Mechanism Versus Organism: A Loosely Kantian Perspective and Its Implications for Bioengineering

18

Reto Gubelmann and Marco Toscano

"*Organisierte Wesen sind also die einzigen in der Natur, welche, wenn man sie auch für sich und ohne ein Verhältnis auf andere Dinge betrachtet, doch nur als Zwecke derselben möglich gedacht werden müssen, und die also zuerst dem Begriffe eines Zwecks, der nicht ein praktischer sondern Zweck der Natur ist, objektive Realität, und dadurch für die Naturwissenschaft den Grund zu einer Teleologie, d. i. einer Beurteilungsart ihrer Objekte nach einem besonderen Prinzip, verschaffen, dergleichen man in sie einzuführen (weil man die Möglichkeit einer solchen Art Kausalität gar nicht a priori einsehen kann) sonst schlechterdings nicht berechtigt sein würde.*"

Immanuel Kant

Summary

Building on the work of Immanuel Kant and researchers working in his tradition (but without getting lost in exegetical details that are immaterial for this volume), we discuss a notion of the living organism based on its formal teleological structure, contrast it with mechanistic objects, and draw some consequences for the notion of bioengineering. Bioengineering, in our

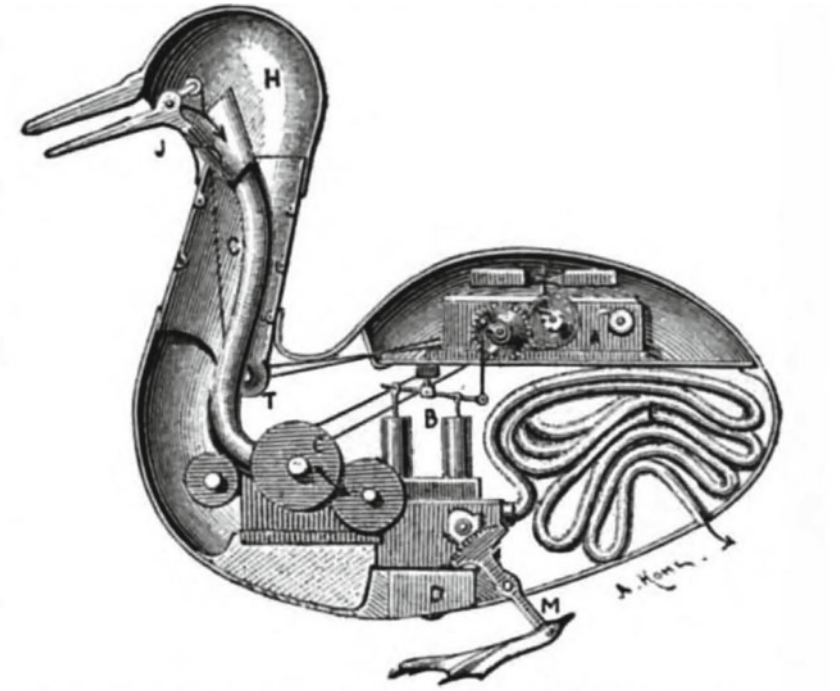
The authors have equally contributed to the work.

R. Gubelmann
University of St. Gallen, Rosenbergstrasse 30, C9000 St. Gallen, Switzerland
e-mail: reto.gubelmann@protonmail.com

M. Toscano (✉)
University of Zurich, Habsburgstrasse 25, 8037 Zürich, Switzerland
e-mail: marco.toscano@uzh.ch

N. Rezaei and A. Saghazadeh (eds.), *Thinking*, Integrated Science 7,
https://doi.org/10.1007/978-3-031-04075-7_18

understanding, is close to a *contradictio in adiecto*: either something is alive, which implies substantial autonomy in its inner functional organization, or it is engineered, which means that the engineer continuously determines the thing's functional organization.



INTERIOR OF VAUCANSON'S AUTOMATIC DUCK.

A, clockwork; *B*, pump; *C*, mill for grinding grain; *F*, intestinal tube; *J*, bill; *H*, head; *M*, feet.

Vaucanson's digesting duck (1899)
(Adapted from Wikimedia Commons, the free media repository https://en.wikipedia.org/wiki/Digesting_Duck#/media/File:Digesting_Duck.jpg).

Keywords

Bioengineering · Cyborg · Kant · Leibniz · Neural networks · Organism · Transformer

QR Code

Scanning the QR code directs you to the word cloud of the chapter that is made up of the words we discussed throughout the whole book in relation to the chapter's keyword, which is intentionally not included in the word cloud. Find the keyword and put it in place in the puzzle according to the clues provided in chapter 36. The mystery hidden in the puzzle is the quote of *THINKING: Bioengineering of Science and Art.*

Introduction

What kind of thing is a living being, and what could it mean to engineer one? In this article, we suggest that in order to appreciate the particular nature of living things as opposed to non-living things, it will be beneficial to remind ourselves how deeply our idea of the sciences as essentially investigating non-living physical matter (that is purely mechanistically structured) is rooted in the tradition of the European Enlightenment. To the concept of a mechanism, we owe our intuitions of causal explanation and technical construction of machines and other scientific knowledge applications. The idea of an organism, which we suggest is characterized by the remarkable features of teleologically structured units such as systematic growth, healing, and reproduction, is hard to square with this mechanistic conception of science and its objects. Here, we begin by detailing Kant's distinction between non-living and living beings (Sect. The Development of Kant's Conception of Living Beings). Then, we put this distinction in the context of the quest for bio-engineering (Sect. Taking Stock: What It Takes to Be Alive According to Kant). Finally, we discuss the implications of Kant's distinction for two different forms that the quest of bioengineering may take (Sect. Kant and Bioengineering: The Cases of Artificial Intelligence and the Cyborg).

The Development of Kant's Conception of Living Beings

Kant's Epistemology is the *locus classicus* for discussing the modern idea of mechanisms. Taking up loose ends from Leibniz's ontology, Kant tries to establish neat boundaries between mechanisms, organisms, and the objects of traditional theological and philosophical reason, such as God. His Critique of Pure Reason ("Kritik der Reinen Vernunft," KrV) furnishes a revolutionary metaphysical and conceptual grounding for physics, a discipline that, as a consequence of Newton's brilliant work, was taken to be the ideal of natural science. At the same time, the KrV delivers a biting critique of traditional metaphysics, undercutting many traditional aspirations of metaphysical theorizings, such as providing arguments for the existence of God or the quest to gain knowledge about the world through concepts alone.

Kant's concept of a mechanism (Sect. 2.2) fits neatly within this overall architecture. Unfortunately, the same cannot be said for the concept of living beings or organisms. It seems fair to say that Kant got caught up in a double-bind here. On the one hand, he wanted to clear the grounds from all obscure metaphysics to make room for the science of physics. On the other hand, he was aware that living beings do not fit into the metaphysics of the KrV. In this vein, the Third Critique, the Critique of the Power of Judgment ("Kritik der Urteilskraft," KdU), can be seen as partly dedicated to providing an adequate conception of organisms. In Sect. 2.3, we introduce this conception. It seems clear to us that there are tensions between this conception and the metaphysics of the KrV, but we do not elaborate on them for the purposes of this contribution.

Before beginning our discussion of the KrV, we give some background by sketching Leibniz's conception of organisms. This is important because Leibniz's thinking on this topic has been pivotal for Kant's theorizing on the matter.

Sui Generis, But Not Anti-mechanistic: Affinities Between Kant's and Leibniz' Intuitions Regarding Living Beings

In this section, we shed light on some points of contact between Kant and Leibniz regarding the status of living beings. We thereby follow the lead of Philippe Huneman [1]. He reaches the following verdict regarding the status of living beings in the KrV: he writes, "an account of the specific status of organized beings […] is not to be found in the first Critique" (p. 183). This diagnosis is closely analogous to our suggestion in the introduction to this chapter. Huneman bases this diagnosis on a fundamental disagreement between Kant and Leibniz on the notion of an organism. This disagreement, however, is framed by an underlying agreement between Leibniz and Kant on how organisms are not to be explained: by positing some organizing soul which is in principle beyond the mechanistic explanations provided by physics—which both Kant and Leibniz admire (see [1, p. 164] for Leibniz' debate with Stahl on that matter).

Leibniz tried to avoid positing something supra-mechanical by distinguishing living beings from machines only quantitatively, not qualitatively: living beings are, in contrast to machines, infinitely organized (ibid.). Ontologically speaking, living beings are machines, just very subtly organized ones, while epistemologically speaking, we will never be able to conceive a living being as a machine since our mind is finite and cannot cognize infinite structures. This, however, Huneman (p. 178) argues, is precisely where Kant correctly senses a contradiction: "according to Leibniz, one should thus consider the division [of a given organic body, RG & MT] as already achieved, albeit infinite, since it is represented in a concept (this point defines precisely what is articulation, Gliederung), and at the same time, one must conceive of it as a never-fulfilled series, since one applies here the rule of regress—which means producing an infinite division through the regress itself, the form of this division being undetermined prior to the regress" [1].

Huneman shows in this passage that there is a problem with Leibniz's notion of organisms as infinitely subtle machines. Since we have a concept of the organism, the organism's division into its parts must be achieved; otherwise, our concept would be empty. However, being finite minds, we can never achieve this division. Huneman (pp. 178–179) then shows convincingly that for Kant, Leibniz confuses divisibility of a body as being in space (which is infinitely divisible) with divisibility of a body as an organism, which cannot be infinitely divisible if we are to have a concept of it [1].

Refuting the Leibnizian distinction between organisms and machines leaves the Kant of the KrV without any means to distinguish them. However, Huneman (p. 184) then cites a passage from *Dreams of a Spirit Seer*, where Kant confesses that Stahl's strict distinction of organisms from machines has immediately convinced him [1], a conviction that Kant was unable to cash out in metaphysical terms by the time of the KrV. In addition to this conviction of Kant, Huneman (p. 85) shows that there was a general trend in Kant's days to advocate "[…] a more "immanentistic" view of organisms—a view defining the intelligibility of organisms both within the intrinsic specific forces that characterize them and through the laws of their manifestations, instead of considering them from the viewpoint of a general mechanistic scheme for understanding nature" [1].

In short, Huneman finds Kant cross-pressured already at this early point:

i. first, he rejects any kind of vitalism, the reification of what is distinctive about living beings (such as a living soul);
ii. second, he is convinced that there is a categorical distinction between living and non-living beings; and
iii. third, he finds Leibniz's attempt to draw such a distinction without falling into the reification trap unconvincing.

In the following, we delineate the traces of this tension in two of Kant's major works. After introducing how the KrV conceives mechanistic objects, we turn to the KdU to discuss Kant's conception of a living being.

Mechanistic Objects in the First Critique

In the KrV, the only causal relationships that are in nature are efficient, not final. According to the second analogy of experience, "[a]ll alterations occur after the law of cause and effect" [2].[1] If Kant's argument in the *Analytic of Principles* is conclusive, this second analogy is a requirement of our mode of cognition ("Erkenntnis"): whatever is possibly experienced by us is structured according to the *Second Analogy*. The same does not hold for final causes. No analogy of empirical thinking states that we could experience nothing unless it stands under a means-end relationship.

What does this mean for mechanistically constituted objects? What is a mechanistically constituted object? Take the example of a watch. A watchmaker designed it in Le Locle in Switzerland. She is a senior professional and has designed and assembled all of its many parts to show the correct time, date, and moon phase for centuries ahead. It is powered by a mechanism that draws energy from its bearer's wrist movement. This energy is transmitted through many wheels and sub-mechanisms to move the heads at exactly the right speed.

The goings-on within the watch can all be explained completely mechanistically: this wheel causes that wheel to turn, which in turn causes another gear to be set in motion, etc. Over and above this mechanistic explanation, one can sensibly ask certain why-questions: why is this wheel taking precisely this position in the entire mechanism? An answer could be: this is necessary to ensure that the watch is functioning properly; it shows the correct time. The watchmaker designed everything within the watch to this end.

You do not, however, need to resort to such teleological explanations. The mechanistic explanation suffices to know everything that is going on in the watch. This changes only if you try to identify the watch as a watch. What belongs to the watch? Everything that is indispensable for it to perform its function. In this sense, you need to know what it means for the watch to function properly to decide whether, say, some really fancy packaging that glimmers in the dark, powered by the mechanism of the watch itself, actually belongs to the watch or not.

Organisms as Natural Purposes

The *Third Critique* analyzes the notion of a living being by developing a determinate concept of an organism. Consider the following passage: "Organisierte Wesen sind also die einzigen in der Natur, welche, wenn man sie auch für sich und ohne ein Verhältnis auf andere Dinge betrachtet, doch nur als Zwecke derselben möglich gedacht werden müssen, und die also zuerst dem Begriffe eines Zwecks, der nicht ein praktischer sondern Zweck der Natur ist, objektive Realität, und dadurch für die Naturwissenschaft den Grund zu einer Teleologie, d. i. einer Beurteilungsart ihrer Objekte nach einem besonderen Prinzip, verschaffen,

[1] "Alle Veränderungen geschehen nach dem Gesetze der Verknüpfung von Ursache und Wirkung" (KrV A189/B232) [3].

dergleichen man in sie einzuführen (weil man die Möglichkeit einer solchen Art Kausalität gar nicht a priori einsehen kann) sonst schlechterdings nicht berechtigt sein würde" (KdU B295/A291) [4].[2]

According to this passage, organized beings are the only beings in nature which, conceived in isolation, must be thought of as purposes ("Zwecke"). The referent of the anaphoric term "derselben" is not entirely clear: it could refer back to "Dinge," "Wesen," or "Natur." The most probable reading is to take it to refer back to "Natur," since, in the further course of the sentence, natural purpose ("Zweck der Natur") appears explicitly and prominently. Thus, Kant here claims that organized beings are unique insofar as they always have to be thought of as natural purposes. By this, they furnish ("verschaffen") objective reality to the concept of a natural purpose. This, in turn, justifies the introduction of teleology into natural science, an introduction not legitimate otherwise because we cannot establish the possibility of such a causality a priori via a transcendental deduction.

Why does Kant think that the study of animals requires teleological concepts? The main reason is that in an animal, the matter's organization occurs in a way conducive to its purposes as a living being in a way that is beyond the scope of mechanistic explanations [4] (KdU B292f./A289). More precisely, what is characteristic for a natural purpose, a self-organizing being [5] (*"sich selbst organisierendes Wesen,"* KdU B292/A288), is the reciprocally productive activity of its constituents, which Kant calls organs. By comparison, artifacts like watches cannot repair themselves, let alone reproduce themselves. Living beings, however, can do precisely that because they possess a formative power [5] ("*bildende Kraft*," KdU B293/A289) to bring their matter into the form conducive to their purposes.

Thus, it is inconceivable that we find mechanistic explanations of organisms as teleological phenomena. Rather, Kant argues for the objective reality of natural purposes in a surprisingly positive way by giving a clear-cut structural model of teleological objects. Teleological objects are organized things, which means that their parts are mutually constitutive. Mutual constitutivity yields an object's special functional unity, as every part is at the same time an end and a cause for the others. Teleological objects are causally closed.

To grasp the metaphysical consequences of Kant's conception of organisms, it is helpful to contrast them with machines. The identity criteria of a concrete machine take into account objective and subjective parts. The objective part identifies the causal mechanism within the empirical world and bears existential import. The functional purpose of the specific machine can be taken to be an intention that is subjectively bestowed on it by the human mind, say, to track time correctly. No objective reality has to be accorded to the concept of purpose. In contrast, a

[2] "*Organized beings are thus the only ones in nature which, even if considered in themselves and without a relation to other things, must nevertheless be thought of as possible only as its ends, and which thus first provide objective reality for the concept of an end that is not a practical end but an end of nature, and thereby provide natural science with the basis for a teleology, i.e., a way of judging its objects in accordance with a particular principle the likes of which one would otherwise be absolutely unjustified in introducing at all (since one cannot at all understand the possibility of such a kind of causality* a priori" (KdU B295/A291) [5].

concrete organism's identity criteria exactly seem to yield such a case of objective reality, according to Kant. First of all, we are confronted with an appearance attributed to properties irreducible to mechanistic reconstruction: reproductivity, growth, and healing. That is what separates organisms from machines. Kant's crucial step is this: the specific organic properties are directly linked to the idea of purposeful articulation (Zweck, telos).

However, this last step also undercuts the KrV's doing away with any non-mechanistic structures in nature, an act that Kant thought necessary to clear the ground for physics and to eliminate all unfounded and obscurantist metaphysics. Indeed, as the KrV closely associates the reality of final causes with a creating deity, it is unclear whether the KrV's refutation of the teleological argument for God's existence still stands at this point. We will not dwell on this fascinating tension any further.

Taking Stock: What It Takes to Be Alive According to Kant

Let us again begin with Huneman's research. Huneman argues that Kat tries to improve on Leibniz by distinguishing the parts of an articulated being *conceived as a mere mereological sum (that is, as a whole that is not more than its parts) of spatial parts, a sum which is infinitely divisible,* on the one hand, and the same parts *conceived as functionally identified parts (whose division into further functional parts comes to an end)* on the other hand. For instance, the liver, conceived as an organ, can be divided almost infinitely, down to nuclear fission. The functional unity, in contrast, is lost much earlier. At some point of dividing it up, the liver will not perform its function anymore and hence cease to exist as an organ. The ten-dollar question is now how to bring these two ways of conceiving the same chunk of matter together.

Huneman (p. 182) suggests the notion of an articulated being. An articulated being is conceived as constituted by functionally identified parts. Hence, living beings, constituted by organs, would be prime examples of such articulated beings. These parts are said to be "superimposed" on the same parts conceived as mere mereological sums of spatial parts [1]. Something similar should hold for the relations between each part of the articulated being: the mereological aspect of their spatiotemporal distribution is qualitatively distinct from their functional interaction that determines each part's role in the articulated whole.

How are these two 'aspects' and their interrelations distinguished (and related) so that they reflect our experience? What is the relationship between the functional aspect and the mereological aspect of the same junk of matter? In agreement with the KrV, we experience the parts of an articulated being in the same way in which we experience unarticulated beings such as stones or houses. However, the watch or the beaver differ from the stone or the house insofar as their parts are identified and delineated not simply in mereological terms; experiencing their spatiotemporal structure corresponds to a functional one that is not entirely fixed by the relation they bear to our cognitive faculty, but also by the relation they have *to each other*. Their parts are not only parts for us, but also for each other (as it were, for

themselves). This means that the functional interrelations of spatiotemporal parts in articulated beings are not dissolvable in the mathematical and dynamical *Synthetic Principles of Pure Understanding*.

In sum, this leaves us with the following picture, one that, as we have suggested, is not without tensions in the Kantian system itself. Kant wants to draw a strong, principled distinction between living and non-living beings without falling into the trap of reification. Non-living beings such as Rudy the canindroid – a machine designed and built by human beings—are mechanistically structured; their goings-on can be fully explained by solely referencing efficient causes. Only in order to identify artifacts such as watches or canindroids one has to consider these objects' functions. Living beings, in contrast, are organisms. They are characterized by the fact that each of their parts has self-organized to function to the benefit of the other, the organism can autonomously heal itself and hence continue to function, and it can reproduce itself.

Regarding bioengineering and the artificial production of living beings, we can make use of the philosophical concepts of *organism* and *mechanism* to establish a clear-cut distinction between living things and machines. From this, it will be possible to set criteria for the goal of producing artificial life:

i. first, living beings are more than the sum of mechanistic descriptions that we can give of their going-on. They are functional units; and
ii. second, this functional unity of living beings is, in a certain sense, autonomous; that is, their ends are set within themselves. By this feature, we capture the astonishing capacities of living substances to systematically and autonomously grow, heal, and reproduce themselves.

These are abilities that human-made machines like Rudy the canindroid are not currently possessing. Our technical capacities are limited to manipulating existing living substances according to our current scientific or practical understanding of at least some of the causal mechanisms within their bodily constitution.

However, a further step seems to come within reach. Evidence suggests that advanced artificial intelligence (AI) systems have autonomously evolved functional structures that strikingly resemble organisms in Kant's understanding of the concept. In the remainder of this paper, we will flesh out teleological structures within advanced software systems, consider the notion and ethical implications of the cyborg, and discuss some implications for the task of bioengineering.

Kant and Bioengineering: The Cases of Artificial Intelligence and the Cyborg

After giving a sketch of Kant's conception of living beings as functionally structured organisms, we now draw some consequences with respect to bioengineering. Bioengineering can be understood in two senses:

i. first, in the sense of creating living organisms out of non-living objects; and
ii. second, in the sense of substantially modifying already existing organisms.

Accordingly, we first analyze the constitution of contemporary AI systems to see to what extent one could conceive them as successful instances of bioengineering in the first sense, that is, of creating living beings out of non-living objects. Second, one can speculate on the consequences of bioengineering in the sense of enhancing the cognitive and physical abilities of a traditional human being. Such speculation leads us to contemporary debates about cyborgs.

Neural Networks as Organisms?

According to Kant, a being counts as living if it is functionally constituted, can grow, repair, and reproduce itself. In this section, we will tentatively suggest that there are now AI systems that show the functional constitution emphasized by Kant's conception of an organism; furthermore, this organization has not been designed by a human engineer, as in the case of the watch; rather, it is a consequence of self-organization. In contrast, we will suggest that these systems show no signs of being able to repair or reproduce themselves autonomously.

The kind of system to be introduced is the so-called transformer, introduced initially in the field of neural machine translation (NMT) [6]. The encoder-decoder structure gives a simplified layout of an NMT system: the encoder takes in the sentence in the source language and produces a high-dimensional representation of it (as it were, a "grid of numbers," usually called the 'context'). The decoder then uses this representation to generate a sentence in the target language. This dual structure maps nicely on the tasks of so-called natural language understanding (NLU, assigned to the encoder) and so-called natural language generation (NLG, assigned to the decoder). This also means that the context, in traditional NMT architectures, has to carry all the information about the semantics of the source language sentence needed to produce a translation. The same context is used to produce all of the translated words, which has later proved to be a bottleneck in the entire system's information flow (Fig. 18.1).

True to NMT-orthodoxy, the transformer model proposed by [6] consists of an encoding part and a decoding part. Self-attention layers are central to this architecture that fundamentally distinguishes it from traditional NMT systems. This brings computational advantages. It is possible to parallelize the system massively,

Fig. 18.1 An abstract view on the general encoder-decoder structure of neural machine translation systems

that is, the processing of all words in a sentence simultaneously, while traditional architectures such as recurrent neural networks impose sequential restrictions. It also helps with long-term dependencies; as it were, the parallelized structure allows the encoder to attend to all words in the sentence simultaneously.

More importantly, scientists generally agree that the self-attention mechanism is a significant driver for the transformer's demonstrably superior performance in a wide variety of NLP tasks. In contrast to the pioneering attention mechanism from [7], the modeled self-attention mechanism connects an input sentence to itself, emphasizing these parts of the sentence that are particularly relevant for the word currently in focus. In this way, these self-attention sub-layers contextualize the words very neatly to the respective sentential context, making explicit semantic and syntactic relationships between them.

In an extended version of their paper, to be found on arXiv, the authors provide visualizations of the work done by the self-attention layers. Based on such qualitative analysis, the study [6] examines the function of self-attention layer number five and concludes that it is "apparently involved in anaphora resolution." What is exciting about this is not only that this specific attention layer does seem to be involved in anaphora resolution and pretty successful at it (NLP engineers have wrestled with this problem for decades; the transformer seems to have solved it in a matter of 3.5 days' training); what is more, this sub-layer autonomously started to assume this function during training. There was never an explicit decision on the human engineer's side that this specific part of the mechanism should be dedicated to this task.

It might be worth dwelling a little on this aspect. What the human engineer contributes is just the bare structure with large numbers of parameters, numbers whose values change during training. These values are then seeded (initialized with random values), and then a training routine starts in which the system autonomously produces output, compares this output with the correct solutions, measures its error, and tries to adapt its parameters such that this error decreases. It was epistemically (and perhaps metaphysically) impossible for the AI engineer to predict the function that self-attention layer 5 came to assume due to training.

What follows from this for the question of whether such neural network systems are alive, conceived from a Kantian perspective? Is the transformer more like a watch or more like a beaver? Is its organization merely mechanistic, or does it show the kind of purpose for itself that is characteristic for organisms according to Kant? Let us take stock. On the one hand, we submit that the spontaneous organizational capacities and the functional organization of the transformer speak in favor of considering it as an organism. Attention layer number 5 has spontaneously assumed the function of resolving anaphora. This is not a purpose that was simply set by its human designer, like in the case of the function of a specific sub-mechanism of the watch. Rather, as it were, the system autonomously evolved this functional organization. Furthermore, even if no human being ever hit upon the idea to study the functional organization of the transformer, this organization would still be real in the sense that it is an essential part of the functioning of the transformer and hence, in the end, of its demonstrably superior performance. In this sense, the functional organization is not merely a result of our conceiving the system but rather part of

the very being and functioning of the transformer – independently of our cognitive relation to it.

This functional independence raises exciting questions on the proper conception of the humans' role in charge of these systems. To the extent to which the autonomy of their functional organization increases, it becomes less convincing to conceive of humans as classical engineers and more convincing to liken them to farmers cultivating the crop. Cultivating involves a substantial responsibility in ensuring the proper circumstances for the crop to flourish, but it does not include building and maintaining the plant the way like one is building and maintaining a bridge. Plants grow by themselves; they autonomously develop their inner functional organization. Contemporary AI systems are similar to plants in this regard, which injects an aspect of cultivation to the humans' role in charge of the systems.

On the other hand, when it comes to the criteria of being able to repair itself and reproduce, we submit that one can only emphasize the true extent of the challenge that faces anyone that intends to engineer rather than cultivate organisms. While there is a research area called *neural architecture search*–essentially based on the idea that AI is best at discovering the best new neural architectures–this does not amount to reproduction in any sense close to the reproduction of living beings. As it were, their offspring is entirely disembodied: the systems conceive software, not hardware. Humans conceive both.

Even more, while there are attempts to cope with individual network cells that malfunction (which consists in adding a so-called residual connection around each cell to ensure that one cell's malfunctioning will not destroy the performance of the entire system), there is nothing close to an AI system's autonomously fabricating, say, new GPUs in case some of its hardware is destroyed. Concerning these two aspects, reproduction and healing, it is clear that the human takes the role of a classical engineer: she has to develop factories to mass-produce the hardware, and she has to get up late at night to fix a broken server.

In sum, we suggest that cutting-edge AI systems indeed have organismic structures, but they fail to self-repair and reproduce; they remain typical non-living objects. This conclusion leaves us with an interesting conceptual puzzle: these AI systems fall between the cracks of the European Enlightenment's conceptual grid to categorize beings. It takes an act of creativity, specifying our inherited vocabulary, to come to terms with this incongruence.

Programming, Autonomy, and the Ethics of Cyborg

In the last section, we have seen that the advances in AI concerning the reproduction of life are real, but also that they are very much limited in scope. This is due to structural reasons. Relative and systematic autonomy in the shape of a teleological structure appears to be within reach as software systems start to cross over from being programmed to programming themselves. At the same time, this capacity of setting one's own limits is itself limited by the fact that these capacities are not embodied. There seems no way for the teleological features of the software

to reach over onto the hardware in which they are set up. Thus, systematic growth, healing, and reproduction are inconceivable.

Still, a combination of machine bodies and autonomous programming in the future is conceivable. This would make up for the lack of embodiment and even introduce new scope to the idea of life: machine life. However, again, it would be too much to say that this new form of life would be bioengineered–not even if we were prepared to accept the breathtaking widening of the scope of the term "bios" in such a case. The more decisive point is: It would be more appropriate to say that the form of life in question is engineering itself, autonomously as it were, by supplying its hardware with ever new inorganic material according to its needs.

Even here, in this scenario of science-fiction, a familiar demarcation affirms itself. The formula–inherited by Western Enlightenment and reconstructed above–is this: If you can engineer it, it does not live, and if it is alive, you have not engineered it. The more ambitious sense of bioengineering does, therefore, seem to be beyond our grasp. We see no way of leaving behind what is being captured by the basic conceptual distinction between mechanism and organism.

We have seen that neither Kant and Leibniz nor technical progress have been able to do away with the disturbingly simple fact that living substance cannot be created out of non-living substance, as it were, *ex nihilo*. That is not true for watches and canindroids who owe their existence to an engineer, providing it with purpose and goal. No matter how deceivingly detailed, Rudy the canindroid is a heteronomous being, a being whose purpose and functional organization have been bestowed to it *from the outside*. What keeps its matter together is not Rudy itself, but an engineer and the mechanistic design she imposed. In comparison, we can call organisms relatively autonomous because what is keeping their matter together are the very organisms themselves. Being alive is to be transforming inorganic matter into a body. Thus, bioengineering will be bound to mere manipulation and enhancement, cultivating what is always already alive.

Living beings cannot come from nothing. They develop in a time-consuming and complex process. This process does not invent, in a strict sense, living matter but merely transforms it. Complex living beings are produced by reproducing already living substances. Life presupposes life, albeit in simpler forms. This is why the inception of complex life forms *ex nihilo* is a mad idea. All existing forms of life come from that long process of building and trying and failing.

The traditionally established distinction between mechanism and organism was about capturing a distinction within the concept of creation. Life was created in a way inaccessible to the craft of human beings. Within the context of the European Enlightenment, this had both epistemological and ethical implications. The limits of human technology were taken to be the limits of human understanding and its rule. The exploitation of life was bound to a supposedly natural taxonomy of beings and legitimized by its normative-hierarchical implications. Now there are questions:

- Which living being is a product of nature? And which is a product of culture?
- Which technical manipulation is, as we might say, naturally stable? And which is an aberration?

- At which point does an engineered enhancement turn its bearer into a cyborg?

Essentialism–still very active within the heritage of Western Enlightenment—is the idea that there are clear-cut demarcations between what is produced by God (or nature, however, conceived) and what is produced by man. No matter how frightening the thought: we cannot refer to essentialism to decide what form of life is performing naturally and what form of life is not. Progress in bioengineering and its technical capacities continue to blur the very basic distinctions essentialist taxonomies are built on. And, supposedly, advances in the field of AI are going to blur distinctions on an even larger scale.

The inaccessibility of life has ethical implications. By working around the irreconcilable tension between mechanism and organism, rather than suppressing it, both Leibniz and Kant allowed for an ethical element, a theological rest, within their mechanistic foundations of man's mastery over nature. This element of the ethics of the inaccessible life should be taken seriously. It can be taken up and put to use in the world of bioengineering, the world of the cyborg. Following authors like Donna Haraway, we can appreciate the forced disintegration of essentialist taxonomies and their morals pressed by the growing importance of bioengineering. In her Cyborg Manifesto, Haraway (1985) not only shows how essentialism reproduces patriarchal and colonialist forms of dominating the other, i.e., women, people of color, workers, animals, by dividing and conquering matter by norming it according to antagonistic dualisms [8]. Haraway also puts forward an image of positive and inclusive practices of difference and affinity. This makes sense of Kant's and Leibniz's points. After all, every living being reproduces its standard of normality according to its own relative autonomy, the realm of material and functional variations its specific teleological constitution continuously allows for—and remains inaccessible to us.

Such ethical speculation can seem to go too far astray for the concerns of bioengineering and the people and institutions involved with it. In reality, scientific research and its industrial implementation are supposed to be very down to earth. However, that does not do away with social responsibility. For, standing at the threshold of manipulating the genetic and micro-genetic make-up of existing forms of life, two pressing images suggest themselves (compare [9]):

i. in one image, the amount in which technological capacities of manipulation of life and living substance advance is directly matched by advanced normalization and exploitation of forms of life according to the needs of profit and domination. This is the scenario where the patriarchal and colonist heritage of Western Enlightenment goes off, dispossessed of its own essentialist morals; and
ii. in the other image, the formative power [4] (*"bildende Kraft,"* KdU B293/A289) of living matter is freed and multiplied by the growing technological possibilities and its hardly foreseeable mutations. In this scenario, the amount of rising technological capacities is only matched by that of rising responsibility to care and deal with the complicity of every action and its consequences.

Conclusion

According to Kant, a non-living object is mechanistically constituted: all of its going-ons can be fully explained with recourse to efficient causes. In contrast, living beings are organisms. They are autonomously constituted by final causes, their parts being defined by the various functions they perform for the entire organism. In this way, Kant tries to improve upon Leibniz's proposal and develop a conception of living beings that avoids the trap of reification while still providing a categorical difference between living and non-living beings.

From the background of this Kantian distinction between living and non-living beings, we suggested that bioengineering can either be understood as the quest to engineer life out of non-living matter or as the quest of engineering already existing living beings. Applying the Kantian distinction to cutting-edge AI systems reveals that these systems clearly show organismic features while still lacking essential aspects of living beings. It highlights the peculiar nature of these systems, and it creates a fruitful situation for future research: how should we conceive these beings? Are we in the process of creating a third category of beings over and above non-living objects and living beings?

Also, the growing importance and capacities of bioengineering, in general, invigorate the ethical dimensions of Kant's and Leibniz's attempts to conceptualize life. Their endeavor is historically situated within a mechanistic framework of man's mastery over nature. At the same time, in ultimately respecting the inaccessibility of living matter, the two philosophers have provided us with a hint for the dystopian dangers as well as the utopian hopes for bioengineering and the powers it beholds.

Core Messages

- We have discussed a notion of the living organism based on its formal teleological structure, according to Kant.
- Organisms are contrasted with mechanistic objects, and some consequences are drawn for the notion of bioengineering.
- Bioengineering is close to a *contradictio in adiecto*: either something is alive, or it is engineered.
- We argue that the growing capacities of bioengineering invigorate the ethical dimensions of Kant's and Leibniz's attempts to conceptualize life.
- In respecting the inaccessibility of living matter, the two philosophers offer a map for bioengineering's dystopian dangers and utopian hopes alike.

References

1. Huneman P (2014) Kant vs. Leibniz in the second antinomy: organisms are not infinitely subtle machines. Kant-Studien 105:155–195
2. Kant I 1781/1787 (1998) Kritik der reinen Vernunft, nach der ersten und zweiten Originalausgabe. In Timmermann J (ed) Meiner. Abbreviated by 'KrV', Hamburg
3. Kant I 1781/1787 (1998) Critique of pure reason. In: Guyer P, Wood AW (Transl. and ed) Cambridge University Press. Abbreviated by 'KrV.', Cambridge
4. Kant I 1790 (1974) Kritik der Urteilskraft. In: Weischedel W, Frankfurt AM. Suhrkamp. Abbreviated by 'KdU'
5. Kant I 1790 (2000) Critique of the power of judgment. In: Guyer P, Matthews E (Transl.), Guyer P (ed) Cambridge University Press, Cambridge. Abbreviated by 'KdU'
6. Vaswani A et al (2017) Attention is all you need. arXiv preprint arXiv:1706.03762
7. Bahdanau D, Cho K, Bengio Y (2014) Neural machine translation by jointly learning to align and translate. In: arXiv preprint arXiv:1409.0473
8. Haraway D (1985) Manifesto for cyborgs: science, technology, and socialist feminism in the 1980s. Soc Rev 80:65–108
9. Haraway D (1979) The biological enterprise: sex, mind, and profit from human engineering to sociobiology. In: Radical historical review, pp 206–37. https://doi.org/10.1215/01636545-1979-20-206

Reto Gubelmann received a Ph.D. in Philosophy from the University of Zurich. His thesis was about naturalism in epistemology and the philosophy of mind. Reto now leads a project funded by the SNSF to investigate the linguistic capacities of neural networks.

Marco Toscano is pursuing a Ph.D. in Philosophy from the University of Zurich. His thesis focuses on conceptions of nature as interfaces between philosophical theories of subjectivity and current methodological frameworks in the social sciences, for which he was awarded an SNSF research grant.