on the assumptions of the argument (see [3]). This is repeated by all of his followers (see [1] and [5]). This application of the necessitation rule can seriously harm the consequence relation of the logic of the ontological argument. It seems that the only way to preserve the modal logic S5 for the ontological argument is to assume some of its axioms in the necessitated form.*

Keywords: *Godel, ontological argument, modal logic, necessitatio*

Ontological argument

Ontological argument belongs to the family of arguments that establishes the existence of God by relying only on pure logic. Argument proceeds from the idea of God to the reality of God and was first clearly formulated by St. Anselm in his Proslogion (1077–78). Later famous versions were given by Descartes, Leibniz and others. These ontological arguments were clearly not formal, but they show a striking similarity in their form: they argue that God exists (actually, really, necessarily) if God is possible (consistent, present in our mind), and then proceed to prove that God is indeed possible. Kant use the term "ontological argument" having in mind their ontological context. Their formal and modal substance are systematized by Hartshorne (see [6]) in S5 modal propositional logic as the following theorem:

$$\Diamond Q, Q \rightarrow \Box Q \vdash \Box Q$$

Gödel expressed his version of the ontological argument in S5 second-order modal logic by deriving $\Diamond Q$ from the definition of the "God-like being" as having all „positive attributes", positive in both moral and aesthetic sense, independently of the accidental structure of the world (see [3]).

Gödel's reasons for his interest in the ontological argument are most clearly expressed in the following quote (see [5]): "I believe, already to be possible to perceive by pure reason (without appealing to the faith in any

 DOI 10.5937/BPA2235111J

religion) that the theological worldview is thoroughly compatible with all known data (including the conditions that prevail on our earth). The famous philosopher and mathematician Leibniz already tried to do this 250 years ago, and this is also what I tried in my previous letters (ontological argument)."

Consequence and proof in modal logic

There is no doubt that the propositional skeleton of the logic of Gödel's argument is the modal logic S5, or something close to it. Besides the axioms of classical propositional logic, the modal axioms of the logic S5 are

$$\Box (A \to B) \to (\Box A \to \Box B)$$
$$\Box A \to A,$$
$$\Box A \to \Box\Box A,$$
$$\Diamond\Box A \to \Box A$$

where $\Box$ is the necessity operator and, $\Diamond$ is the possibility operator defined by $\Diamond A \leftrightarrow \neg\Box\neg A$, and the inference rules are modus ponens and necessitation: from A infer $\Box A$. The last axiom is usually called the S5 axiom. This elegant and simple axiomatization of modal logic was made possible by Gödel's introduction of the necessitation rule (see [4]).

The notions of consequence and proof in modal logic are different from those in classical logic. The relation that a sentence A is a consequence of assumptions Σ can have two meanings in modal logic: A is true at each world at which the members of Σ are true, and A is true in every model in which Σ holds. The two notions are not equivalent and to distinguish between them some authors (see [2]) are using the terms *local* assumption and *local* consequence in the first, and the terms *global* assumption and *global* consequence in the second case. We shall show how this semantical distinction is reflected in the syntax of modal logic.

Assume a sentence A globally; if A is true at an arbitrary world w in some model, then A is true at every world accessible to w (since A holds in every world), so $\Box A$ holds at w. Since w is arbitrary, $\Box A$ holds at every world of a model. This means that if A is a global assumption, the necessitation rule can be applied to A. On the other hand, if we assume A locally, so that A is known to be true at a world w of some model, there is no reason to expect that $\Box A$ is also true at w. If A is a local assumption, the necessitation rule cannot apply to it.

The distinction between global and local assumptions in formal deductions comes down to the applicability or nonapplicability of the necessitation rule. A formal proof or derivation in modal logic does not allow the use of the necessitation rule to local premises and their consequences. To insure this, some authors define modal derivations as finite sequences divided in two separate parts, global and local (see [2]). The global part comes first, containing only global premises and the necessitation rule is allowed, while the local part comes second containing local premises, but without the necessitation rule.

Necessitation in Gödel's argument

It is well known that Gödel was involved in the foundation of the modern approach to modal logic. He was among the first logicians who introduced the necesitation rule that made possible the simple and elegant modal axiom systems that are in use today. But in the early 1970s, at the time Gödel wrote his note about the ontological argument, the idea of possible world semantics was new and perhaps not well appreciated. Gödel argument is modal and is presented in at least second-order logic, however the exact logic is not specified.

According to what we have told about consequence in modal logic, to allow the unrestricted use of the necessitation rule in the logic S5, we have to assume the axioms of our theories globally. In the modal logic S5, where $\Box A \leftrightarrow \Box\Box A$, assuming A globally we assume $\Box A$. Formally, this means that all axioms of the theories in the S5 logic must come in the necessitated form, i.e. with $\Box$ prefixed.

Gödel's argument is a particular version of the general ontological argument that usually means two things: to prove that God's existence is possible and to prove that God exists necessarily if He exists. If Q is the statement that God exists, this means that in the general ontological argument we have to prove $\Diamond Q$ and $Q \rightarrow \Box Q$ (Anselm's principle). It is generally accepted that with these assumptions within S5 logic one can prove $\Box Q$: the necessitation of Anselm's principle gives $\Diamond Q \rightarrow \Diamond\Box Q$, the S5 axiom gives $\Diamond Q \rightarrow \Box Q$ and the first assumption finally gives $\Box Q$ (see [1], [3], and [5]). But the use of necessitation in this proof was not correct. It seems that the only way to overcome this incorrectness is to formulate Anselm's principle in the form $\Box(Q \rightarrow \Box Q)$: it is necessary that God exists necessary if He exists.

At some point in his note, relying on axioms that are not formulated in necessitated form, Gödel presents the theorem

$$G(x) \rightarrow \Box \exists y G(y)$$

where $G(x)$ means that "x is godlike being" (see page 403 in [3]), and without any comments proceedes in the following three steps:

$$\exists x G(x) \rightarrow \Box \exists y G(y)$$
$$\Diamond \exists x G(x) \rightarrow \Diamond \Box \exists y G(y)$$
$$\Diamond \exists x G(x) \rightarrow \Box \exists y G(y)$$

In the first step the existential quantifier is introduced, the second step comes from the necessitation rule, and the third uses of the S5 axiom. Since he was able to prove $\Diamond\exists x G(x)$, Gödel finally concludes $\Box\exists y G(y)$.

Gödel, as well as his followers and commentators in this matter, say nothing about the local or global character of the ontological argument axioms. They present these axioms in the unnecessitated form (see [2], [3], and [5]), and use the necessitation rule on them and on their consequences. Perhaps they have in mind global axioms?

Acknowledgement. We express our thanks to Kosta Došen who warned us about the problem concerning the correctness of the necessitation rule in the ontological argument.

References

[1] Fitting, M., 2002, *Types, Tableaux and Gödel's God*, Dordrecht, Kluwer.

[2] Fitting, M. and Mendelsohn, R. L., 1990, *First-Order Modal Logic*, Dordrecht, Kluwer.

[3] Gödel, K., 1995, *Ontological Proof*, In *Collected Works, Volume III, Unpublished essays and lectures*, New York, Oxford University Press, pp. 403–404.

[4] Gödel, K., 1933, *An interpretation of the intuitionistic propositional calculus*, In *Collected Works, Volume I*, New York, Oxford University Press, pp. 301.

[5] Háyek, P. 2011, *Gödel's Ontological Proof and Its Variants*, In *Kurt Gödel and the Foundations of Mathematics – Horizons of Truth*, Cambridge, Cambridge University Press, 2011, pp. 307–324.

[6] Sobel, J. H., 1987, *Gödel's Ontological Proof*, In *On Being and Saying: Essays in Honor of Richard Cartwright*, Cambridge, MA, MIT Press, pp. 24–61.

CIP – Каталогизација у публикацији
Народна библиотека Србије, Београд

1

FILOZOFSKI godišnjak = Belgrade philosophical annual / editor Voin Milevski. – God. 1, br. 1 (1988)– . – Belgrade : Institute of Philosophy, Faculty of Philosophy, 1988– (Belgrade : Službeni glasnik). - 24 cm

Polugodišnje. – Glavni stvarni naslov od br. 28 (2015) Belgrade philosophical annual. – Tekst na engl. jeziku.

ISSN 0353-3891 = Filozofski godišnjak

COBISS.SR-ID 15073792

BELGRADE PHILOSOPHICAL ANNUAL 35 (1)/2022

SELECTION EFFECTS AND THE DEEP PAST

Articles