



**HAL**  
open science

## On selves and I's – A lecture apropos of Douglas Hofstadter's "I am a strange loop"

Dimitris Ginosatis

► **To cite this version:**

Dimitris Ginosatis. On selves and I's – A lecture apropos of Douglas Hofstadter's "I am a strange loop" . Maitrise. Athens School of Fine Arts' annex in Hydra, Greece. 2015. cel-01247890v5

**HAL Id: cel-01247890**

**<https://hal.science/cel-01247890v5>**

Submitted on 24 Mar 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Dimitris Ginosatis\*

---

On Selves & I's

A lecture apropos of Douglas Hofstadter's  
*I am a strange loop*

---

Lecture (13.000 words) delivered on Tuesday 17<sup>th</sup> March 2015, before an audience of faculty members and digital arts & design undergraduate and postgraduate students of the Athens School of Fine Arts (Athens-Greece) and OCAD University (Toronto-Canada), as part of the "Future of Storytelling and Open Worlds Masterclass", organized by the Digital Futures Program of the OCADU in collaboration with ASFA. The Masterclass was held at ASFA's annex in Hydra, Greece (15<sup>th</sup>-20<sup>th</sup> March 2015).

Keywords: *self, selfhood, consciousness, strange loop, logic, epistemology, cybernetics, neurophenomenology.*



\*Since 2009 to present, Dr. Dimitris Ginosatis has been teaching, as an Adjunct Lecturer and, consequently, as an Adjunct Assistant Professor of Philosophy & Aesthetics, at the "Digital Arts" Postgraduate Studies Program of the Athens School of Fine Arts (Greece). He has written on the subject of philosophy, aesthetics, epistemology and technical media and has delivered lectures in (inter)national conferences, symposia, workshops and masterclasses. His essays and lectures can be found at the digital archive of the *Centre for Direct Scientific Communication*, part of the *French National Centre for Scientific Research* (<https://hal.archives-ouvertes.fr/>) as well as at the website of the *Social Sciences Research Network* ([www.ssrn.com](http://www.ssrn.com)). He has translated into Greek and published works in the fields of 20th century continental philosophy (Jacques Derrida, Michel Foucault, Sarah Kofman, Pierre Klossowski), 20th century contemporary art theory (Nicolas Bourriaud) and 18th-19th-20th century french literature (Marquis de Sade, Guy de Maupassant, Jacques Tournier). All the titles of the works translated by Dimitris Ginosatis can be found at the website of the national Book Centre of Greece ([www.biblionet.gr](http://www.biblionet.gr)).

"The cells inside a brain are not the bearers of its consciousness; the bearers of consciousness are patterns. The pattern of organization is what matters, not the substance. It ain't the meat, it's the motion!"<sup>1</sup>

Douglas Hofstadter



Students and colleagues: it is my great pleasure to be part of this international masterclass, here at the Athens School of Fine Arts' annex in Hydra. I hope that you enjoy it as much as I do.

Our general starting point, as regards the issues we will be discussing in today's lecture, can be formulated as follows: how does a "self" come into being and in what ways does that relate to the concept of "strange loop", as developed in Douglas Hofstadter's *I am a strange loop*?

To put the point in perhaps more properly Hofstadterian terms: how is it that self-reflecting, intelligent beings come out of non-intelligent matter? How can a "self" come out of things (carbon, molecules, atoms, proteins and so on) "as selfless as a stone or a puddle"?<sup>2</sup> How does all that "meaningless" stuff that makes us up in the physical universe, grow into an entity that can refer "meaningfully" to itself, perceive itself, talk about itself, become self-aware?

I say "meaningfully" and I stress that word, for, as far as our human world-system is concerned, it's all about "meaning": in the field of immanence that is human existence, everything –even "meaninglessness"– is meaningful.

Before I proceed to the presentation of certain theoretical hypotheses with regard to the above puzzling questions, allow me first to recount a short entertaining incident, which occurred to me last night. It will serve as the stage for what I am going to say:

I was reviewing my lecture notes, when a random thought suddenly dawned on me, to be more precise, a reminiscence of my childhood readings on mythology. As some of you may know, in ancient Greek mythology "Lernaean Hydra" was the name of a monstrous, serpent-like water creature, which was said to possess many heads. According to the legend, for each head cut off it grew two more: "Cut off one head, two more shall take its place".

In Modern Greek, the name of the aforementioned mythological beast is being used in everyday languaging as a common expression that defines a task, an activity or a problem that is difficult to complete or resolve due to constant inflow of new unpredictable factors that suspend the completion of the resolving process.

This morning, I pondered over the incident and I provided myself with the following simple explanation:

Would I be exaggerating if I claim that the reminiscence of the multi-headed monster myth was triggered –not, as many among you might reasonably assume, by the name of the island we are sitting on right now– but by the task I had undertaken, that is, the presentation of a number of ideas that would elucidate as perspicuously as possible such a complex subject of reflection as the "strange loop" concept?

Would it be regarded as a too far-fetched scenario, if I say that, while trying to come up with a way to simplify matters, I found myself confronted with a Hydra-like, multi-headed monster-subject, which

was growing more heads, becoming even more complex, the more I was struggling to reduce its complexity?

To cut a long story short: being aware that I would be addressing a non-expert audience and that I should keep it light and simple, I found myself walking down one of those loopy paths, first described by the Greek presocratic philosopher Zeno of Elea in his famous dichotomy paradox:<sup>3</sup> the further I moved forward, the closer I got to its starting point.

The problem was lying in the fact that I was attempting to convey something that is difficult to communicate to those non-initiated in the philosophical and logical mysteries of loopiness, circular causality and recursiveness. In that way, every act of communication demanded even more communication that, in turn, complicated the whole process, instead of simplifying it. Which proves, at least in my eyes, that "communication" is a highly paradoxical state of affairs, an almost impossible task: contrary to what is commonly believed, it is not in its nature to solve problems; it can only temporarily minimize their impact, inducing at the same time new unpredictable ones. And yet, as the great system theorist Paul Watzlawick suggests, "we cannot not communicate". Communication is simultaneously unavoidable *and* unattainable, not teleological, but paradoxical: a kind of strange loop.

I do not plan to pursue this reasoning further. I shall simply stress the problem, by quoting another father figure of system theory, Niklas Luhmann: "psychic systems or individual minds can think, but cannot communicate; only communications can communicate"!

One may justly object to this claim by asking: Can "communications" communicate outside a thinking "processing device" or hardware: psychic system, individual brain or a computer's Central Processing Unit (well, the latter is far from being adequately "self-reflective" so that it could be labeled as "thinking" in an anthropogenic sense)? And, inversely: Can an entity think independently of a communicative "medium" (natural or symbolic languages, analogue or digital signal sequences etc.) that would formalize the thinking process?

What, at least in my view, would count as an answer to these absolutely legitimate questions, with respect to Luhmann's aphorism, is the following: human thoughts can never be communicated in their own terms, that is, in terms of the laws that govern the endogenous, biochemical organization of the mechanisms that produce them, but in terms of the laws that are inherent in a communicative medium's operational structure. Human brains –like cells– are contained in "membranes": they share no direct relationship with their environments or other entities; they can only "connect", but they cannot "access". However, their "connections" define their very conditions of possibility to such an extent that the medium ("communications") is in essence the message ("thought"). As stated by Ludwig Wittgenstein, "thinking and language belong together. A child learns a language in such way that it suddenly begins to think in it";<sup>4</sup> "Language is [itself the vehicle of] thought".<sup>5</sup>

Ultimately, I realized that I had been gradually enclosing myself in a mental space where every concept I was elucidating was not really pushing me forward in my presentation, but, on the contrary, it was bringing me back full circle to the starting point of the examined problem. I had been caught in a loop of end-less productivity, a self-perpetuating loop, in which every communicative event produced further never-ending communication. In that loop, I was not the "subject", but only the "channel".

The Hydra monster was gradually turning into an "Ouroboros", the self-eating serpent from Egyptian mythology, a tail-devouring snake, which symbolizes self-reflexivity or cyclicity, in the sense of something constantly re-creating itself. From a variant on the classic Ouroboros, with the snake looped once before eating its own tail, may be derived the current mathematical symbol for infinity [ $\infty$ ].

May I, in passing, also note that an immortal, self-eating, circular being is described by Plato as the first living thing in the universe in his *Timaeus*, the most "bizarre" of platonic dialogues (or rather a long monologue), where among other things is being discussed how the world came into existence:

"The living being had no need of eyes when there was nothing remaining outside him to be seen; nor of ears when there was nothing to be heard; and there was no surrounding atmosphere to be breathed; nor would there have been any use of organs by the help of which he might receive his food or get rid of what he had already digested, since there was nothing which went from him or came into him: for there was nothing beside him. Of design he was created thus, his own waste providing his own food, and all that he did or suffered taking place in and by himself. For the Creator conceived that a being which was self-sufficient would be far more excellent than one which lacked anything; and, as he had no need to take anything or defend himself against any one, the Creator did not think it necessary to bestow upon him hands: nor had he any need of feet, nor of the whole apparatus of walking; but the movement suited to his spherical form was assigned to him, being of all the seven that which is most appropriate to mind and intelligence; and he was made to move in the same manner and on the same spot, within his own limits revolving in a circle. All the other six motions were taken away from him, and he was made not to partake of their deviations. And as this circular movement required no feet, the universe was created without legs and without feet".<sup>6</sup>

Thus, had I witnessed my being caught in the very loopiness-state I was thinking and writing about? Was my thought as a *form* becoming increasingly loopy, identifying itself with its *content*, that is, the examined concept of the strange loop?

What kind of dynamic network of interlacing patterns of neural activity had generated this "knot", tying together my unconscious associations (the random reminiscence of the mythological Hydra monster), my conscious thought (the examined Hydra-like problem of strange loops) and my physical reality (the real island of Hydra)?

That's what happens, I guess, when you get too much involved in such elusive matters as strange loops: you may as well end up becoming one.

However, the inaugural question remains: What made me become conscious of all those troubling complications? And what makes me capable, at this very moment as we sit here, of reflecting upon "reflecting upon them"? How does this multi-level-crossing, linguistic meta-awareness come into being?



Since this lecture is about "selves" and "I's", perhaps you may wish to know a few things about your speaker's "I". So let me introduce my –academic– self:

Institutionally speaking, my area of expertise is philosophy, with special emphasis on certain fundamental, ontogenetic issues concerning the material substrate –the biotechnological ground strata– of being in a/the world. The central set of questions that encapsulates the *esprit* of my intellectual interests is an almost undecidable (and, for that reason, recurring) one of pure metaphysics, except that, contrary to traditional metaphysical mode of inquiry, the emphasis here is not put on the *what* (*things are*), but on the *how* (*things become*). For instance: How does a reality emerge? How is it constructed? How such things as "matter", "energy", "information", "things", "beings", "realities" et cetera are possible at all? How do they come into being? How do they work? How do they acquire meaning? But, also, other less abstract, but hardly decidable questions, such as: How is a brain wired up in a skull and how is the hand –that prehensile, multi-fingered organ we primates use in order to seize, move or touch objects– related to the development of rational thinking? How are brain neurons equipped for two and only two reactions, a positive and a negative one (a neuron either fires completely or it does not fire)? How does formal thinking arise? In what ways do media technologies (prehistoric pictographic images, phonetic alphabets, writing, numerical and musical notation systems, programming languages, synthetic images et cetera) that a species has invented, shape not only the species' experience of a world, but also its conceptualizations of that experience? In short, what interests me is not *what a thing is* (i.e. classic ontological issues of identity and non-identity), but *how a thing becomes* (i.e. ontogenetic processes and modes of differentiation).

If you think that these are totally absurd, or even superfluous questions not worth debating, you are most probably right! There is nothing in life –or so it seems at least– obliging us to venture into this kind

of inquiries. They are insubstantial, transitory, arbitrary, unnecessary constructs of a human all too human intellect. So, why bother? "There were eternities during which it did not exist; and when it has disappeared again, nothing will have happened. For this intellect has no further mission that might extend beyond the bounds of human life. Rather, the intellect is human, and only its own possessor and progenitor regards it with such pathos, as if it housed the axis around which the entire world revolved".<sup>7</sup> But, on the other hand, you are also wrong! Firstly, because –despite Socrates' opposite opinion– an unexamined life may be worth living after all, but it can be a crashing bore. Secondly, because the most compelling questions in life –those that invite us to delve more deeply into the fabric of things– are usually the most absurd, undecidable ones. Thirdly, because, as postulated by Heinz von Foerster, only those questions that are in principle undecidable we can decide upon: decidable questions are not, in essence, real questions. Why? "Simply because", Heinz von Foerster continues, "the decidable questions are already decided by the choice of the framework in which they are asked, and by the choice of the rules used to connect what we label 'the question' with what we take for an 'answer'. In some cases it may go fast, in others it may take a long, long time. But ultimately we arrive [...] at an irrefutable answer, a definite 'yes' or a definite 'no'. But we are under no compulsion, not even under that of logic, when we decide on in principle undecidable questions. There is no external necessity that forces us to answer such questions one way or another".<sup>8</sup> Fourthly, because the *kind* of questions one asks, but also the *way* one asks, determine to a large degree one's relating/opening to the world and *vice versa*. In any case, venturing into such inquiries is a matter of *choice*, i.e. a result of manifest and latent processes that are not so easy to decipher.

To conclude, my overall academic activity up to now has been inscribed in the broad, discursive edifice of the good old European (also known as continental) philosophical tradition. I've surely written hundreds of academic pages on various philosophical issues and I've spent hundreds of worth spending hours, since 2009, teaching philosophy and aesthetics, in conjunction with communication and media theory, to undergraduate and postgraduate students. However, engaging with philosophy may not always prove to be as thought provoking as one might expect. The reason is that some quite sizeable chunk of philosophical discourse restricts thought to an attitude of "detached" contemplation: in its distantiated way of thinking about the reality of "things", philosophy tends to overlook any "hard" data with regard to their materiality, not to mention the "physics" of its own act of contemplation.

That kind of philosophical discourse denigrates "hard" sciences perspectives and empirical inquiries; its thought system is immune to transitivity and transdisciplinarity; it suffers from obsessive-compulsive purism and severe lack of experimentation; its scope rarely reaches beyond philosophy to incorporate non-philosophical references and modes of thinking – whether they touch upon matters of biology, thermodynamics, geology and cosmology or of paleontology, language sciences, musicology and poetry.

Philosophy can only provide "concepts" – which, nevertheless, is definitely *not* a "problem" *per se*. On the contrary, "the art of forming, inventing and manufacturing concepts"<sup>9</sup> constitutes the fulcrum of philosophizing. However, philosophy becomes a "problem", when it refuses to problematize its own activity or, alternatively put, when it projects its conceptual products not as regulative, metaphorical constructs that have been invented, but as reason's vessels for truth *a priori*.

On the other hand, what has been said about philosophy also applies to particular sciences: in most cases, engaging in a truly thought-provoking dialogue with scientists proves to be just wishful thinking. They, too, are inclined to confine themselves to their narrow field of overspecialization, thus repressing any element of theoretical meta-thinking, cross-disciplinarity and transitivity. They do science in the same way bees do bee dances or the way computers talk to each other, that is, mechanically, without reflecting upon the very historical, cultural, onto-epistemological conditions of possibility of their own mental activities. Their scientific practices, while impressive in their application of novel methods and analytical tools, suffer from too little philosophical inquiry; they provide "data" – which, nevertheless, is

also *not* a "problem" *per se*. On the contrary, science's primary function is to name the "real" and deal with it.<sup>10</sup> However, science becomes a "problem", when it fails to problematize its own activity or, alternatively put, when it projects its rigid formulae and preordained categories not as regulative, descriptive constructs that have been invented, but as irrefutable states of affairs that have been discovered.

In modern times, this discursive divide between "hard" scientific positivism, dealing with un-historical, raw matter (brains/bodies) and "soft" philosophical theoreticism, dealing with historically and culturally in-formed entities (minds/souls) can be traced back to the work of major philosophical figures of the western tradition, as well as to the practices of most empiricist physiologists, anatomists and natural scientists from the 17<sup>th</sup> century and on. Let me remind you, for example, how persistently Descartes refused to conflate the "mind/soul/spirit" with the gray-white gelatinous mass known as the brain and, consequently, how he managed to isolate it from the rest of the "body machine", as he called it, by locating the former in a small pineal gland in the middle of the brain matter (see *The Passions of the Soul*). Or, how artfully Kant managed to skip the "brain obstacle" in his account of the transcendental consciousness of the self (see the *Critique of pure reason*). Or Hegel who, in his *Phenomenology of mind* (or *spirit*), attacks natural sciences' methods, by arguing that if you seek to understand human thought, "don't place the brain on a dissecting table or feel the bumps on the head for phrenological information. If you want to know what the mind is, examine what it does".<sup>11</sup>

Thenceforth, natural science and philosophy have been relentlessly negating each another – the former dissecting "human meat" (brains/muscles/nerves) and the latter studying its cultural and historical formatting: "The two discourses went separate ways: philosophy of the mind and physiology of the brain remained, for the most part, as blind to the activities of one another as the two hemispheres of a 'split-brain' patient are oblivious to the operations of each other – arguably to the detriment of both".<sup>12</sup>

A virtual debate between a scientist and a philosopher brings to mind those farcical dialogues one finds in Molière's plays or in Woody Allen's scripts:

Scientist contradicting a philosopher on the subject of what "thought" is:

- You, my friend, are the least competent to decide upon the nature of any thing, living or inanimate. How do you really expect to learn a damn thing about what "thought" is, if you have no firm understanding of the way nerve cells are wired up in the central and peripheral systems? It's all about the "meat", my friend: cells, tissues and biochemical secretions. Everything else is but clueless, unprincipled assumption. Wake up from your metaphysical slumber, cleanse your mind of your philosophical delusions, refine your science and have your lancets sharpened!

Who could actually dispute our scientist's argument? Its validity is unquestionable: how can one know anything, without immersing into the "hardware" of the "things" examined? It stands to reason that rendering something knowable requires that knowledge about it be based, captured or inscribed through minute, detailed examination techniques to render it knowable.

Let's hear now our philosopher's reply with regard to the matter:

- I see your point. You have made yourself believe that, by cutting into the "meat", you actually cut into the "thing-in-itself". You have learned to think –what an illusion!– that by studying dead brain tissues and cells, you actually get a grasp of "what" a brain "is". So, tell me –you poor ignoramus– where in this amorphous mass of meat, which, by the way, does not differ a bit from that of a sea slug, do you spot the seat of "thought", "consciousness", "emotion", "feeling" and all that elusive stuff that makes a human being what it is? For I see nothing in there except a pile of organic matter. What is more, you presuppose an external, objective reality that you deem independent of your actions! But, know this: not only your so precious "meat", as object of observation, but also you, my friend, as knowing observing subject, are already always formatted by historical, cultural and other discursive agencies. Study these and leave the cutting and slicing to the butcher.

Who could really dismiss our philosopher's stance on the issue? Indeed, the brain and its nervous system and sub-systems are definitely not "contained within the body's limits. The circuit from sense perception to motor response begins and ends in the world. The brain is thus not an isolable anatomical body, but part of a system that passes through a person and her or his (culturally specific, historically transient) environment. As the source of stimuli and the arena for motor response, the external world must be included to complete the sensory circuit (sensory deprivation causes the system's internal components to degenerate)".<sup>13</sup>

Such has been the situation for centuries. There have been, of course, exceptions to the rule: such are the cases of philosophers like Fr. Nietzsche (primarily a classical philologist with a particular focus on ancient greek and roman rhetoric, who was expressly more interested in physiology and natural history than in modern philosophical hermeneutics) or Ed. von Hartmann (metaphysician and psychologist, author of the emblematic *Philosophy of the unconscious*, published in 1869), but also scientists such as R. L. C. Virchow, A. V. Espinas, C. W. von Nägeli and others who played a key role in the development of scientific experimentation in the 19<sup>th</sup> century. More recently, during the 20<sup>th</sup> century, a considerable number of scientifically informed philosophers and philosophically informed scientists, from the most diverse areas of expertise, have been contributing explicitly or implicitly, each one affected in his own distinctive way by widely varying issues, to a certain "re-dif-fusion" of the discourses: R. Magnus,<sup>14</sup> D. de Barenne, W. McCulloch, G. Günther, H. von Foerster, J. Lettvin, G. Bateson, P. Watzlawick, K. L. von Bertalanffy, E. von Glaserfeld, H. Atlan, A. A. Moles, E. Morin, M. Serres, G. Deleuze, F. Guattari, H. Maturana, Fr. Varela, N. Luhmann, T. Winograd, H. Dreyfus, P. Smith Churchland, J. Z. Young, R. M. Young, J. Haugeland, D. Dennett and D. Hofstadter, to name a few.

However, the problem remains: an overwhelming majority of philosophers and scientists are reluctant to come to grips with "alien" *modi cogitandi*. Broadly speaking, trotting along multiple (disciplinary) paths is not a viable way to be in the world. Which brings to my mind a statement by a man of many disciplines, Heinz von Foerster, a pioneering mathematician, biophysicist, cybernetician and philosopher:

"I don't know where my expertise is; my expertise is no disciplines. I would recommend to drop disciplinarity wherever one can. Disciplines are an outgrowth of academia. In academia you appoint somebody and then in order to give him a name he must be a historian, a physicist, a chemist, a biologist, a biophysicist; he has to have a name. Here is a human being: Joe Smith—he suddenly has a label around the neck: biophysicist. Now he has to live up to that label and push away everything that is not biophysics; otherwise people will doubt that he is a biophysicist. If he's talking to somebody about astronomy, they will say 'I don't know, you are not talking about your area of competence, you're talking about astronomy, and there is the department of astronomy, those are the people over there' and things of that sort. Disciplines are an aftereffect of the institutional situation".<sup>15</sup>

To expel any misunderstandings, it should be remarked that "a-disciplinarity" does not mean "bypassing" or "abolishing" the particular disciplines: to drop a discipline, one needs to have mastered it first. In that sense, "a-disciplinarity" amounts to treating disciplines as complementary, to cutting *across* them. It is, therefore, equivalent to a kind of constant inquiry, a singular stance that is forever "self-enriching its relationship with the world".<sup>16</sup> True "thinking" starts where common sense and standardized conceptions of knowledge end.

However, this matter of cross-disciplinarity and transitivity has another crucial aspect which is of paramount significance: it's neither about aspiring to a holistic *weltanschauung* (world-view), nor an issue of mutual enrichment that would strengthen both philosophical and scientific discourses, incorporating them to an absolute, synthetic unity. The problem seems to be a bit more complex: it's inextricably intertwined with the production mechanisms and the very existential status of *living* in our era of unprecedented technologization and compartmentalization of knowledge, disguised under a false cloak of "productive creativity"; with the uniformization that the latter have imposed on modes of

thinking, feeling, behaving and being; with the de-singularizing ways by which *living* is being calibrated on the basis of normative practices underpinning *life* in its entirety and, inevitably, with the overriding need for the invention of further mutant life-forms: forms that would re-singularize existence and foster new fields of reference; agencies that would help to invent new relations "to the body, to phantasy, to time passing, to the 'mysteries' of life and death",<sup>17</sup> to *living* itself. And here is where the significance of transitivity and cross-disciplinarity emerges: by grafting the *philosophical* into the *scientific*, by injecting human sciences perspectives to the "hard" sciences and vice-versa, one can create a "polyphonic interlacing", that undermines the hegemony of the scientific and the philosophical "superego", aspiring to a virtual reshaping of both on the model of a transmuted ecology of mind as well as a reassessed ecosophy of space and time, according to a third paradigm, an *ethico-aesthetic* one.

However, this is a huge, extremely complicated and delicate issue that calls for separate treatment in a future lecture. For the moment, let's stick to our subject matter.



The general, conceptual framework of the topics we are discussing today has been already predetermined by the contents of an exceptional book, titled *I am a strange loop*, published in 2007, by a *sui generis* thinker, writer and scientist, Douglas Hofstadter, eminent professor of cognitive science at Indiana University and author of another emblematic book titled *Gödel Escher Bach—An eternal golden braid*, an 800-page masterpiece, published in 1979, that I first read twenty years ago, as a third year undergraduate (majoring in Communication, Media & Culture) with an intense interest in circular causality, self-reference, paradoxes and vicious circles, combined with an enthusiasm for J. S. Bach's art of the fugue. With respect to *GEB*, *I am a strange loop* could be considered a further development of the central idea amply elaborated in the former. I thus suggest that if anyone seeks to experience to the fullest Hofstadter's conceptual universe, he'd better immerse into *GEB*'s vortex-like architecture.

Now, I should warn you that my intention is in no way to restrain myself to the themes developed in *I am a strange loop*. I am sure you've all read it—as you were supposed to do— and "wrestled" with its contents. You can also find tons of related articles and scientific papers worth reading out there.

What I shall do—in the simplest way possible—is to provide you with an insight into Hofstadter's core idea, by reconstructing a general introduction to a philosophy of the so-called "consciousness", with a little help from certain key figures in the development of western rationality: the most self-disciplined philosopher of modern times, Immanuel Kant; Friedrich Nietzsche, who was indeed so wise and wrote such good books; Gotthard Günther, who I deem to be the most interesting 20<sup>th</sup> century logician and metaphysician to ever walk this earth; Heinz von Foerster, an extraordinary man of many disciplines and father of 2<sup>nd</sup> order cybernetics and, finally, two leading figures in the domain of neurophenomenology, Humberto Maturana and Francisco Varela.

The subject under examination falls into the category of what John Haugeland has aptly called "Mind design":

"Mind design is the interdisciplinary endeavor to understand mind (thinking, intellect) in terms of its design (how it is built, how it works). It amounts, therefore, to a kind of cognitive psychology. But it is oriented more toward structure and mechanism than toward correlation or law, more toward the "how" than the "what", than is traditional empirical psychology. [...] Mind design has always been an area of philosophical interest, an area in which the conceptual foundations—the very questions to ask and what would count as an answer— have remained unusually fluid and controversial".<sup>18</sup>

This is exactly the direction we are going to follow today: the fluid and controversial one. What I am going to do is to focus on the philosophical aspects of my subject rather than the mathematical or

technical ones. There is certainly a lot of mathematics talked about in those two books. Yet, I will leave it aside focusing mostly on the central philosophical problem and its multifaceted implications.

Allow me to begin in the most stereotypical manner: The quest for an explanation of the inner workings of the mind, the constitution of consciousness and the self as a unitary agent of knowledge, of willing, of thinking, of feeling and so on, is a very old project –undertaken *en masse* by philosophers and scientists in ancient and modern times– which is impossible to outline in a single lecture. Yet, we have to determine a starting point. And that point will be Immanuel Kant's theory of transcendental logic.

Intimidating as the term "transcendental logic" may sound, I urge you to not get discouraged. Kantian theory of transcendental logic is the academic code name of a considerable portion of Immanuel Kant's philosophical project. Kant's main concern was to illuminate what "knowing" is, how it is acquired, processed and elaborated by the human cognitive apparatus, what is false knowledge and what is true knowledge, finally how can we be sure that our perceptions, our intuitions and our convictions about our external as well as our internal reality are real or a mere illusion.

Kant was the first modern western philosopher to formulate a truly radical and innovative theory of the mind thoroughly demonstrated in his *Critique of pure reason*,<sup>19</sup> published in 1781. It's a theory of reflexivity and recursiveness, according to which the mind, the human cognitive apparatus, is presented as a set of complex recursive functions through which the act of conscious knowing and the "I" emerge.

Let me clarify this vague description of kantian transcendental logic via some enlightening excerpts and examples concocted by the German logician Gotthard Günther.

Günther left Hitler's Germany in 1937 and after brief stays in Italy and South Africa came to the United States in 1940. From 1961 till 1971 he was given the unique chance –through the agency of Warren McCulloch, eminent experimental epistemologist– to conduct research on early artificial intelligence and bio-inspired computing research programs at the then renowned Biological Computer Laboratory (BCL), an independent division within the Department of Electrical Engineering at the University of Illinois, a top research institute founded in 1957/1958 by Heinz von Foerster. Driven by a certain metaphysical interpretation of the foundations of logic, Günther would pursue his research on a non-Aristotelian many-valued (poly-contextural) logic.

My admiration for Günther's way of thinking things that seem almost unthinkable demands that I dedicate to him a few more words.

As mentioned in Charles Parsons' incisive introductory note<sup>20</sup> to Günther's correspondence with Kurt Gödel:

"[Günther's] original philosophical background was Hegelian and he continued to see philosophy from that point of view, though he was also influenced by Leibniz and by twentieth-century German figures. [...] A project that he pursued for many years [...] was how formal logic ought to be revised to accommodate what he took to be insights about the nature of thought and its relation to reality from the German idealist tradition. He also became interested in and wrote about cybernetics. Norbert Wiener, who publicized the term, defined cybernetics as the science of 'control and communication, in the animal and the machine'. Its concerns derived from engineering and theoretical biology, but what seems to have most interested Günther was the idea of artificial intelligence. He was one of the earlier thinkers to write from a philosophical point of view on that subject. He was thus a very unusual intellectual figure for his time, a Hegelian philosopher with an interest in modern logic and involvement in what later came to be called 'computer science'".



But, let's proceed to the excerpts. What follows is from Günther's paper titled "Can mechanical brains have consciousness?",<sup>21</sup> published in 1953:

"Till the publication of *The Critique of Pure Reason*, philosophers and scientists had entertained the following ideas about the origin of consciousness: they said, our mind is like a jug into which you pour water. The water while it

is poured is in a rather chaotic state. The jug, however, stills it, and forces the fluid to adopt its own hollow form. According to this theory, then, our consciousness is a system of hollow forms into which are poured all the sensations, impressions and stimulations which our nerve system transmits from the outer world. But these transmissions arrive in a rather chaotic state. They become conscious only by being submitted to a forming and ordering mechanism, which gives them their final (i.e., conscious) shape. [...] They say: our mind has two fundamental components, namely contents and forms, and if the two come together the result is consciousness. If we talk about the universal reservoir of possible contents of our consciousness, we say: 'material world'; if we talk about the jug these contents are poured into, we say: 'formal logic'. The first description of our forms of consciousness and how they work, in order to shape the incoming material was originally given by Aristotle.<sup>22</sup> Since then, 'formal logic' and 'Aristotelian logic' have been historically equivalent terms. However, the 'jug' Aristotle described was comparatively small. The Stoics, later, enlarged it a bit and since the introduction of Boolean algebra it has been discovered that all our previous conceptions about the size of our 'jug' have been ridiculously conservative. The 'jug' is still growing. Now it is usually called: mathematical logic; but it is still of course the same venerable vessel of ancient origin: a formal logic – meaning the theory of a mechanism that forms and orders contents. The only trouble is that if you pour water in a jug, this vessel does not become water-conscious; and if you charge a battery the battery does not become electricity-conscious. This did not disturb the philosophers of the Platonic and Aristotelian tradition. They said: it is different with man. Man has a soul. The inanimate object has not; and you need in addition to that synthesis of forms and contents, a Self that watches that synthesis, thus finally producing that miraculous phenomenon we call 'consciousness'. To the scientist, of course, the introduction of the term 'soul' is nothing but a very polite way of saying: there is something in addition to this form-and-content business, but we don't know what it is. It was the German philosopher Immanuel Kant who in his Critique of Pure Reason eliminated the concept of 'soul' from the theory of logic (earning him an indictment of "atheism") and who stated that beyond the mechanism of formal logic there is in our brain a second mechanism which works on entirely different principles".

The kantian design for consciousness is as follows – pay attention to the way Mr. Günther binds together the concepts of "reflexivity" and "consciousness":

"Consider your own consciousness, a sensitive 'screen'. This 'screen' receives, through your 'I' sensorial system messages from the outer world. Neuronic impulses coming from your eyes, your ears, your skin, your muscles, etc. impress themselves upon that 'screen' and are reflected. But this reflection is not thrown back at the world-system from which it came... Instead, it is thrown into a deeper recess of your brain, turns around and appears a second time on your brain-'screen', superimposing a second reflection on the first. This second appearance establishes the miraculous phenomenon, which we call 'consciousness'. Let's illustrate this process with a simple example: you are aware of a flower. This object of the outer world sends messages through your senses to your brain-'screen', where a picture of the object is formed. The picture bounces off the 'screen' as unconscious message: 'a rose'. Then it goes to some other part of your brain, and returns to the first place with the superimposed content 'acknowledged'. Now the image on your brain-'screen' has a functional depth-dimension, which is expressed in the statement: 'I see a rose'. The original message 'a rose' does not establish consciousness, because it is a simple reflection, not unlike the one in the mirror; but the returning message does, for it is a reflection-in-itself – or as we moderns should rather be inclined to say, it is a reflection upon itself. Now, it is obvious that we should be able to design consciousness technically if we could find out what happens to the message after it has been first received on the screen of our brain and before the later moment, when it returns to it with the stamp 'acknowledged' and produces consciousness by its second impact on the screen (Incidentally, the time-lag between the two moments is so small that it is unobservable by the normal method of introspection). Fortunately we know what happens to the message during this reflexive intervall and it is this theory of the brain processes during the round-trip of our message that is called 'transcendental logic'".

According to Günther, the second mechanism, which works on entirely different principles beyond the first mechanism of formal logic:

"... does not form messages any more but carries them through processing stages and finally returns them to the original 'screen', the identity level of the formal logic. Insofar as this carrying capacity, which transports the messages first beyond the screen, is the most outstanding feature of this second brain-mechanism, Kant called the theory of it

'transcendental' logic. This theory is capable of demonstrating that if the message 'a rose' is carried beyond the original 'screen' and processed in a well defined manner, then the concepts 'I' and 'perception' are added. These additions, however, do not by themselves produce consciousness. They are pre-consciously attached. Only when the thus modified message returns to the screen is consciousness actually produced. This happens in the following way: The returning message does not return to all parts, of the screen, but only to two sections of it, called "memory" and "identification" (the classical axiom of identity). The memory still retains the original pattern (unconscious): 'a rose'; on which is superimposed (unconscious): 'I see a rose'. Identification now produces a confrontation by attempting to establish a one-to-one correspondence relation between the original pattern and the enriched second message. This does not work! It turns out to be impossible to establish, by confrontation, a one-to-one correspondence between 'a rose' and 'I see a rose'. The first part of the second sentence: 'I see...' overlaps. In other words: the reflection-in-itself produces something that cannot be identified with the mere content 'a rose'. A tension of meaning is created – a tension between identity and non-identity. And this is the moment when consciousness and conscious thought come into existence. No mysterious soul is necessary to explain the workings of consciousness. It should, however, be stressed that transcendental logic demonstrates only that consciousness is a mechanical process. Consciousness is that state in which a person is aware of the objective world. In other words: consciousness is equivalent to being aware of objects located outside the system of awareness. It is quite a different story whether self-consciousness is also mechanical [...]. Self-consciousness is not awareness of objects, but awareness of awareness of objects".

In Summa, Günther's interpretation of the kantian theory of the mechanics of consciousness is as follows:

- The cognitive apparatus of a human living system consists of an "interface" (an imaginary projector "screen" inside the brain) and a series of mechanisms.
- The system receives a stimulus from the outside world and replicates it as "object-image" onto that intermediary surface.
- This "object-image" is then filtered through a 1<sup>st</sup> mechanism and is transferred to the logical processing center.
- At that point, the "object-image" is "acknowledged" and is filtered through a 2<sup>nd</sup> mechanism that translates the "object-image": instead of being "an object-image", the entry becomes "I perceive an object-image".
- The percept "I perceive an object-image" is transferred back to the original surface and superimposes a copy of itself: "I perceive an object-image" onto the preexisting imprint of "an object-image".
- Consciousness is then produced when the system notices the difference between the "object-image" and "I perceive an object-image". In effect, they are equivalent, because the latter is just a logical reflection of the former back onto itself, but not identical.
- Consciousness is the state in which a living system is aware of objects located outside its system of awareness.
- Self-consciousness is not awareness of objects, but awareness of the awareness of objects located outside the system of awareness.
- Once the system is aware of its ability to be aware of objects located outside its plane of being, it develops a sense of self.



Not bad at all for an 18<sup>th</sup> century philosopher whose surrounding technological reality did not consist of super-advanced computational systems and artificial neural networks, but of bedchamber mirrors, i.e. simple reflective surfaces!

The story, however, does not end here. And the reason is that, in his *Critique of pure reason*, Kant enriched his innovative contribution to a consistent philosophy of cognition with a constructivist theory

that is considered the cornerstone of his philosophical edifice and the birthplace of contemporary cognitive sciences.

He said: our human cognitive apparatus is biologically structured in such a way that it cannot access what he calls the "things-in-themselves", that is, the external objective reality. Instead, it can only have access to "appearances", that is, to the specific way things appear to our cognitive apparatus or, to be more precise, to the way our cognitive apparatus constructs and shapes the incoming data providing it with a form, thus permitting it to appear to the intellect as a representation of an object. Cognition reaches appearances only, leaving the thing in itself as something actual for itself but uncognized by us.

In short, he claimed that our representation of things as they are given to us does not conform to these things as they are in themselves, but rather that these objects, as appearances, conform to our specifically human manner of mental representation.

Therefore, Kant describes the human cognitive apparatus in terms of a somehow closed system that processes indeterminate incoming stimuli or impressions by providing them with a form, which permits them to appear to the intellect as objects. Yet, these external stimuli or impressions, what Kant calls "things-in-themselves", cannot be cognized, due to the specific biological organization of human cognition.

Kant's constructivist foundation for scientific knowledge restricts science to the realm of appearances and implies that *a priori* knowledge of things-in-themselves that transcend possible human experience is impossible.

Those among you who are intuitive enough must have noticed that both the phenomema of consciousness (awareness of objects located outside the system of awareness) and self-consciousness (awareness of awareness of objects located outside the system of awareness), as described by Kant, share a common characteristic. They both behave as a reflexive loop, in the sense that they exhibit a sort of circularity of processes. In the context of our discussion, the "loop" concept is to be understood according to its topological meaning: a path or a process that starts and ends at the same point. A famous paradigm of a loop is the "Möbius strip", a continuous, one-sided surface formed by twisting one end of a rectangular strip through 180° about the longitudinal axis of the strip and attaching this end to the other. Another example is the "Klein bottle": a closed non-orientable surface that has no inside or outside, originally described in 1882 by the German mathematician Felix Klein.

However, the most decisive twist to the historical course of the "loop" conceptualizations was given by Warren McCulloch who, before and during World War II, was struggling with his ambitious project to model the brain's neural functions with a formal logical system, a sort of Leibnizian logical calculus. A colossal project that was finally completed with the help of Walter Pitts and was published in the *Bulletin of Mathematical Biophysics* in 1943, under the famous title: "A logical calculus of the ideas immanent in nervous activity". Having digested Russell's and Whitehead's endeavor to show, in their *Principia Mathematica*, that all of mathematics could be built up by means of basic logic –that is, either true or false propositions and logical operations ("and", "or", "not"), which link the former into increasingly complicated networks–, but also inspired by Alan Turing's foundational 1936 paper "On Computable Numbers, with an Application to the *Entscheidungsproblem*":

"which proved the possibility of a machine that could compute any function (so long as it was possible to do so in a finite number of steps), McCulloch became convinced that the brain was just such a machine – one which uses logic encoded in neural networks to *compute*. Neurons, he thought, could be linked together by the rules of logic to build more complex chains of thought, in the same way that the *Principia* linked chains of propositions to build complex mathematics.

[McCulloch] knew that each of the brain's nerve cells only fires after a minimum threshold has been reached: Enough of its neighboring nerve cells must send signals across the neuron's synapses before it will fire off its own electrical spike. It occurred to McCulloch that this set-up was binary – either the neuron fires or it doesn't. A neuron's

signal, he realized, is a proposition, and neurons seemed to work like logic gates, taking in multiple inputs and producing a single output. By varying a neuron's firing threshold, it could be made to perform 'and', 'or' and 'not' functions. [...] Before Pitts' arrival, McCulloch had hit a wall: There was nothing stopping chains of neurons from twisting themselves into loops, so that the output of the last neuron in a chain became the input of the first – a neural network chasing its tail. McCulloch had no idea how to model that mathematically. From the point of view of logic, a loop smells a lot like paradox: the consequent becomes the antecedent, the effect becomes the cause. McCulloch had been labeling each link in the chain with a time stamp, so that if the first neuron fired at time  $t$ , the next one fired at  $t+1$ , and so on. But when the chains circled back,  $t+1$  suddenly came *before*  $t$ . Pitts knew how to tackle the problem. He used modulo mathematics, which deals with numbers that circle back around on themselves like the hours of a clock. He showed McCulloch that the paradox of time  $t+1$  coming before time  $t$  wasn't a paradox at all, because in his calculations 'before' and 'after' lost their meaning. Time was removed from the equation altogether. If one were to see a lightning bolt flash on the sky, the eyes would send a signal to the brain, shuffling it through a chain of neurons. Starting with any given neuron in the chain, you could retrace the signal's steps and figure out just how long ago lightning struck. Unless, that is, the chain is a loop. In that case, the information encoding the lightning bolt just spins in circles, endlessly. It bears no connection to the time at which the lightning actually occurred. It becomes, as McCulloch put it, 'an idea wrenched out of time'. In other words, a memory".<sup>23</sup>

I remind you that earlier we defined consciousness as that state in which a person is aware of the objective world. In other words: consciousness is equivalent to being aware of objects located outside the system of awareness. But we also noted that the biological constitution of human cognition does not permit the latter to cognize the object or the thing-in-itself, but only an appearance of it – an appearance, which is produced and shaped by cognition itself.

The word "appearance" ("Erscheinung") is not to be confused with "deception" or "illusion" ("Schein"). Our mind does not deceive us. It only obeys the laws and restrictions of its own internal organization: "Kant refines the traditional philosophical account of appearance by distinguishing between appearance, phenomenon and illusion. He insists on these distinctions in order to redeem appearance from the obloquy it suffered at the hands of the philosophical tradition: [appearance] is not simply illusion –the deceptive semblance of sensible perception– but rather experience within the limits of human intuitions of space and time".<sup>24</sup> Kant suggested that "we cannot have legitimate knowledge outside these forms of intuition and consequently that we can only properly know appearances in space and time. Appearances then are not potentially deceptive sensible impressions, but possess their own order and organization".<sup>25</sup> They are as real as our mind. On the other hand, "Appearance" ("Erscheinung") becomes a "deceptive semblance" or "illusion" ("Schein"), whenever our understanding takes appearances as if they were "things-in-themselves" ("Ding an sich").

Kant actually showed us, almost two and a half centuries ago, that what we nowadays call "information" is something totally different from what we usually mean by that word. The usual notion of information is that on your computer screen or on your wristwatch, for example, there is information. In other words, we think information as something external to our cognition that is to be decoded by the latter.

Yet, on your screen there are only linearly ordered meaningless symbols. And on your watch there are only mechanical hands and numbers. If there is something there, this is definitely not information. Only when "you" are observing the screen or the watch do you generate the information, by interacting with your own sensory experience. So, as Heinz von Foerster would say, information is generated in the one who looks at things. In that sense, you can also understand why there is not such thing as "information processing", as if information were a commodity we could pass on. I would dare to say that information is a misleading notion with respect to what really happens in the workings of human cognition. Information is not external to the cognitive system, but immanent to it. Sensory forms are produced only in the cognitive system of the perceiver. In a nutshell, consciousness is something that emerges from a self-generating cognitive system that behaves as a reflexive loop. Consciousness emerges when our cognitive system interacts with its own sensory experience.

In the case of self-consciousness loopiness is even more complex and abstract: if consciousness is a set of recursive functions and processes characterized by circularity and loopiness, if consciousness is self-reflection, then self-consciousness is self-reflection reflected upon itself, a loop looped upon itself – an inconceivably outrageous topology.

No wonder then that Kant's philosophy of the mind has been defined as a sort of "Copernican revolution" in the domain of human sciences. Actually, if you attempt to trace back the ancestral roots of contemporary cognitive science and artificial intelligence research, I can assure you that in the end you will find standing there the figure of Kant accompanied by those of Aristotle and Hegel, whose *Phenomenology of mind/spirit*, published in 1807, might be as well considered the first "handbook" to artificial intelligence and robotics.



At this point, it would be interesting to see this matter through another "non-Kantian" (yet implicitly relevant to Kant) radical perspective, in that case a nietzschian one. Listen to the following –much debated– insight offered by Friedrich Nietzsche in one of the most staggering pieces of philosophical prose, where consciousness is presented as a superficial –and even superfluous– phenomenon (produced under the pressure of the human need to communicate), inextricably linked with cultural technologies (speech, writing), which translate the singular (the unconscious) in the gregarious sign-system of the species (consciousness). In that piece, Nietzsche's argument echoes Gottfried Wilhelm Leibniz's notion of "Unconscious Mind", expressed in the form of infinitesimal perceptions ("petites perceptions"), a sort of calculus of knowledge, presented in his preface to the *New essays on human understanding* (1765).<sup>26</sup> Long before Freud established the term in the context of his psychoanalytic discourse, Leibniz's idea can be found "growing through Immanuel Kant's 'a priori forms of sensibility' (space and time), Fichte, Schelling, Lessing and Schopenhauer, to become in Eduard von Hartmann's work, not merely everything performed by animals for their surviving, but the very *Geist* of evolution itself".<sup>27</sup>

What follows is from the aphorism §354 of the *Gay Science*,<sup>28</sup> published in 1882. It is titled: "On 'the genius of the species'":

"The problem of consciousness (or rather, of becoming conscious of something) first confronts us when we begin to realize how much we can do without it; and now we are brought to this initial realization by physiology and natural history [...]. For we could feel, think, will, remember, and also 'act' in every sense of the term, and yet none of all this would have to 'enter our consciousness' (as one says figuratively). All of life would be possible without, as it were, seeing itself in the mirror; and still today, the predominant part of our lives actually unfolds without this mirroring – of course also our thinking, feeling and willing lives, insulting as it may sound to an older philosopher. To what end does consciousness exist at all, when it is basically superfluous? If one is willing to hear my answer and its possibly extravagant conjecture, it seems to me that the subtlety and strength of consciousness is always related to a person's (or animal's) ability to communicate; and the ability to communicate, in turn, to the need to communicate. [...] Assuming this observation is correct, I may go on to conjecture that consciousness in general has developed only under the pressure of the need to communicate; that at the outset, consciousness was necessary, was useful, only between persons (particularly between those who commanded and those who obeyed); and that it has developed only in proportion to that usefulness. Consciousness is really just a net connecting one person with another – only in this capacity did it have to develop; the solitary and predatory person would not have needed it. That our actions, thoughts, feelings and movements –at least some of them– even enter into consciousness is the result of a terrible 'must' which has ruled over man for a long time: as the most endangered animal, he needed help and protection, he needed his equals; he had to express his neediness and be able to make himself understood – and to do so, he first needed 'consciousness', i.e. even to 'know' what distressed him, to 'know' how he felt, to 'know' what he thought. For, once again: man, like every living creature, is constantly thinking, but does not know it; the thinking which becomes conscious is only the smallest part of it, let's say the shallowest, the worst part – for only that conscious thinking takes place in words, that is, in communication symbols; and this fact discloses the origin of

consciousness. In short, the development of language and the development of consciousness (not of reason, but strictly of the way in which we become conscious of reason) go hand in hand. One might add that not only language serves as a bridge between persons, but also look, touch and gesture; without our being conscious of our sense impressions, our power to fix them and as it were place them outside of ourselves, has increased in proportion to the need to convey them to others by means of signs. The sign-inventing person is also the one who becomes ever more acutely conscious of himself; for only as a social animal did man learn to become conscious of himself – he is still doing it and he is doing it more and more. My idea is clearly that consciousness belongs not to man's existence as an individual, but rather to the community and herd-aspects of his nature; that accordingly, it is finely developed only in relation to its usefulness to community or herd; and that consequently each of us, even with the best will in the world to understand ourselves as individually as possible, 'to know ourselves', will always bring to consciousness precisely that in ourselves which is 'non-individual', that which is 'average'; that due to the nature of consciousness – to the 'genius of the species' governing it– our thoughts themselves are continually as it were outvoted and translated back into the herd perspective. At bottom, all our actions are incomparably and utterly personal, unique and boundlessly individual, there is no doubt; but as soon as we translate them into consciousness, they no longer seem to be... This is what I consider to be true phenomenism and perspectivism: that due to the nature of animal consciousness, the world of which we can become conscious is merely a surface and sign-world, a world turned into generalities and thereby debased to its lowest common denominator – that everything which enters consciousness thereby becomes shallow, thin, relatively stupid, general, a sign, a herd-mark; that all becoming conscious involves a vast and thorough corruption, falsification, superficialization and generalization. In the end, the growing consciousness is a danger; and he who lives among the most conscious Europeans even knows it is a sickness. As one might guess, it is not the opposition between subject and object, which concerns me here; I leave that distinction to those epistemologists who have got tangled up in the snares of grammar (of folk metaphysics). Even less am I concerned with the opposition between 'thing in itself' and appearance: for we 'know' far too little to even be entitled to *make* that distinction. We simply have no organ for *knowing*, for 'truth': we 'know' (or believe or imagine) exactly as much as is *useful* to the human herd, to the species: and even what is here called *usefulness* is finally also just a belief, a fiction and perhaps just that supremely fatal stupidity of which we some day will perish".

In summa, according to Nietzsche:

- The phenomenon of consciousness cannot be grasped outside the realm of language; consciousness equals to linguistic meta-awareness.
- As such, it is inextricably intertwined with intersubjectivity; it is relational.
- Its main function is to reduce the "singular" to the "general" (sign code of the herd).
- Consciousness is not an organ for knowing, but a precondition for there being knowledge in the first place.
- The conscious 'I' is a collective narrative fiction and not a Master Judge Transcendental "Self" as Kant believed.

Incipit tragoedia!



After this short –but not superfluous– digression from Kantian transcendental philosophy through Nietzsche's physio-logical interpretation, I propose that we turn to self-reference and recursiveness and try to understand how these notions relate to the neuro-biological functional substrate of the human cognitive apparatus.

In terms of contemporary neurobiology, we could say that the human cognitive apparatus is an autopoietic system. "Autopoiesis" is a scientific term coined by two leading neurobiologists from Santiago Chile, Humberto Maturana and Francisco Varela, famous representatives of the Santiago School of biology. The noun "autopoiesis" is a compound word that comes from the Greek word "autos" meaning "self" or "oneself" and "poiesis" meaning "to produce", "to create". Autopoiesis means then "self-creation", "self-production". Simply put by Heinz von Foerster: when characterizing the cerebral mechanisms as an autopoietic system we are saying that this system is continuously producing itself;

"autopoiesis" means that the brain is a network of processes of production where the products generate their interactions in the same network that produces them.

According to Maturana and Varela,<sup>29</sup> human brains are characterized by what they call *operational closure*. The latter "designates that internal operations of the brain work in such a way that the by-product of its development remains within the neurobiological processes in the brain. Notice that to say that the brain is operationally closed is not to say that it is isolated from the external world; [...] the brain is in harmony with its surrounding environment. By *closure* we note that the human brain is endogenous; the brain is openly in synchronization with the world".<sup>30</sup> Being "openly in synchronization with the world" simply means that "over time, that is, both phylogenetically and developmentally, people establish interdependencies between the environment and their internal dynamics such that the formal becomes part of their external structure: their boundary conditions".<sup>31</sup> That being said, what then is an *environment* and how should we picture it? An *environment* is all those "perturbations, or triggers, that actuate, but do not determine changes in consciousness or behavior. The human brain is unintelligible without serious consideration of the way it is always already primed and expectant of any triggering stimuli. [...] A perturbation then is anything in the environment that triggers, but does not necessarily determine reactions from the agent. [...] The perturbation does not act causally; it does not determine action or thought. A perturbation is observable and only takes part in the inception of action and thought";<sup>32</sup> "As long as an autopoiesis is running, there appears an *environment*, which is of a special significance for the system. Contact with the environment brings about exposure in the system".<sup>33</sup> However, the relationship between the brain and the environment is not *causal*, but *interactive*, which is a totally different thing: non-causal means that despite incoming stimuli, the cause of cerebral activity is the cerebral system itself. What is abolished here is "causality": a constraint, which David Hume, the Scottish empiricist philosopher, characterized as a "habit of the mind" and which Warren McCulloch simply called a "superstition": "...let us be perfectly frank ... causality is a superstition".<sup>34</sup>

Now, you may ask, aren't all loopy systems, even the artificial ones, autopoietic systems? For example, aren't we entitled to call a state-of-the-art artificial intelligence system or a robot of the latest technology, autopoietic? The answer is negative. Only living systems are autopoietic and only autopoietic systems can be defined as living. For a system to be autopoietic it has to be organized as a network of processes of production of components that produces the components. It has to produce the very components it is made of. Only living beings exhibit this kind of recursion. When all the components of a system (cells, for example) are productively interacting with each other, producing the very components that constitute the system itself, then this system is autopoietic.

In our current historical level, artificial intelligent systems and robots do not fulfill the above condition. They cannot be considered autopoietic, because they are "biologically non-self-productive and [do not] have any self-sustenance of [their] own. [Their] actions are deliberately programmed by man so as to make [themselves] behave like a human being. The robot is simply an assemblage of parts, which cannot reproduce themselves. When it breaks down, the broken parts must be replaced from an outside system. Therefore, it is not a system running *by itself*".<sup>35</sup> Artificial intelligences and robots are *allopoietic* systems; they are not (yet) capable of constituting a "self" region of immanence.

So, as you can see, as far as our cognition is concerned, we, humans, are literally feeding on ourselves. The idea that we are harvesting fruits from reality's tree of knowledge is the greatest deception in the history of mankind: "we" are the harvesters, "we" are the eaters and "we" are the tree. By "we" I mean our common biologically determined and physically constrained cognitive system in its interaction with a community of other cognitive systems of the same kind.

Reality is what emerges through intersubjective interaction; it is a collective invention, a product that in its turn determines the network of relations that produced it – that is, human minds.

In terms of computational logic and mathematical recursion, *autopoiesis* is defined as:

"that organization which computes its own organization. [...] Autopoiesis is a notion that requires systemic closure. That means organizational, but not necessarily thermodynamic, closure. Autopoietic systems are thermodynamically open, but organizationally closed. [...] The concept of closure has recently become very popular in mathematics by calling upon a highly developed branch of it, namely, *Recursive Function Theory*. One of its concerns is with operations that iteratively operate on their outcomes, that is, they are operationally closed. Some of these results are directly associated with notions of self-organization, stable, unstable, multiple and dynamic equilibria, as well as other concepts... However, traditionally there have always been logical problems associated with the concept of closure, hence the reluctance until recently to take on some of its problematic aspects. Consider, for example, the relation of an observer to the system he observes. Under closure, he would be included in the system of his observation. But this would be anathema in a science where the rule is 'objectivity'. Objectivity demands that the properties of the observer shall not enter the descriptions of his observations. This proscription becomes manifest when you submit to any scientific journal an article containing a phrase like 'I observed that...'. The editor will return it with the correction 'It can be observed that...'. I claim that this shift from 'I' to 'it' is a strategy to avoid responsibility: 'it' cannot be responsible; moreover, 'it' cannot observe! The aversion to closure, in the sense of the observer being part of the system he observes, may go deeper. It may derive from an orthodox apprehension that self-reference will invite paradox, and inviting paradox is like making the goat the gardener. How would you take it if I were to make the following self-referential utterance: 'I am a liar'. Do I speak the truth? Then I lie. But when I lie, I speak the truth. Apparently, such logical mischief has no place in a science that hopes to build on a solid foundation where statements are supposedly either true or else false".<sup>36</sup>

This problematization of logical paradoxes and self-referentiality, as formulated by Heinz von Foerster, brings us ideally to our next section.



Keeping what has already been said about reflexive loops and consciousness in mind, I suggest that we move a little bit further and try to finally approach Hofstadter's territory.

Loops and especially strange loops can be found anywhere: in a living organization, as we saw before, in logical paradoxes, in the very structure of language, in mathematical recursion, finally in everyday activities as in the case of a simple dialogue between two persons. Let's examine a few examples:

a) Paradox

Consider the following story:<sup>37</sup>

In a little village lives a restaurant owner, who cooks only for those villagers who do not cook for themselves. If you live in the village and don't cook for yourself, the restaurant owner will cook for you. If you do cook for yourself, the restaurant owner will not cook for you. Now, in order to generate the logical paradox, one only has to pose the right question: should the restaurant owner cook for himself? The logic used to think through this question goes like this: if the restaurant owner were to cook for himself, he would belong to the class of those who "cook for themselves". If this were the case, then he should not cook for himself, because he only cooks for people who do not cook for themselves. But, if he does not cook for himself, he would then belong to the class of those who do not cook for themselves and, hence, should cook for himself.

Just like the famous "Liar paradox" invoked earlier by Heinz von Foerster, the above "restaurant owner" story shows how a logical paradox works: a paradox is generated when a statement contains a proposition, which is true when it's false and false when it's true. Consequently, paradoxical reasoning emerges whenever statements are self-referential:

"For instance: 1) This statement is false. 2) I am lying. 3) Please, ignore this notice. 4) It is forbidden to forbid. Each statement comments on itself. The moment you make self-referential statements, the logicians will immediately protest: 'You can't do that!'. 'But why not?' you might ask. 'Because', say the logicians, 'self-referential statements

produce paradoxes. They contaminate logical systems!'. Why the logician's objection to paradox? The answer is quite simple. Logicians work with declarative statements called propositions. Over 2.000 years ago, Aristotle taught that if a proposition makes sense, it must be *either true or false*. Each proposition must meet this criterion to qualify for membership in a scientific doctrine. Otherwise, it is unacceptable. Paradox renders a proposition's truth-value indeterminable. Paradoxical statements or propositions are neither true nor false. [...] The word paradox has two Greek roots: *para* meaning 'outside' and *doxein* meaning 'to point, to show, to teach'. So, paradox means 'outside of the teaching'. 'Orthodox' (from the Greek root *ortho* meaning straight) means the straightforward or inside teaching. For thousands of years, the orthodox teaching was Aristotelian".<sup>38</sup>

In short, paradoxical statements defy the foundational laws of the classical two-valued, bivalent Aristotelian logic (either [true/false] or [false/true]) as well as the principle of the excluded middle or of non-contradiction (*tertium non datur*). Actually, the first who had turned paradoxical reasoning into a profession were the Sophists:

"These Sophists were rather like traveling magicians or first-rate circus performers of our days. You paid your admission and watched the 'artist' perform his tricks. He would, for instance, single out a man from the audience and address him as follows: 'You admit, sir, that you have that which you have not lost?'. The innocent answer was: 'Of course'. 'Then, my friend', the Sophist blandly continued, 'as you never lost a tail, you must have a tail'. The performer might select a woman known to be a shrew and ask her: 'Madam, have you stopped beating your husband? Answer *yes* or *no*'. This proved an embarrassing alternative. Among those laughing at the befuddled woman was a man with a dog. The Sophist turned to him and inquired: 'Is this your dog?'. 'Yes'. 'I see it is a female dog. Has she had puppies?'. The proud owner of the dog affirmed it. Diabolically the Sophist concluded, 'This dog has two properties. First, it is your dog, and second it is mother. Let's add up the predicates: this dog is your mother'. The performer's mental gymnastics were successful because it was little known in pre-Aristotelian times that formal logic is based on a strict technique, and that the skillful 'logician' can do amazing tricks when using (or misusing) that technique. Audiences today more sophisticated and not so easily fooled".<sup>39</sup>

Thenceforth, beginning with Aristotle, philosophical and scientific discourse has been struggling to legislate upon itself for the purpose of chasing away the demons of self-referentiality and paradox: using propositions to make scientific explanations, the latter should be logically consistent. A truly difficult, if not impossible, task! As if life itself were logically consistent and non-contingent: a contradiction-free contextuality.

#### b) Language

The next example comes from cognitive linguistics. It is drawn from a fascinating study by Maurício Dias Martins, Sabine Laaha, Eva Maria Freiburger, Soonja Choi and W. Tecumseh Fitch, published in the October 2014 issue of the scientific journal *Cognition*. It is titled: "How children perceive fractals: Hierarchical self-similarity and cognitive development".<sup>40</sup>

We start with a noun, such as "committee". We all know what a "committee" is. Yet, we know nothing! And that's because there are hundreds of different kinds of committees. In order to further understand the noun's specific contextual meaning, we only have to learn one rule: that each extra noun embedded in the initial noun concretizes the meaning of the nouns to its right. So you start asking questions:

- Q: What type of committee?
- A: A film committee.
- Q: What type of film committee?
- A: A student film committee.
- Q: What type of student film committee?
- A: A graduate student film committee.
- Q: What type of graduate student film committee?
- A: A physics graduate student film committee.

Q: What type of physics graduate student film committee?

A: A particle physics graduate student film committee.

This example reveals an important part of the way we think: more precisely our capacity to generate multiple hierarchical levels with a single rule. These hierarchical levels act as a recursive development looped on itself. Each descriptor refers to the one on its right and ultimately to the noun "committee", from which may spring a potentially infinite number of descriptors. Both the noun and the descriptors, but also the descriptors between themselves, are interdependent. None can exist without the other. It's a loop that consists of productions of recursive relations.

### c) Mathematics

Now let's jump into another field, that of mathematical recursion. I suppose that you've all heard of Fibonacci numbers or Fibonacci sequences. It's a really simple procedure. You start with two numbers (1 and 1) and then you construct the next number by summing the previous two. So we have:

$$\begin{array}{l} 1+1=2 \\ 2+1=3 \\ 3+2=5 \\ 5+3=8 \\ \text{and so on...} \end{array}$$

And you can create what is called a recursive definition, where you define the thing in terms of itself:

$$f(n)=f(n-1) + f(n-2)$$

What it "is" is really itself on a smaller level.



The above examples portray adequately what is defined as a strange loop in the context of Hofstadter's *GEB* and *I am a strange loop*: a strange loop arises when, by moving only upwards or downwards through a hierarchical system, one finds oneself back to where one started.

In pages 101 and 102 of *I am a strange loop* Hofstadter gives the following explanation:

"And yet when I say 'strange loop', I have something else in mind – a less concrete, more elusive notion. What I mean by 'strange loop' is –here goes a first stab, anyway– not a physical circuit but an abstract loop in which, in the series of stages that constitute the cycling-around, there is a shift from one level of abstraction (or structure) to another, which feels like an upwards movement in a hierarchy, and yet somehow the successive 'upward' shifts turn out to give rise to a closed cycle. That is, despite one's sense of departing ever further from one's origin, one winds up, to one's shock, exactly where one had started out. In short, a strange loop is a paradoxical level-crossing feedback loop".

As we are running out of time, I'd like to close this lecture with a passage from a wonderful short article on *I am a strange loop*, published in the review *Philosophy now*:

"Hofstadter subscribes to the concept known as *the narrative self*: the notion that the idea of the self is ultimately a hypothetical construct – a story our brains spin which generates the illusion that there is a single, stable and unified locus of willing, thinking and choosing which constitutes our 'I'. We are all like Scheherazade, the queen narrating the *1001 Arabian Nights*, who postponed her execution by seducing the king with one fantastic tale after another. Similarly, our 'I' can only be sustained through an act of perpetual storytelling on 'our' brain's part. Yet who –or what–

is doing this storytelling? According to Hofstadter, the threads that make up the tapestry of a self are patterns of active symbols ('neurological patterns') that mirror the outside (and also the inside) world. [...] Perhaps Hofstadter's most intriguing argument is that the complexity and extensibility of active symbols in the brain inevitably leads to the same kind of self-reference which Gödel proved was inherent in any complex logical or arithmetical system. [...] Hofstadter argues that the psychological self arises out of a similar kind of paradox. We are not born with an 'I' – the ego emerges only gradually as experience shapes our dense web of active symbols into a tapestry rich and complex enough to begin twisting back upon itself. According to this view the psychological 'I' is a narrative fiction [...]. "It is the 'I', it is the 'I', that is deeply mysterious!" exclaimed Wittgenstein. A perspective (a mind) is therefore a consequence of a unique pattern of symbolic activity in our nervous systems. [...] Each of us is a more than just a self; we are a collection of selves. In addition to a core self, which we identify as our 'I', each of us contain neuronally-based symbolic models that mirror and reflect the other people in our lives. [...] The Cartesian prison of isolated and monadic selves is demolished, in favor of selves that are deeply enriched and entwined by their relationships to other points of view".<sup>41</sup>

Thank you for your attention and your patience.

## NOTES

---

<sup>1</sup> Douglas Hofstadter, *I am a strange loop*, Basic Books, 2007, p. 257.

<sup>2</sup> Douglas Hofstadter, "Preface to GEB's Twentieth-anniversary Edition", in: *Gödel Escher Bach – An eternal golden braid*, Penguin Books, 2000, pages xx-xxi.

<sup>3</sup> According to Zeno's "dichotomy paradox", for an object to travel a given distance  $d$ , it must first travel a distance  $d/2$ . Similarly, in order to travel a distance  $d/2$ , it must first travel a distance  $d/4$  and so on. Since this sequence goes on forever, it therefore appears that motion is impossible, since the distance  $d$  cannot be travelled.

<sup>4</sup> Ludwig Wittgenstein, *Philosophical Remarks*, trans. R. Hargreaves & R. White, Basil Blackwell, 1975, p. 5.

<sup>5</sup> Ludwig Wittgenstein, *Philosophical Investigations*, trans. G. E. M. Anscombe, Basil Blackwell, 1963, p. 329.

<sup>6</sup> Timaeus, 33 & 34, in: *The dialogues of Plato Vol. III*, Trans. B. Jowett, Oxford University Press, 1931.

<sup>7</sup> Friedrich Nietzsche, "On Truth and Lying in a Non-Moral Sense", in: Friedrich Nietzsche, *The birth of tragedy and other writings*, Cambridge University Press, 1999.

<sup>8</sup> Heinz von Foerster has provided us with a remarkable elaboration of what "becoming a metaphysician" is, which deserves to be quoted at length: "Today, when I invoke Metaphysics, I do not seek agreement with anybody else about her nature. This is because I want to say precisely what it is when we become metaphysicians, whether or not we call ourselves metaphysicians. I say that we become a metaphysician any time we decide upon in principle undecidable questions. For instance, here is a decidable question: 'Is the number 3,396,714 divisible by 2? It will take you less than two seconds to decide that indeed this number is divisible by two. The interesting thing here is that it will take you exactly the same short time to decide if the number has not 7, but 7000 or 7 million digits. I could of course invent questions that are slightly more difficult; for instance: Is 3,396,714 divisible by three? Or even more difficult ones. But there are also problems that are extraordinarily difficult to decide, some of them having been posed more than 200 years ago and remain unanswered. Think of Fermat's 'Last Theorem' [...]. Or think of Goldbach's 'Conjecture' which sounds so simple that it seems a proof cannot be too far away [...]. One of the most remarkable examples of such a crystal of thought is Bertrand Russell's and Alfred North Whitehead's monumental *Principia Mathematica*, which they wrote over a 10 year period between 1900 and 1910. This 3 volume magnum opus of more than 1500 pages was to establish once and for all a conceptual machinery for flawless deductions. A conceptual machinery that would contain no ambiguities, no contradictions and no undecidables. Nevertheless, in 1931, Kurt Gödel, then 25 years of age, published an article whose significance goes far beyond the circle of logicians and mathematicians. The title of this article I will give you now in English: 'On formally undecidable propositions in the *Principia Mathematica* and related systems'. What Gödel does in his paper is to demonstrate that logical systems, even those so carefully constructed by Russell and Whitehead, are not immune to undecidables sneaking in. However, we do not need to go to Russell and Whitehead, Gödel, or any other giants to learn about in principle undecidable questions. We can easily find them all around. For instance, the question about the origin of the universe is one of those in principle undecidable questions. Nobody was there to watch it. Moreover, this is apparent by the many different answers that are given to this question. Some say it was a single act of creation some 4 or 5,000 years ago. Others say there was never a beginning and that there will never be an end, because the universe is a system in perpetual equilibrium. Then there are those who claim that approximately 10 or 20 billion years ago the universe came into being with a 'Big Bang' whose remnants one is able to hear over large radio antennas. [...] I could go on and on with other examples, because I have not yet told you what the Burmese, the Australians, the Eskimos, the Bushmen, the Ibos, etc., would tell you

---

about their origins. In other words, tell me how the universe came about, and I will tell you who you are. I hope that I have made the distinction between decidable and, in principle, undecidable questions sufficiently clear so that I may present the following proposition which I call the 'metaphysical postulate': Only those questions that are in principle undecidable, we can decide. Why? Simply because the decidable questions are already decided by the choice of the framework in which they are asked, and by the choice of the rules used to connect what we label 'the question' with what we take for an 'answer'. In some cases it may go fast, in others it may take a long, long time. But ultimately we arrive after a long sequence of compelling logical steps at an irrefutable answer; a definite 'yes or a definite 'no'. But we are under no compulsion, not even under that of logic, when we decide on in principle undecidable questions. There is no external necessity that forces us to answer such questions one way or another. We are free! The compliment to necessity is not chance, it is choice! We can choose who we wish to become when we have decided on an in principle undecidable question. [...] Objectivity, Pontius Pilate, hierarchies, and other devices are all derivations of a choice between a pair of in principle undecidable questions which are: 'Am I apart from the universe?', meaning whenever I look, I'm looking as if through a peephole upon an unfolding universe; or: 'Am I part of the universe?', meaning whenever I act, I'm changing myself and the universe as well. Whenever I reflect on these two alternatives, I'm surprised by the depth of the abyss that separates the two fundamentally different worlds that can be created by such a choice. [...] I have a dear friend who grew up in Marakesh. The house of his family stood on the street that divides the Jewish and the Arabic quarters. As a boy, he played with all the others, listened to what they thought and said, and learned of their fundamentally different views. When I asked him once who was right he said: 'They are both right'. 'But this cannot be', I argued from an Aristotelian platform, 'Only one of them can have the truth'. 'The problem is not truth', he answered, 'The problem is trust'. I understood. The problem is understanding. The problem is understanding understanding! The problem is making decisions upon in principle undecidable questions". In: Heinz von Foerster, *Understanding understanding: Essays on cybernetics and cognition*, Springer, 2003, pp. 291-295.

<sup>9</sup> G. Deleuze-F. Guattari, *What is philosophy?*, trans. H. Tomlinson & G. Burchill, Verso 1994.

<sup>10</sup> How really "real" is that "real"? Isn't the latter always already conditioned, historically-culturally mediated and formalized?

<sup>11</sup> Susan Buck-Morss, "Aesthetics and Anaesthetics: Walter Benjamin's Artwork Essay Reconsidered", in: *October*, Vol. 62 (Autumn, 1992), pp. 3-41.

<sup>12</sup> *Ibid.*

<sup>13</sup> *Ibid.*

<sup>14</sup> A "turn of the century" German pharmacologist and physiologist (1873-1927), who best embodies the *Geist* of the philosopher-scientist. He was convinced that he could actually spot the kantian *a priori* in the brain's physiology: "Inspired by Immanuel Kant, [Rudolf Magnus] made his last great lecture one on 'the physiology of the a priori', by which he meant the go of those mechanisms that determine for us the three-dimensional nature of our world, its axes and its angles, and that give to us our sense of velocity and acceleration, from which he held out notion of time to be in large measure derived" (Warren McCulloch, "Through the den of the metaphysician", in: *Embodiments of mind*, The MIT Press, 1988, pp. 142-156).

<sup>15</sup> "From Vienna to California. A journey across disciplines – An Interview with Heinz von Foerster", Stefano Franchi, Güven Güzeldere and Erich Minch (2005). In: *Kybernetes*, 34(1/2), 15-32.

<sup>16</sup> Félix Guattari, *Chaosmosis: An Ethico-Aesthetic Paradigm*, trans. P. Bains & J. Pefanis, Indiana University Press, 1995.

<sup>17</sup> *Ibid.*

---

<sup>18</sup> John Haugeland, "What is mind design?", in: *Mind Design II – Philosophy, Psychology, Artificial Intelligence*, edited by John Haugeland, The MIT Press, 1997, pages 1-28.

<sup>19</sup> There exist several English editions of the *Critique of pure reason*. Yet, the one that is widely considered the best is the following: P. Guyer, and A. Wood (eds.), *Critique of Pure Reason*, Cambridge: Cambridge University Press, 1998.

<sup>20</sup> Kurt Gödel, *Collected Works Volume IV. – Correspondence A-G*, edited by Solomon Feferman, John W. Dawson Jr., Warren Goldfarb, Charles Parsons & Wilfried Sieg, Oxford University Press, 2003, pages 457-476.

<sup>21</sup> Originally published in: *Startling Stories*, Vol.29, no.1, pp. 110-116, New York 1953. Published online in: <http://www.vordenker.de/>

<sup>22</sup> The term "consciousness" does not appear in Aristotle's body of work, at least not in the context we are using it for the purposes of our presentation. However, Aristotle has bequeathed us a significant number of concepts and arguments important in philosophy of consciousness especially in his treatise *On the Soul*. For a detailed exposition and analysis of Aristotle's theses on the issue, see Victor Caston's paper titled "Aristotle on consciousness", published in *Mind*, Vol. 111, 444, October 2002, pp. 751-815 (Oxford University Press): "Aristotle's discussion of perceiving that we perceive (*On the Soul*, 3.2) has points of contact with two contemporary debates about consciousness: the first over whether consciousness is an intrinsic feature of mental states or a higher-order thought or perception; the second concerning the qualitative nature of experience. In both cases, Aristotle's views cut down the middle of an apparent dichotomy, in a way that does justice to each set of intuitions, while avoiding their attendant difficulties".

<sup>23</sup> Amanda Geffer, "The man who tried to redeem the world with logic", in: *Nautilus* science magazine, "Information – The anatomy of everything", Issue 021, 02/2015.

<sup>24</sup> H. Caygill, *A Kant dictionary*, Blackwell Publishers Ltd., 1995, pp. 79-81.

<sup>25</sup> *Ibid.*

<sup>26</sup> Gottfried Wilhelm Leibniz, *New essays on human understanding*, ed.: P. Remnant & J. Bennett, Cambridge University Press, 1996, pp. 3-19.

<sup>27</sup> Warren McCulloch, *The past of a delusion*, Chicago Literary Club, 1953, p. 22.

<sup>28</sup> Friedrich Nietzsche, *The Gay Science*, Trans. J. Naucköf, ed. B. Williams, Cambridge University Press, 2001, pp. 211-214.

<sup>29</sup> The themes discussed in this section were introduced and developed by H. Maturana and Fr. Varela in their two major works: *Autopoiesis and cognition: The realization of the living*, Dordrecht, UK: Reidel, 1980 (Original work published 1972); *The tree of knowledge: The biological roots of human understanding*, Boston, MA: Shambhala Publications, 1987.

<sup>30</sup> Robert Garfield McInerney, "Neurophenomenological praxis: Its applications to learning and pedagogy", in: Susan Gordon (ed.), *Neurophenomenology and its applications to psychology*, Springer, 2013, p. 40.

<sup>31</sup> A. Juarrero, *Dynamics in action: Intentional behavior as a complex system*, Cambridge, MA: The MIT Press, 1999, p. 197.

<sup>32</sup> Robert Garfield McInerney, "Neurophenomenological praxis: Its applications to learning and pedagogy", in: Susan Gordon (ed.), *Neurophenomenology and its applications to psychology*, Springer, 2013, pp. 40-41.

<sup>33</sup> Moriyama Shigeru, "What is Life and Environment? Autopoiesis of Life World" (<http://www.asahinet.or.jp/~ns3s-mrym/autopo.htm>).

---

<sup>34</sup> Cited in: Heinz von Foerster, "Metaphysics of an experimental epistemologist", 1995. Published online in: <http://www.vordenker.de/>. See also: Warren S. McCulloch, "What is a number, that a man may know it, and a man, that he may know a number?", in: *Embodiments of mind*, The MIT Press, 1988, pp. 1-19.

<sup>35</sup> Moriyama Shigeru, "What is Life and Environment? Autopoiesis of Life World" (<http://www.asahi-net.or.jp/~ns3s-mrym/autopo.htm>).

<sup>36</sup> Heinz von Foerster, "Disorder/order: Discovery or invention?", in: *Understanding understanding: Essays on cybernetics and cognition*, Springer, 2003, pp. 273-282.

<sup>37</sup> Modified replication of the "Barber's paradox", mentioned in: Lynn Segal, *The dream of reality—Heinz von Foerster's constructivism*, Springer, 2001, page 39.

<sup>38</sup> Lynn Segal, *The dream of reality—Heinz von Foerster's constructivism*, Springer, 2001, page 39.

<sup>39</sup> Aristotelian and non-Aristotelian logic", first published in: *Startling Stories*, N.Y.: 1954, vol. 32, p. 102-108. Published online in: <http://www.vordenker.de/>

<sup>40</sup> *Cognition*, Volume 133, Issue 1, October 2014, pp. 10–24.

<sup>41</sup> Scott O' Reilly, Book presentation of *I am a strange loop*, in: *Philosophy now*, Vol. 78, April/May 2010, pages 39-41.