The Hidden Logic of The Mind and the Universal Learning Problem.

Misha Gromov

May 3, 2015

Contents

1 Memorandum Ergo.

It is not knowledge, but the act of learning,... which grants the greatest enjoyment.

CARL FRIEDRICH GAUSS

The ultimate aim of our *ergo project* is designing a *universal learning program* that upon encountering an *interesting flow of signals* e.g. representing a natural language, starts *spontaneously interacting* with this flow and will eventually arrive at *understanding* of *the meaning* of messages carried by this flow.

We do know that such programs exist, we carry them in the depths of our MIND, in what we call *ergo-brain*, but we have no inkling of what they are.

Prior to embarking toward design of such programs we proceed with

• assessing *flows of signals* commonly encountered in life from a mathematical perspective and formalising what we find *interesting* about them;

• describing, let it be in general, yet, mathematical, terms, what the words *learning, understanding, meaning* signify;

• working out general conceptual guidelines for ergo-learning.

When approaching these issues, we shall be following the principles of what we call the *ergo logic*, thus, *distancing ourselves* from the *common sense* ideas about the Human Mind that are dominant in our *ego-mind* and that are pervasive in our *(Popular) Culture.*

Ego-mind is a part of a greater MIND; in fact, it is the part you normally perceive as your mind, but MIND, as we understand it, also contains ergo-brain that, unlike ego-mind, is inaccessible to your mind eye.

Schematically, MIND is a *finite connected graph* (network of ideas), that decomposes into two subgraphs (very roughly) corresponding to the ergo-brain and the ego-mind,

$$\text{MIND} = M_{ergo} \cup M_{ego}$$

where $M_{ergo} \subset$ Mind, a kind of a *core* of the MIND, is a *union of cycles* and $M_{ego} \subset$ Mind, a *periphery*, is *a disjoint union of trees* T such that each T meets M_{ergo} at a single vertex – the root of T from where it grows.[1]

Common (popular) culture is a kind of the collective ego-mind while science and mathematics are parts of our *collective ergo.*

Ergo is *irrational*; it is after beautifully interesting structures in the world, not practically useful ones, it is enthralled by play, art, science.

Ego is *rational*: common sense – the logic of "ego" carries accumulated evolutionary wisdom needed for our personal survival and that of our genes.

Common sense ideas and opinions are unquestionably self-evident you are not suppose to overrun them. For instance,

if something heavy falls on you – dodge out of the way as fast as you can: heavy objects fall faster than light ones and they hit you harder.

This is great for your survival. But if the source of you idea *in science* is common sense, search for another idea.

Much of what we said must be mystifying to you; do not worry, just remember the ergo-words – they will acquire meaning as we go on.

1.1 Universality, Simplicity and Ergo Brain.

Out of chaos God made a world,
and out of high passions comes a people.
BYRON.

Our fascination by learning systems comes from what may seem as an almost godlike ability of a human (and some animal) infant's brain of building a *consistent* model of *external world* from an *apparent* chaos of *flows of electric/chemical signals* that come into it.

Imagine, you see on a computer screen what a baby brain "sees": a "throbbing streaming crowd of electrified shifting points" encoding, in an incomprehensible manner, a certain never seen before, not even imaginable, "reality". Would you reconstruct anything of this "reality"? Would you be able to make such concepts as "shadow", "roundness", "squareness"?

[1] Of course, all finite graphs decompose this way.

Could you extract any *meaning* from a Fourier-like transform of the sound wave the brain auditory system receives?

No, this ability is lost by "mature minds". One can not even recognize 2-dimensional images by looking at graphical representations of the illumination levels, which is a much easier problem. What a baby chimpanzee's brain does is more "abstract" and difficult than the recently found solution of the *Fermat's Last Theorem*.

Yet, we conjecture that an infant's ergo-brain operates according to

an universal set of simple learning rules.

The ergo-brain *extracts structural information* "diluted" in flows of signals following these rules and continuously *rebuild itself* by incorporating this structure.

(It would be unrealistic making any conjecture on how such rules could be implemented by the neurophysiology of the human brain, although it seems plausible that they are incorporated into the "architecture of pathways" of signal processing's in the brain. But we shall try to guess as much as possible about these rules by looking at the universal learning problem from a mathematical perspective.)

This *idea of universality* is supported by capability by deaf and deaf blind people to learn native languages even when these are embedded *not in auditory* flows of signals. but in *visual* and/or *tactile* ones.

Also human capacity for *mathematics* (capacity to understand the proof of the Last Fermat Theorem, for instance) points toward simplicity and universality of the programs run by the ergo-brain. A specialised and/or complicated learning program, besides being evolutionary unfeasible, could hardly do mathematic that is far removed from the mundane activities the ergo brain was "designed" for.

Yet, at the moment, one may only speculate in favor of universality by appealing to "evolutionary thrift of Nature" and to "brain plasticity".

Universality is the most essential property we require from the learning systems/programs which we want to design – these programs must *indiscriminately* apply to *diverse classes* of incoming signals regardless of their "meanings" using the same toolbox of rules for learning languages, chess, mathematics and tightrope walking.

Without universality there is no chance of a non-cosmetic use of mathematics;[2] and only "clever mathematics" may furnish universality in learning.[3]

Ultimately, we want to write down a *short* list of *general* guidelines for "extracting" *mathematical structures* from *general* "flows of signals". And these flows may come in many different flavors – well organized and structured as mathematical deductions processes, or as unorderly as *"a shower of little electrical leaks"* depicted by Charles Sherrington in his description of the brain.

[2] This is meaningless unless you say what kind of mathematics you have in mind. Mathematical creatures, such, for example, as *Turing machine* and *Pythagorean theorem* differ one from another as much as a single-stranded RNA virus form a human embryo.

[3] Our objectives are different from those taken by *mathematical psychologists* (see e.g. [?], [?],...) as we are not so much concerned with modelling Human Mind but rather the "invisible" processes that shape the Mind.

Of course, nontrivial structures can be found by a learning system, (be it universal or specialized) only in "interesting" flows of signals. For instance, nothing can be extracted from fully random or from constant flows. But if signals are modulated by *something* from "real world" we want to reconstruct as much of the *mathematical structure* of this *something* with these rules as the brain of a human infant can do.

. UNIVERSAL PATTERNS IN ANIMAL BEHAVIOUR.

Why Universal? It would be unrealistic to expect that the evolution had time enough to select for many long sequential learning programs specific to different goals, but "clever" universal programs may accommodate several different situations.

Example: Hawk/Goose Effect. A baby animal distinguishes *frequently observed* shapes sliding overhead from those that appear *rarely.*

The former eventually stops soliciting HIDE! response, while every unusual kind of a shape, e.g. that of a hawk, makes an animal run for cover.

Such "universal programs" develop in the environment of evolutionary older *general "ideas"* that are kind of *tags*, such as dangerous/harmless, edible/useless, etc., selectively associated in an animal/human brain with a variety of *particular* "real somethings" in the world.

NO universality, however, in biology is mathematically perfect – "laws of biology" are no brothers of laws of physics.

The genetic code – a *specific* correspondence between the two set accidentally frozen in time,

$\{A,T,C,G\}^3 \leftrightarrow$ {❀, ✲, ❊, ✾, ❁, ❉, ✳, ✹,✶, ✤,✺,✷, ✿,✳,✶, ❃,❄,❁, ✱,✿,✖}

may be the most general law of Life from a biologist point ov view.

But the primal universality in biology for an ergo-minded mathematician is seen in

one-dimensionality of polynucleotides and polypeptides,
digital nature of genetic code and
"*3D information transfer principle*" implemented by *folding of heteropolimers*.

Technical (Im)Practicality of Universality. Multi-purpose gadgets are not among Greatest Engineering Achievements of the Twentieth Century: *flying submarines*, if they were a success, then only in James Bond films.[4] On the other hand, the 20th century machine computation has converged to universality; the basic machine learning will, most probably, follow this path in the 21st century.

1.2 Freedom, Curiosity, Prediction and Goal Free Learning.

The essence of mathematics lies in its freedom.
GEORGE CANTOR

These words by Georg Cantor equally apply to learning in place of mathematics. Universality necessitates *non-pragmatic* character of learning. Indeed, *formulat-*

[4]There are sea birds, e.g. *pelagic cormorants* and *common murres* who are (reasonably) good flyers and who also can dive, some up to more than 50(150?)m. The technology for building comparably universal/adaptable machines may be waiting ahead of us.

ing each utilitarian goal is *specific* for this goal – there is no universal structure on the "set of goals". Thus,

the essential mechanism of learning is goal free
and independent of an external reinforcement,[5]

where the primary example of free learning is the first language acquisition.

The ability of native learning systems to function with

no purpose, no instruction, no reinforcement

is no more paradoxical than, say, a mechanical system moving in an absence of a force.

External constrains and forces change the behaviour of such systems, but *inertia* remains *the source* of motion. (The use of metaphors in science leads to confusion. The force of gravity is what *makes* things fall but it can hardly be called the *source* of motion of Earth around the Sun.)

Closer to home, think of your digestive system. The biochemistry of metabolic networks in the cells in your body needs no teacher instruction, albeit hunger initiates the digestive process.

Similarly, you may start learning to play chess or to walk a tightrope in order to impress your peers, but the learning program(s) in you (ergo)brain carries no trace of this purpose.

Ergo-Systems. These are universal learning systems that we want to design. They also must be self-propelled learners that learn spontaneously with no need for instructions and reinforcement. (Strictly speaking, our concept of ergo-system is broader, in particular it does nor exclude native ergo-brain learners.)

Cuiriosity as Intrinsic Motivation. The idea of what we call *ergosystems* is close to what was earlier proposed by Schmidhuber and by Oudeyer, Kaplan and Hafner,[6] in the context of robotics under the name of *Intrinsically Motivated Curiosity Driven Robots.*

This "motivation" is implemented by a class of *predictor programs*, that depend on a parameter B which is coupled with (e.g. by being a function of) the behaviour of robots.

These programs $Pred = Pred(H, B)$ "predict" in a certain specified way incoming signals on the basis of the history H, while the robots (are also programed to) optimize (in a specific formally defined way) the quality of this prediction by varying B.[7]

This "freedom" for an ergo-brain is not just a possibility to generate any kind of signals it "wants", but rather to have "interesting" environmental responses to these signals.

For instance, a bug crawling on an *infinite* leaf has *zero* freedom: no matter where it goes it learns nothing new. But an accessible edge of the leaf, adds to

[5]Feeling of pain when you fall down or bump into something may be helpful in learning to run – this is debatable; but contrary to what a behavioristically minded educator would think, *reward/punishment reinforcement* does not channel the learning process by *reinforcing* it, but rather by *curtailing and constraining* it. Compare [13] [11].

[6]See [Schmidthuber] –*Formal Theory of Fun and Intrinsic Motivation and Creativity* on http://www.idsia.ch/~juergen/ and [OKH] – (Oudeyer, P., Kaplan, F., Hafner, V.V.: Intrinsic Motivation Systems for Autonomous Mental Development. IEEE Transactions on Evolutionary Computation 11:1, (2007) and [www.pyoudeyer.com].

[7]Universal learning problems are being designed by Jürgen Schmidhuber and his team and *curiousity driven robots* and build in Ouduer's lab.

bug's "freedom".

Thus, an ergo-brain comes to "understand" the world by "trying to maximize" its "predictive power" but what the ergo-brain exactly predicts at every stage depends on what structure has been already built.

In order to maximise anything, one needs some freedom of choice, e.g. your eye needs a possibility to run along lines/pages or, in a chess game, you can choose from a certain repertoire of moves.

When this repertoire becomes constrained, the ergo-brain feels *bored* and *frustrated*. This happens to you when a pedantic lecturer curbs your curiosity by displaying slides on the screen line by line, preventing you from seeing the whole page.

1.3 Ego, Ergo, Emotions and Ergo-Moods.

One may understand the cosmos, but never the ego; the self is more distant than any star.

GILBERT K. CHESTERTON

Our main premises is that learning mechanisms in humans (and some animals) are *universal, logically simple and goal free*. An organized totality of these mechanisms is what we call *ergobrain* – the essential, albeit nearly invisible, "part" of human mind – an elaborate mental machine that serves as an *interface* between the neuro-physiological brain and the *(ego)*mind.

Metaphorically, this "invisible" is brought into focus by rewriting the Cartesian

I THINK therefore **I** AM

as

cogito ERGO sum.

"I think" and *"I am"* are what we call *ego-concepts* – structurally shallow products of *common sense*. But ERGO – a mental transformation of the seemingly *chaotic flow* of electric/chemical signals the brain receives into a coherent picture of *a world* that defines your personal idea of existence has a beautifully organized *mathematical structure*.

Apparently, MIND contains two quite different separate entities, that we call *egomind* and *ergobrain*.

Ego-mind is what you see as your personality. It includes all what you perceive as your conscious self – all your thoughts, feelings and passions, with subconscious as a byproduct of this *ego*.

Ego is rational. The *core ego-mind* is shaped by the *evolutionary selection* that had been acting on tens of millions of generations of our animal forebears. The ideas (and actions) generated by the ego-mind serve your survival and reproduction needs.

Besides the ego-mind carries imprints of *the popular culture* of the social group an individual belongs to.

Ego-processes are observed in the behavior of human and animals and some are perceived by retrospection.

Egomind is "real", large and *structurally shallow*. Most (all) of what we know of egomind is expressible in *the common sense* language that reflects the logic of ego-mind. This language is adapted to our social interaction; also it suffices for expressing ideas in *the theory of mind* of *folk psychology*.

Ego-mind is responsible for WHYs about your thougts; if you want to understand HOWs you must turn to the *ergo-brain*.

Ergobrain, logically, mediates between electrochemical dynamics of neuronal networks in the brain and to what we perceive as our "thinking".

Ergobrain is something abstract and barely existing from ego's point of view. Ultimately, ergobrain is describable in the language of what we call (mathematical universal learning) *ergosystems* but it is hard to say at the present point what ergobrain truly is, since almost all of it is invisible to the conscious (ego)mind. (An instance of such an "invisible" is the mechanism of *conditional reflexes* that is conventionally regarded as belonging with the brain rather than with the mind.)

Ergobrain, unlike egomind, is a structural entity, which underlies deeper mental processes in humans and higher animals; these are not accessible either to retrospection or to observations of behavior of people and/or animals. This makes the ergobrain difficult (but not impossible) for an experimental psychologist to study. (Folk psychology, psychoanalysis and alike are as unsuitable for looking into the depths of the mind as astrology for the study of the synthesis of heavy atomic nuclei in supernovae.)

The ergobrain and the egomind are autonomous entities. In young children, human and animals, the two, probably, are not much separated; a presence of *ergo in the mind* is visible in how children think about play.

As the egomind ("personality", in the ego-language) develops it becomes protected from the ergobrain by a kind of a wall. This makes most of ergobrain's activity invisible.

In grown ups, ergo, albeit reluctantly, may comply with demands by ego:

"*Concentrate and solve this damn problem! – I need a promotion.*"

But the two can hardly tolerate each other.

Human ergo has a seriousness of a child at play. As a child, it does not dutifully follow your instructions and does not get willingly engaged into solving your problems. This irritates ego. From the ego perspective what ergo does, e.g. composing utterly useless chess problems, appears plain stupid and meaningless.

Reciprocatory, utilitarian ego's activity, e.g. laboriously filing in tax return forms, is dead boring for ergo.

Certain aspects of ergo may be seen experimentally, e.g. by following *saccadic eye movements*, but a direct access to ergo-processes is limited.[8]

But there are properties of the working ergo in our brain/mind that are, however, apparent.

For example, the *maximal number* $N_\circ$ of concepts our ergobrain can manipulate with *without structurally organizing them* ("chunking" in the parlance of psychologists) equals three or four.[9] This is seen on the conscious level but such

[8]This is similar to how it is with the cellular/molecular structures and functions, where the "ergo of the cell", one might say, is the machinery controlled by the *housekeeping genes that is not directly involved in any kind of production by the cell.*

[9]Some people claim their $N_\circ$ is as large as (Miller's) "magical seven" but this seems unlikely

a bound is likely to apply to all signal processing by the ergobrain.

For instance, this $N_\circ$ for (the rules of) chess is between three and four: the three unorganized concepts are those of "rook", "bishop" and "knight", with a weak structure distinguishing king/queen.

Similar constrains are present in the structures of natural language where they bound the number of times operations allowed by a generative grammar may be implemented in a single sentence. [10]

Animal (including human) emotional responses to external stimuli seem rather straightforward with no structurally elaborate ergo mediating between neuronal and endocrine systems.

We think of emotions as colors or typefaces – a few of dozen of different kinds of them, which the brain may choose for writing a particular message, such as

run! run! RUN! RUN!

On the other hand, distinct *ergo-moods*, such as being

curious, interested, amused, amazed, perplexed, bored,

serve as indicators as well as dynamic components, of the activity of the ergo-brain.

These indicators tell us how far our ergo-brain is from animal rationality.

Our visual system is *amused* by optical illusions, *amazed* by tricks of magicians, *fascinated* by performance of gymnasts.

Our auditory system is *enchanted* by music.

Our olfactory system is attracted by exotic perfumes.

Our gustatory system is hungry for strange and often dangerously bitter foods.

Our motor/somatosensory system plays with our bodies making us dance, walk on our hands, perform giant swings on the high bar, juggle several unhandy objects in the air, climb deadly rocks risking our lives, play tennis, etc.

Ergo-moods, being independent of the pragmatic content of the signals received by the ergo-brain, serve as universal signatures/observable of ergo-states.

These moods are apparent as reactions to *external* signals by the ergo-brain; we conjecture that similar signatures mark and guide *the internal ergo-processes* as well.

1.4 Common Sense, Ergo Ideas and Ergo Logic.

Einstein, when he says that

common sense is the collection of prejudices acquired by age eighteen

does not try to be intentionally paradoxical. There is a long list of human conceptual advances based on *non-trivial* refutations of the *old way which is*

from our mathematical perspective; also some psychologists also find the number four more realistic.

[10]An often repeated statement that "one *can* potentially produce an *infinite* number of sentences in any language" is, to put it politely, a logical misdemeanour.

The only meaningful concept of "infinite" belongs with mathematics while there is no no room for the concept of "can" within mathematics proper. (Hiding behind "potentially" or appealing to such definition as "a language is as sets of strings..." does not help.)

also the common-sense way.[11] The first entry on this list – *heliocentrism* – was envisioned by by Philolaus, albeit not quite as we see it today, twenty four centuries ago. The age of enlightenment was marked by the counterintuitive idea of Galileo's *inertia*, while the 20th century contributed *quantum physics* – *absurd from the point of view of common sense* – in Richard Feynman's words. (Amusingly, Einstein sided with common sense on the issue of *quantum*.)

The core of your *ego-mind self* with its *pragmatic ego-reasoning* – common sense as much as your emotional self, is a product of evolutionary selection with the final touch accomplished by the cultural pressure. The two "selves" stay on guard of your survival, social success and passing on your genes.

But *ergo*, unlike ego, was not specifically targeted either by evolutionary selection nor by the pressure of any the popular culture – it was adopted by evolution out of sheer logical necessity as, for example, the *1-dimensionality* of DNA molecules. And ergo is often in discordance with the dominant cultural traditions of one's social environment.

A pragmatically teleological ego-centered mode of thinking that was installed by evolution into our conscious mind along with the caldron of *high passions* seems to us intuitively natural and logically inescapable. But this mode was selected by Nature for[12] our social/sexual success and personal survival, not at all for a structural modeling of the world including the mind itself.

The self-gratifying ego-vocabulary of

intuitive, intelligent, rational, serious, objective,
important, productive, efficient, successful, useful.

will lead you astray in any attempt of a rational description of processes of learning; these words may be used only metaphorically. We *can not*, as Lavoisier says,

to improve a science without improving the language or nomenclature which belongs to it.

The intuitive common sense concept of *human intellegence* – an idea insulated in the multilayered cocoon of *teleology* –purpose, function, usefulness, survival, is a persistent human illusion. If we want to to understand the *structural essence* of the mind, we need to to break out of this cocoon, wake up from this illusion and pursue a different path of thought.

It is hard, even for a mathematician, to accept that your conscious mind, including the basic (but not all) mathematical/logical intuition, is run by a blind evolutionary program resulting from "ego-conditioning" of your animal/human ancestor's minds by million years of "selection by survival" and admit that mathematics is the only valid alternative to common sense.

Yet, we do not fully banish common sense but rather limit its use to concepts and ideas *within* mathematics. To keep on the right track we use a semi-mathematical reasoning – we call it *ergologic* – something we need to build along the way. We use, as a guide, the following

ERGOLIST OF IDEAS.

interesting, meaningful, informative, funny, beautiful,
curious, amusing, amazing, surprising,

[11]This is the way of thinking by *a plain, reasonable working man* as Lev Tolstoy tells his readers.

[12]This embarrassing "for" is a fossilized imprint of the teleological bent in our language.

confusing, perplexing, predictable, nonsensical, boring.

These concepts, are neither "objective" nor "serious" in the eyes of the egomind, but they are *universal*. By contrast, such concept as "useful", for instance depends on what, *specifically*, "useful" refers to.

Hopefully ergo logic and ergo-ideas direct us toward developing ergo-programs that would model leaning processes in a children's minds. After all, these minds can hardly be called serious, rational or objective.

It is difficult to bend you ego-mind to the ergo-way of thinking. This, probably, why we have been so unsuccessful in resolving the mystery of the Mind.

Chimpanzee Model.

We can not learn much about ergo by a study of animal behaviour,[13] but out egos are similar to those of animals. This is seen in the following experiment performed by Sarah Boysen more than 20 years ago.

X (Sarah) and Y (Sheba) were Chimpanzees who learned the concepts "more than" and "less than" and who adored gumdrops, the more the better.

While Y watched, X was asked to point to one of the two plates on the table: a "large" one, with many gumdrops and a "small" one, with few of them. Whichever plate X pointed to was given to Y.

Try after try, X was pointing to the "large" plate and receiving only a few gumdrops. Apparently X realised it was behaving stupidly but could not override the "grab what you can" drive.

Then gumdrops were replaced by plastic chips. Now X was invariably pointing to the "small" plate thus receiving more gumdrops than Y.[14]

‘

1.5 Ergo in our Minds.

Those who dance are often thought mad
by those who hear no music.
Tao Te Ching.

The most dramatic evidence for the existence of an unbelievably powerful survival indifferent mental machinery in our heads comes from the rare cases where *the ergo insulating wall* has *"leakages"*.

The legacy of evolution keeps "ergo-power" in our minds contained by an *ego-insulating wall*: a hunter-gatherer whose ergobrain had overrun his/her pragmatic egomind did not survive long enough to pass on his/her genes.

But if in older times, people with such "leakages" in their ergo brains had no chance for "survival", in today's civilised societies they may live; they shine like "mental supernovas" unless their fire has been stifled by educational institutions.

Srinivasa Ramanujan (1887 – 1920) was the brightest such supernova in the Universe of Mathematics; only accidentally, due to intervention of Godfrey Harold Hardy, he's got a chance to became visible.

---

[13] There are exceptions. Orangutangs, for instance, have propensity for *3D topology*. They may enjoy playing with knots as much as human mathematicians do.

[14] Abstract "more" and "less" are not ingenious of ergo-brain as we shall argue later on. Chimpanzees' "more"/"less" for food (depending on the intensity of the smell?) and for non-edible items might be located in mutually disconnected parts of their brains/minds.

We invite the reader to find yet another interpretion of this experiment besides this and the obvious one.

When he was 16, Ramanujan read a book by G. S. Carr. "A Synopsis of Elementary Results in Pure and Applied Mathematics" that collected 5000 theorems and formulas. Then in the course of his short life, Ramanujan has written down about 4000 new formulas, where one of the first was

$$\sqrt{1+2\sqrt{1+3\sqrt{1+4\sqrt{1+5\sqrt{1+\cdots}}}}}=3.$$

During his life, Ramanujan recorded his discoveries in four notebooks. The fourth notebook–a bunch of loose pages– the so-called "lost notebook" with about 650 of Ramanujan's formulas, most of them new, was rediscovered in 1976 by George Andrews in the Library at Trinity College. George Andrews and Bruce Berndt collaborated on publishing for volumes (appearing in 2005, 2009, 2012, 2013) of the proofs of Ramanujan's formulas included in the lost notebook. http://www.math.uiuc.edu/~berndt/lostnotebookhistory.pdf

Writing off "Ramanujans" and "Mozarts" to "mere accidents",[15] is like judging *explosions of supernovae* "irrelevant", just because only a dozen of supernovae were recorded in our galaxy with billions stars (none since October 9, 1604).

The hidden mental power of everybody's (ergo)brain, not only of Ramanujan's brain, must be orders of magnitude greater than what is available to the ego-mind, since *rare* mental abilities could not have been evolutionary selected for and structurally complex functional features (be they anatomical or mental) can not come by an accident.[16]

What kind of mathematical structure could adequately describe "mysterious something" in the human brain/mind that caused the transformation from the flow of written symbols from Carr's book to the formulas written by Ramanujan?

Unless we develop a fair idea of what such a structure can be, we would not accept any speculation either on the nature of mathematics or of the human mind, be it suggested by psychologists or by mathematicians.

Further evidence in favor of ergobrain – a universal mathematically elaborate machine hidden in *everybody's* head that is responsible for non-pragmatic mechanism(s) of learning can be seen in the following.

1. *Spontaeneous learning mother tongues by children.*

Albeit human speech depends our inborn ability to distinguish and to articulate a vast variety of phonemes, the structural core of learning mother Language goes according some *universal rules* that are not bound to a particular physical medium supporting a "linguistic flow". Learning languages and writing poetry by deaf-blind people is a witness for this.

2. *Learning to to read and to write.*

This, unlike learning to speak, has no evolutionary history behind it.

3. *Mastering bipedal locomotion.*

[15]Ergo-logic, unlike the insurance companies, assigns significant weights to *miraculously improbable* events.

[16]The development of the brain is a random process, where only its general outline is genetically programmed. Rare fluctuations of some average "connectedness numbers" can be further amplified by "Hebbian synaptic learning". To properly account for this one has to argue in terms not of individual ergo-brains but of *(stochastic) moduli spaces of ergo-brains.*

One is still short of designing bipedal robots that would walk, run and jump in a heterogeneous environment.

• *Human fascination by sophisticated body movements: dance, acrobatics, juggling.*

4. *Playful behaviour of some animal, e.g. human, infants during the periods of their lives when the responsibility for their survival resides in the paws of their parents.*

5. *Attraction to useless from survival perspective activities by humans, such as climbing high mountains and playing chess.*

Albeit rarely, adult animals, e.g. dolphins, engage in similarly useless playful actions.

6. *Creating and communicating mathematics.*

Probably, several hundred, if not thousands or even millions, people on Earth have a mental potential for understanding *Fermat's last theorem.*[17] by reading a thousand-page *written proof* of it.

The following example demonstrates human ergo in all its illogical beauty.

A 4-5 years old child who sees somebody balancing a stick on the tip of the finger, would try to imitate this; eventually, without any help or approval by adults, he/she is likely to master the trick.[18]

What is the mathematics behind this?

A naive/trivial solution would be reformulating the problem in terms of classical mechanics and control theory. The balancing problem is easily solvable in these terms but this solution has several shortcomings:

• It does not apply where the external forces are unknown.

• It does not scale up: no such robot came anywhere close do a healthy human in its agility.

• It suggests no universality link between balancing sticks and $\sqrt{1+2\sqrt{1+\cdots}}$.

• It distract you from the key issue:

What on Earth drives children to try to perform such the tricks?

(Younger children at play enjoy putting pencils vertically on their non-sharp ends on the table. And if captured and caged by an an extraterrestrial, you would have no better way to prove your "non-animal mentality" but by putting a stick vertically in the *centre* of you cage.)

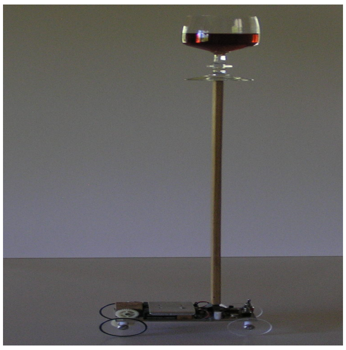
A simple (ergo-style) solution of the balance problem with a single degree of freedom – the inclination angle α, may be obtained by following $grad_v(T)$ for $T = T(\alpha, \alpha', v)$ being the empirical "falling time" where α' denotes the angular velocity and v is the control parameter – the (horizontal) velocity of the support.

A Few Words about Stars.

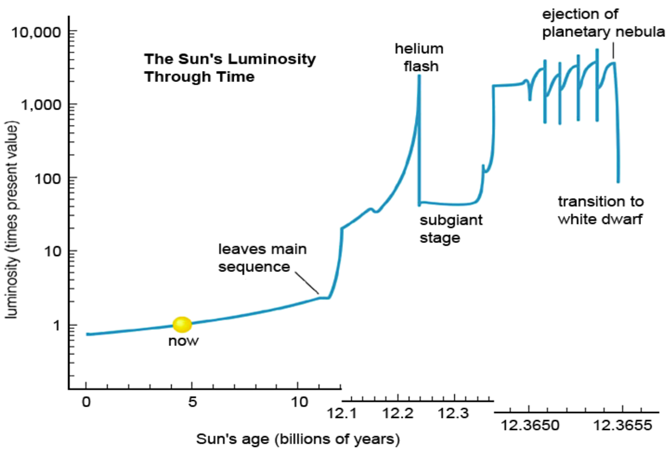
There are between 100 and 1000 billion stars in our Galaxy with less than 10 thousand visible by the naked eye from Earth. One estimates that there are ≈3 supernovae explosions per century in our Galaxy.

[17] No integers $x > 0, y > 0, z > 0$ and $n > 2$ satisfy $x^n + y^n = z^n$.

[18] Possibly, let it be rarely, a baby ape may also try to do it, but a *reasonable* human or non-human grown-up animal would have none of this nonsense.



Majority of stars do not turn to supernovae. For instance the life expectancy of the Sun like stars is about 10 *billion* years and their ends are *relatively* peaceful.



Ten times bigger stars shine 10^4 brighter and live only $\approx$10 *million* (=(10 $\cdot 10^{10})/10^4$) years, They end up exploding as supernovae. Also some stars in binary systems turn to supernovae by accreting matter from their companions.

During several weeks a supernova radiates with intensity of 1-100 billion Suns.

1.6 Mathematics and Ergo.

... the object of pure mathematics that of unfolding the laws of human intelligence.
JAMES JOSEPH SYLVESTER

We share our inborn ability to count with pigeons... but

what is Mathematics from the ergo perspective?

The triple-answer below is, by necessity, circular.

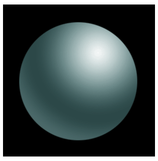
Mathematics, at least at its core, is "just" an instance of an ergo-structure.

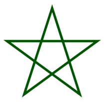
Mathematics is a product of human ergo-brain. It is "just" a fragment of the collective human ergo.

Mathematics is the tool and the language for a study of ergo-structres where the latter are "just" particular mathematical structures.

Let is say a few more words about these 1, 2, 3.

The shape of the heaven is of necessity spherical.
ARISTOTLE.





1. *The core mathematics* is all about *amazing structures* clustered around *symmetries*: perfect symmetries, hidden symmetries, supersymmetries, partial symmetries, broken symmetries, generalized symmetries, linearized symmetries, stochastic symmetries. Two thirds of this core along with most theoretical physics would collapse if the symmetry axes had been removed.

A fantastic vision, unimaginable to ancient mystics and to mediaeval occultists, emerges in the *Langlands correspondence* between arithmetic symmetries and the *Galois symmetries* of algebraic equations, where much of it is still in the clouds of conjectures.

It is amazing how mathematics manages to contain these symmetries and to encode, for example, "roundness" in the combinatorial, nearly digitalised, form of axioms, lemmas, theorems and proofs.[19]

This may seem not surprising, since the (collective) ergo brain that created mathematics – represents the external world in this manner. But it may be *an endogenous* property of mathematics.[20]

This "combinatorial" nature of mathematics may be also compared to that used by Life for encoding shapes of organisms by DNA sequances, except that there is no(?) mathematical counterpart to *transfer of information by 3D-folding*. (A primitive form of "embryonal development" may be discerned in organisation of some mathematical proofs.)

A mathematician is an ergo-brain's way of talking to itself.
NIELS BOHR [misquoted]

[19] We also poorly understand how and why this happens *within* mathematic, where *transcendental methods* often resolve *purely algebraic* problems; the prominent instance of this is *the Hodge theory* in algebraic geometry.

[20] The proposition – *mathematics exists as an independently entity* – may be understood only metaphorically. No conceivable experiment or argument would make it more or less feasible. But... you can not do mathematics if you do not believe into it. And this also the way how physicists take *reality of the physical word*.

2. Mathematics is the last born child of the ergo-brain, its development is guided by our ergos, we search for *interesting and beautiful* , while *trivial* bores us to death.

Mathematics *shines in the mind of God*, as Kepler says, but we are no gods and our minds are not pure ergo, our thinking is permeated by ego that makes hard for us to tell "true and interesting" from "important" and that makes the (ergo)right choices difficult.

In the eyes of the egomind, much of mathematics appears *abstract and difficult* while what you see in front of you eyes is *simple and concrete*.

But this simplicity is deceptive: what your eyes "see" is *not* simple – it is an outcome of an elaborate image building by your visual ergosystem that is, probably, more abstract and difficult than most of our mathematics.

Compelled by our ergo, we, mathematicians, search for another kind of "simplicity" that is *beautiful and interesting* – not at all trivial; *trivial* bores us to death.

We are trilled when ego-mind's "simple and apparent" is explained in terms of "abstract and difficult", that may not, a priori, even exist. Inventions of *negative* and later of *complex "imaginary"* numbers are witnesses to how it goes.

Our mathematical diamonds have been polished and their edges sharpened – century after century, by scratching away layers of ego from their facets, especially for the last fifty years. Some of what came out of it may appear as "abstract nonsense" but, as Alexander Grothendieck says,

The introduction of the cipher 0 or the group concept was general nonsense too, and mathematics was more or less stagnating for thousands of years because nobody was around to take such childish steps.

You can not apply mathematics as long as words still becloud reality.
HERMANN WEYL

3. In building a mathematical frame for "ergo" and we need to recognize what of our mathematics is ready to serve as "parts" of ergosystems, what should be rejected and what needs to be be made anew

"Ergo-criteria" for these "ergo-parts" are exactly those we use everywhere in mathematics:

NATURALITY, UNIVERSALITY, LOGICAL PURITY CHILDISH SIMPLICITY.

Mathematical universality of learning programs in our ergobrains $\mathcal{EB}$ is seen in how we *enjoy and learn many different* logically complicated games. Thus, for instance a chess learning program in somebody's $\mathcal{EB}$ must be a *specialisation of a universal* learning program.

But why such programs should be simple? After all

The human brain is the most complicated object in the Universe. Isn't it?

The answer is that most general/universal theories are logically the simplest ones.[21] What is not simple is formulating/discovering such theories.

As mathematicians we are ready to accept that we are hundred times stupider than the evolution is but we do not take it for the reason that evolution is

[21]The simplicity of a universal idea, e.g. of *Gödel's incomleteness theorem*, may be obscured by plethora of technical details.

able to make miracles, such as a logically complicated brain at birth.

Believers into simplicity, we seek our own solution to the *universal learning problem* by adapting the *purest* kind of mathematics to the "dirty world" of *flows of signals* and their *transformations"* by out (still conjectural) ergo-brains.

1.7 Ergo in Play, in Humour, jn Art.

Without play and "playful thinking" we would not be human.

Children carry magic lanterns within themselves – the world projects onto the playground screens in their minds. And a similar *play-mode behaviour* of kittens and puppies is familiar to all of us.

Most young mammals play and also some birds e.g. crows and ravens. (It is not always clear what behavior can be classified as "play")

Playfulness retained into adulthood in humans and dogs, goes along with other neonatal characteristics. Some adults animals in the wild also paly, e.g. *dholes*.[22]

An Episode Taken from:

http://www.onbeing.org/program/play-spirit-and-character/feature/excerpt-animals-play/1070

[the dog] wagged his tail, grinned, and actually bowed to the bear, as if in invitation. The bear responded with enthusiastic body language and nonaggressive facial signals. These two normally antagonistic species were speaking the same language: "Let's play!" The romp was on. For several minutes dog and bear wrestled and cavorted.

There is no accepted adaptive evolutionary explanation for the play. Apparently, patterns of play programs reflect some facets in the mental architecture of the ergo-brain that came about *despite* not because of selection.[23]

[22]Dholes, also called *red Asian dogs* and *whistling dogs* are agile and intelligent animals, somewhat one-sidedly depicted by Kipling in *The Jungle Book*.

The systematic killing of dholes was conducted by locals and promoted by British sport hunters during the British Raj. Later, some European "naturalists" called for extermination of dholes, because dholes had no "redeeming feature" but rather hair between their toes. Despite recent measures protecting dholes their population (<2000) keeps declining.

[23]Eventually, selection may win out and populate Earth exclusively with bacteria which would have no risk-prone inclination for play. This would be the most stable/probable state of the biosphere of an Earth-like planet, granted an "ensemble" of $10^{10^{10}}$ such planets.

Ego and Ergo in Play. The drive *to win* originates in ego, but "winning/loosing" is, structurally speaking, a trivial component of play.

A pure ergo-system would not try to win but rather adjust to a weaker player to make the play/game *maximally interesting*.[24]

One's ego-mind approaches the problem of play (as much as everything else) with "why"-questions; purpose-oriented solutions are welcomed by ego and such "explanations" as *Oedipus complex* for chess are acceptable.

We – students of ergo, on the other hand, admit that we do not understand the deep nature of play, but we reject the very idea of any common sense (teleological) explanation.

For instance, contemplating on the "meaning" of a chess-like game, we do not care what drives one to win, but rather think of the architecture of an elaborate network of *interesting game positions*. Algorithms for representation of such networks lie at the core of the *universal learning.*

Sense of humour, laughing at "funny", is closely associated with play – this is apparent in children. This "sense" is an instance of what we call an *ergo-mood* – a reaction of the ergo-brain to "*funny arrangements of ideas"*.

Making a universal style program recognising such "funny arrangements",say on the internet pages, seem easier than recognising *interesting arrangements of pieces* on the chessboard

Performing as well as fine Arts – theater, dance, painting, sculpture, music, poetry, grew out of child's play on the ego-soil of the Human Mind, where *aesthetic perception* – feeling of beauty of nature and of artistic beauty, is shared within our minds with the *sense of the opposite-sex beauty.*

Music, poetry, the architecture of plants, animals and cathedrals, kaleidoscopic symmetry of peacock's tails – all that with no "reproduction" tag on them,[25] trigger in us a feeling similar to that is caused by attraction to opposite sex.

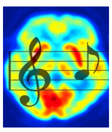
But this may only distract you from what we want to understand. For instance,

what are *universal structures* in Arts that are separate of "ego"?

Well,... the formal study of arts, especially of music, goes back to Pythagorus.

Also, there are active fileds of *Neuroscience of Art, Neuroesthetics, Cognitive neuroscience of music*, with many publications openly accessible on the web. As an instance, let us point to http://www.ncbi.nlm.nih.gov/pubmed/21217764, where a group of neuroscientists present their finding of

endogenous dopamine release in the striatum at peak emotional arousal during music listening" obtained with *"positron emission tomography scanning, combined with psychophysiological measures of autonomic nervous system activity"*.



[24] This may be not very *interesting* to the second player, e.g. in the cat and mouse game.

[25] Peacock's tails are sexually significant for peahens.

And an avalanche of superlatives that music lovers pour on you when they speak of music[26] tells you something about the levels of endorphins release into their blood, but does not help answering the following kind of questions.

What is the starting level of complexity an ergo-system must have, such that, upon unsupervised learning, it will achieve the ability to "correctly" assign aesthetic values to pieces of art?

Probably, this level need not be prohibitively high, if such a "value" is represented *not by a number* $V = V(A)$ assigned to a piece of art A but by a (partial?) order relation

$$V(A_1) >_c V(A_2), c \in C,$$

that depends on c taken from a set C of groups of art critics c.[27]

1.8 Ergo in Science.

The mental set-up that makes the very existence of science possible is of ergo. Henri Poincaré articulates this as follows.

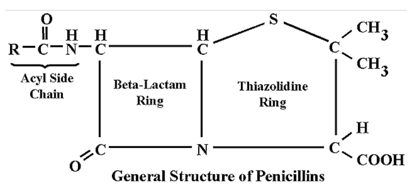
The scientist does not study nature because it is useful to do so. He studies it ... because it is beautiful. [It is] *intimate beauty which comes from the harmonious order of its parts, and which a pure intelligence can grasp.*

But – one objects – Poincaré was a high priest of pure thought. Would experimentalists agree?

The experimental scientist who single handedly contributed most to our electricity-hungry industrial civilisation was Mikhail Faraday.[28] He writes:

It is the great beauty of our science, chemistry..., that ... opens the doors to further and more abundant knowledge, overflowing with beauty and utility.

Yet, medical researchers – doctors and inventors of drugs – were not playing "scientific curiosity games" but were driven by the concern for the wellbeing of their fellow humans. Weren't they?



General Structure of Penicillins

[26]This is also how mathematicians speak of their beloved science.

[27]Owners of art galleries routinely solve the problem of assigning consumer depending prices $P_c(A)$ to pieces A of modern art.

[28]Without him on the scene, the world history would be shifted by a few years backward and would be, of course, quite different from what we know as our world today.

Let us listen to what Alexander Fleming, who discovered penicillin and Howard Florey who brought penicillin to the therapeutic use,

Fleming: *I play with microbes. There are, of course, many rules in this play....but...it is very pleasant to break the rules... .*

Florey: *This was an interesting scientific exercise, and because it was of some use in medicine is very gratifying, but this was not the reason that we started working on it*

Penicillin had saved about 100 million human lives.[29] Without either Fleming or Florey half of us would not be alive today and younger ones would not be even born.

Brief History of Penicillin

[30]

Albeit the history of penicillin[31] was not only about Fleming and Florey,[32] [33] say.

who suggested transferring the active ingredient of penicillin back into water by changing its acidity, this would purify the penicillin.

But Heatley had just the right abilities. He knew how to improvise. He knew how to use his hands. He wasn't just a scientist. He had the inventive sense of an engineer. First he created means for growing penicillin and measuring its output. Then he figured out how to reuse the special fungus it grows in.

But that wasn't enough. We have to separate penicillin from the fungus. That'd stymied everyone for 10 years. Heatley invented a new extraction process to get the penicillin without destroying it. Finally, he used hospital bedpans as mass-production growth chambers.

The Oxford team announced penicillin as an antibiotic in 1940. Their paper had six authors, listed in alphabetical order. Heatley was the fourth.

This was the eve of WW-II. America picked up Heatley's techniques. We mass-produced penicillin during the war and saved countless lives with it. Finally, in 1945, Fleming and Florey got the Nobel Prize in medicine for their part in the work.

---

[29] By comparison, the number of victims of the 20th century "fighters for people happiness" is estimated 180 – 220 million.

[30] Fleming determined that the secretion from the mold *Penicillium notatum* was effective against many bacteria, but its antibacterial activity was short-lived (1928).

Penicillin, the substance yet unknown to Fleming, kills so-called *gram-positive* bacteria by blocking cell wall growth in the course of bacterial replication. (Gram-positive bacteria, e.g. *streptococci* and *staphylococci*, have no outer cell protective membrane.)

[31] Healing properties of mould have been known for 45 centuries. Ernest Duchesne (1874 – 1912) was, apparently, the first who conducted a scientific study of antibacterial properties of mould. His results on *Penicillium glaucum*, similar to what was observed by Fleming, were recorded in his 1897 thesis: *Contribution à l'étude de la concurrence vitale chez les micro-organismes: antagonisme entre les moisissures et les microbes*. The Institut Pasteur did not even acknowledge receipt of his dissertation.

[32] We suggest reading *Howard Florey, The Making of a Great Scientist* and *Alexander Fleming, The Man and the Myth* written by haematologist Gwyn Macfarlane.

[33] In 1939, Florey created and directed a team of scientists for the study of anti-bacterial substances that are produced by mould. Using the sample of Penicillium notatum preserved by Fleming, they extracted and purified *Penicillin* – the active antibacterial agent in the mould and produced therapeutically significant amounts of it (1940). The key roles in this achievements was played by biochemists: Ernst Boris Chain and Norman Heatley) Edward Penley Abraham

Now, 50 years later, Oxford honors the quiet man who made all that possible. And Heatley says, diffidently:

This is an enormous privilege since I am n

In 1938, Chain stumbled across Alexander Fleming's 1929 paper on penicillin in the British Journal of Experimental Pathology, which he brought to the attention of his colleague Florey.7

Heatley devised the solvent-to-water transfer cycle that permitted an impure but stable penicillin to be prepared from the mould culture fluid. He also devised the cylinder plate diffusion technique that provided a much easier, reliable and sensitive assay for penicillin and was later adopted as the standard assay for antibiotic activit

who monitored the first experiment in which the protective effect of penicillin was assayed in mice infected with streptococci.

In order to strengthen the chemical side of the work, Florey attracted Edward Abraham to the Dunn School. Abraham, who had recently completed his doctorate in the Department of Organic Chemistry at Oxford (the Dyson Perrins Laboratory) set about the difficult task of purifying penicillin and determining its structure. He was eventually completely successful in both these aims and was the first to propose the correct chemical structure for penicillin. Abraham?s structure, which involved the novel ? -lactam ring, was at once accepted by Chain, but not by Robert Robinson, Head of the Dyson Perrins Laboratory, or by J.W. Cornforth, another distinguished Australian who was then also working in the Dyson Perrins; they suggested an alternative structure. The matter was settled by Dorothy Hodgkin (Crowfoot at the time), who examined crystals provided by Abraham and confirmed by crystallographic methods the presence of the ? -lactam ring.

For example, the "obvious" properties of light and matter we see everywhere around us make sense *only* in the context of *quantum theory of electromagnetic fields*, with the energy source of the sunlight being inconceivable without the theory of *strong interactions* in *atomic nuclei*.

Something as simple as the air we breath is the product of unbelievably complicated quantum-chemical process of *photosynthesis* and the whole edifice of Life on Earth is based on statistical mechanics of large *heteropolymeric molecules* .

The mental set-up that makes the very existence of science possible is of ergo. Here is how Poincaré puts it:

The scientist does not study nature because it is useful to do so.
He studies it ... because it is beautiful.

... intimate beauty which comes from the harmonious
order of its parts, and which a pure intelligence can grasp.

But grasping, embracing this beauty in your mind may be prohibitively difficult. Nothing in Nature that is worth understanding comes in "a few simple words". If you happen to learn something novel for you in science without much intellectual effort and hard work – this is either not especially novel or it is not science.

Even most familiar and apparently simple things in science are intuitively hard to accept, such as the second law of Newton that presents a manifestly mathematical (ergo)way of thinking about motion:

Lex II: Mutationem motus proportionalem esse vi motrici impressae, et fieri secundum lineam rectam qua vis illa imprimitur.

This law, even more so than the first law, runs against how our visual and somatosensory (mainly *proprioception*– the body sense) systems represent properties of motion in our mind.[34] Yet, some people find themselves comfortable with Newtonian laws; analyzing how they learn to understand them may shade some light on general mechanisms of ergo learning.[35]

Science is a child of the art of *not* understanding.

There is no visible *non-trivial* mathematical structure in what we conciously perceive in our (ego)mind and it is unlikely that there is a *realistically describable* structure (mathematical model) of human (neuro)brain capable to account for such mental processes as learning a language, for example. But we conjecture that such structure(s) does reside in the *ergobrain*.

Mastering accurate throwing, a uniquely human[36] capacity, could have been, conceivably, a key factor in the early hominid brain evolution.[37] According to *the unitary hypothesis*, the same neural circuitry may be responsible for other sequential motor activities, including those involved into the speech production and language. [1], [12].

The unitary hypothesis, albeit quite amusing, is hardly(?) relevant to our "ergo" but drawing the following parallel between mechanics of throwing and ergo-learning may be instructive.

From a thrower point of view the most important is his/her *aim*, that must be achieved with a correct *initial condition* – the velocity vector of a stone – that then will follow the trajectory toward a desired target. You may (and you better do) fully forget the laws of Newtonian mechanics for this purpose.

But from a physicist's point of view, it is the *second law + the force field* (graviton and the air resistance) that determine the motion – the initial condition is a secondary matter.

A mathematician goes a step further away from the ancient hunter and emphasizes the general idea of time dependent processes being described/modeled by *differential equations.*

We – physicists and mathematicians with all our science would not stand a chance against *Homo heidelbergensis*[38] in a stone or spear throwing contest;

[34]The essential logic of this reconstruction is of ergo but it serves the survival of our ego and serves it well, better than mathematical Newtonian model would do.

[35]Majority of us, even if we can correctly recite the three laws of motion, do not believe in these laws. We intuitively reject them in view of the apparent inconsistency of these laws with much of what we see with our own eyes, such as the motion of a pendulum that visibly contradicts to the *conservation of momentum law*.

[36]Elephants may be better than humans at precision throwing.

[37]500 000-year-old hafted stone projectile points, 4-9cm long, were found in the deposits at Kathu Pan in South Africa, http://www.newscientist.com/article/dn22508-first-stonetipped-spear-thrown-earlier-than-thought.html

[38]*Homo heidelbergensis*, a probable ancestor of Homo sapiens as well as of Neanderthals and of Denisovans, lived in Africa, Europe and western Asia between 1 000 000 and 200 000

however, we, at least some of us, shall do better in mathematically designing *gravity-assist trajectories* from Earth to other Solar system bodies.

Homo heidelbergensis would lough at the fools engaged in the useless activity of aiming at inedible targets and Lev Tolstoy who speaks not for Homo heidelbergensis but for *"a plain and reasonable man "* advocates a similar philosophy of life.

> *It* [science] *triumphantly tells him*[/her]*: how many million miles it is from the earth to the sun; at what rate light travels through space; how many million vibrations of ether per second are caused by light, and how many vibrations of air by sound; it tells of the chemical components of the Milky Way, of a new element – Helium – of micro-organisms and their excrements, of the points on the hand at which electricity collects, of X-rays, and similar things.*
> *"But I don't want any of those things," says a plain and reasonable man –"I want to know how to live".*

Well, let us make it clear, *goal free learning* is far from being "plain and reasonable" but it rather follows a mathematical physicist in his view on mechanical motion: there is nothing special, nothing *intrinsically* interesting *neither in the hunter's aim* A, no matter how hungry he/she is, *nor in the initial condition* I, although much skill is needed to achieve it. But the transformation $T = T_L$ from I to A, that incorporates the laws of motion expressed by differential equations, is regarded as something *universal* and the most essential from our point of view.

There are many possible aims A and initial conditions I but not so many fundamental laws L and of transformations $T = T_L : I \mapsto A$ associated to them.[39] This what makes these laws so precious in our eyes.

Similarly, one may think of learning as of a transformation of an initial input and/or of a learning instruction I to the final aim A of learning.

Here we are even in a poorer position than the ancient hunter: we have hardly an inkling of what the corresponding "transformation by learning" T_L does as it brings you from the initial input/instruction I to our your aim A:

What is the "space" where all this happens?

What is the structure of the trajectory that leads from I to A?

And, unlike a teaching instructor, we are not concerned with *observable* I and A, but with mathematical models of *invisible* intrinsic structures of transformations T that are built according to "universal laws of learning".

It is not that we deny importance of goals, instructions and external stimuli for learning, but we relegate them to the secondary roles in the formula $T_L : I \mapsto A$. We try to understand learning processes regardless of their specific aims, or, rather, we want to see general aim generating mechanisms within the "universal laws" of learning.

??????????? ON BEING TRIVIAL.

Triviality is a mathematician's scarecrow but no-trivial constructions are often made with a few trivial constituents.

---

years ago.

[39]This stands in a sharp contradiction with *Cantor's theorem*: there are more *logically conceivable* functions $f : x \mapsto y = f(x)$ than arguments x. But logic should not be taken literally when it comes to the "real life mathematics".

For instance, certain structures *are* "trivial" when taken in isolation, such as highly disconnected (often bipartite) graphs that represent the [object]–[name] relations where the edges join words with the corresponding visual images or [question]–[answer] graphs of human dialogs – the brains of the stupidest animals depend on such graphs. Yet, mathematical derivations issuing from *several* trivial graphs make *non-trivial* structures in the human ergobrain.

What is abstract and what is obvious?

Explaining "simple and apparent things" by means of something "abstract and difficult", that may not a priori even exist, is against common sense, but this is how it is in science and in mathematics.

For example, the "obvious" properties of light and matter we see everywhere around us make sense *only* in the context of *quantum theory of electromagnetic fields*, with the energy source of the sunlight being inconceivable without the theory of *strong interactions* in *atomic nuclei*.

Something as simple as the air we breath is the product of unbelievably complicated quantum-chemical process of *photosynthesis* and the whole edifice of Life on Earth is based on statistical mechanics of large *heteropolymeric molecules* .

???????????

1.9 Ergo in Language and Language in the Brain.

it is unlikely that there is a *realistically describable* structure (mathematical model) of human (neuro)brain capable to account for such mental processes as learning a language, for example. But we conjecture that such structure(s) does reside in the *ergobrain*.

Mastering accurate throwing, a uniquely human[40] capacity, could have been, conceivably, a key factor in the early hominid brain evolution.[41] According to *the unitary hypothesis*, the same neural circuitry may be responsible for other sequential motor activities, including those involved into the speech production and language. [1], [12].

Parrots can be taught to speak without much stimuli: if a mirror is placed between the parrot and the trainer, the parrot, seeing his own reflection in the mirror, fancies another parrot is speaking, and imitates all that is said by the trainer behind the mirror. The teaching goes well if your bird is happy, relaxed and... young.

All neurologically healthy (and even partially impaired) children spontaneously learn languages without any reward/punishment reinforcement [**?**], which appears as miraculous and inexplicable as strong inclinations of some children to play chess, perform/compose music, to study science and to do mathematics.

[40] Elephants may be better than humans at precision throwing.

[41] 500 000-year-old hafted stone projectile points, 4-9cm long, were found in the deposits at Kathu Pan in South Africa, http://www.newscientist.com/article/dn22508-first-stonetipped-spear-thrown-earlier-than-thought.html

A child language acquisition works fine *without* supervision. In fact, a pressure by a teacher inhibits language development in children according to studies by W. Labov and by S. Phillips, see p. 299 in [?]. (The unhappy supposition is that this applies to learning mathematics by children as well.)

Due to the complexity of the problem and the lack of study it remains unclear which kind of teaching practices promote and which suppress the learning process by children, but it is painful for a teacher to accept that his/her intervention may be harmful rather than beneficial.

(This is similar to the dilemma having been faced by medical doctors for centuries. For example, it was established by Almroth Edward Wright, Alexander Fleming and their coworkers during the First World War that antiseptics applied to fresh wounds were more likely to kill phagocytes than bacteria, such as ordinary streptococci and staphylococci as well as deadlier anaerobic tetanus bacilli and Clostridium perfringens causing gas gangrene [?] ch. 9. But it took a couple of decades before the wound treatment suggested by Wright became a standard surgical procedure.)

Noam Chomsky, Eric Lenneberg and their followers believe (see ch.12 in [?]) that children have an evolutionary installed "innate universal grammar" in their heads that facilitates and constrains language learning. (Apparently, [?] Chomsky himself does not adhere to this idea anymore. See [?] for a roboticist's perspective on language generating mechanisms and other emergent cognitive phenomena.) In fact, learning a "general language" even with a simple grammar may be virtually impossible [?].

On the other hand, besides languages, children easily learns chess and bonobos display Pac-Man playing ability. Even more inexplicably, young students can learn mathematical arguments which, if logically expounded in full detail, would cover tens (hundreds?) thousand pages, e.g. the proof of Fermat's Last Theorem [?].

(The results of recent neurological studies of how we learn to *read* [?] can be interpreted in favour as well as against Chomsky-Lenneberg thesis.)

We shall explain later on how the universal structure learning mechanism accounts for the language acquisition along with chess (regarded as a dialog between the players) and mathematics (where ergobrain plays with itself [?]) with agreement with the point of view currently accepted by many psychologists and computer scientists [?].

Besides the ordinary spoken languages there are several types of less common ones.

Whistled languages (Wikipedia) are systems of communication which allow fluent whistlers to transmit and comprehend messages over several km distances, with the whistling being either tone or articulation based.

Whistled languages function by varying the frequency of a simple wave-form and convey phonemic information through tone, length, and, to a lesser extent, stress, where many phonemic distinctions of the spoken language are lost [**?**].

The *Silbo* on the island of La Gomera in the Canary Islands, based on Spanish, is one of the best-studied whistled languages. The number of distinctive sounds or phonemes in this languages is a matter of disagreement, varying according to the researcher from two to five vowels and four to nine consonants. This variation may reflect differences in speakers' abilities as well as in the methods used to elicit contrasts.

The language of 200-300 hunter-gatherer *Pirahã* people, living in a few villages on the banks of the river Maici deep in the Amazon rainforest in Brazil, can be whistled, hummed, or encoded in music with the meaning conveyed through variations in pitch, stress, and rhythm.

The Pirahã language, unrelated to any other living language, has been studied by Keren and Dan Everett who lived with Pirahã people from 1978 to 1983 and from 1999 to 2002. Many unusual, often controversial, features of Pirahã [**?**], identified/claimed by Everetts, have brought this language to the focus of an exceptional interest among linguists and anthropologists.

Whistled languages tend to disappear in the face of advancing telecommunications systems. For example, in the Greek village of Antia, only few whistlers remain (2005) but in 1982 the entire population knew how to whistle their speech.

Sign languages, which commonly develop in deaf communities, have a high *non-sequential* component: many "phonemes" are produced simultaneously via visually transmitted patterns combining shapes, orientations and movements of the hands, arms and body as well as facial expressions.

Before the 1970s, deaf people in Nicaragua were largely isolated from each other. In 1980, a school for adolescent deaf children was opened and by 1983 there were over 400 students enrolled. Initially, the language program emphasized spoken Spanish and lip-reading; the use of signs by teachers was limited to fingerspelling. The program achieved little success, with most students failing to grasp the concept of Spanish words (Wikipedia).

But while the children were *linguistically disconnected* from their teachers they were communicating by combining gestures and elements of their home-sign systems, a pidgin-like form, and then a creole-like language rapidly emerged.

Then the young children had taken the pidgin-like form of the older children to a higher level of complexity, with verb agreement and other conventions of grammar. (Some linguists questions assertions that the language has emerged entirely without outside influence.)

The communication problem seems insurmountable for deaf-blind people; they succeed, nevertheless. Different systems are in use, one is *Tadoma* which is tactile lip-reading (or tactile speechreading). The Tadoma user, feels the vibrations of the throat and face and jaw positions of the speaker as he/she speaks. Unfortunately, this requires years of training and practice, and can be slow, although highly-skilled Tadoma users can comprehend speech at near listening rates, Below are two pieces of poetry written by deaf-blind people.

From "A Chant of Darkness" by Helen Keller
http://www.deafblind.com/hkchant.html

In the realms of wonderment where I dwell
I explore life with my hands;
I recognize, and am happy;
My fingers are ever athirst for the earth,
And drink up its wonders with delight,
Draw out earth's dear delights;
My feet are charged with murmur,
The throb, of all things that grow.

This is touch, this quivering,
This flame, this ether,
This glad rush of blood,
This daylight in my heart,
This glow of sympathy in my palms!
Thou blind, loving, all-prying touch,
Thou openest the book of life to me.

****

“My Hands“ By Amanda Stine.

My hands are . . .
My Ears, My Eyes, My Voice . . .
My Heart.
They express my desires, my needs
They are the light
that guides me through the darkness

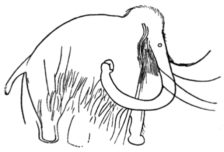
They are free now
No longer bound
to a hearing-sighted world
They are free
They gently guide me

With my hands I sing
Sing loud enough for the deaf to hear
Sing bright enough for the blind to see

They are my freedom
from a dark silent world
They are my window to life
Through them I can truly see and hear

I can experience the sun
against the blue sky
The joy of music and laughter
The softness of a gentle rain
The roughness of a dog's tongue

They are my key to the world
My Ears, My Eyes, My voice



My Heart

They are me

http://www.deafblind.
com/myhands.html

The limits of my language means the limits of my world.
LUDWIG WITTGENSTEIN.

• Da-Da, Ma-Ma, Pa-Pa, Ba-Ba.

• Neanderthals are mobbing a mammoth; their shouts fly throw the air and reappear as cuneiform writings on clay.

• The "run and yell" program in your brain switches to the "sit and read" mode.

What is LANGUAGE? Is it *conversing, writing, reading?*

A mammoth hunter scratches his head and pronounces after a minute of concentration:

"Semiosis that relates *signs* with *things* I can eat."[42]

This translates in our terms to

a *bipartite graph*[43] Σ on two vertex sets[44], call them Th and Si – the sets of "things" th and of "signs" si.

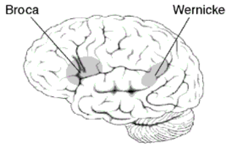
(The edges of Σ correspond to the pairs (th, si), where th and si are related by semiosis).

The hunter interrupts and raises his ax:

[42]However cruel, employment of semiosis by these people in hunting was indispensable for filling their bellies with meat and their life with meaning.

[43]This is not the same as an *existential graph* of Peirce.

[44]On may argue whether "signs" constitute a *set* in the proper meaning of the word and this is even more dubious for "things" and "meanings".



"This mathematics, it is not REAL LANGUAGE."

Since the hunter is safely far away, I may state openly what I think:

If we want to say something general and structurally interesting about LANGUAGE, *we ought to speak* MATHEMATICS *– the only language available to Homo sapience for this purpose.*

(Accidentally, there is a semiotic turn to it:

Mathematics is the art of giving the same name to different things.
HENRI POINCARÉ.

Of course, this a metaphor, it is not meant to be taken literally, nor is it intended as a definition. This sentence brings to one's mind the image of an exquisitely elaborate graph depicting the artful arrangements of patterns of the body of mathematical ideas and where one sees bushes of edges sprouting from buds of *universal* "abstract" concepts toward many seemingly different specific "concrete things".)

But does it make any difference if you say "graph" rather than "semiosis"?

Different names you give to things channel your stream of thought in different directions.[45]

"Semiosis" brings to you mind the images of "meaningful things" of "real world" and "meaningless" scribbles on paper – the signs associated to these objects .

When you say "graph", you focus on the *structure of association* and close your eyes to all what you naively[46] perceived as being real and meaningful.[47]

However, whatever you say, the bare semiosis model is more applicable to vervet monkey alarm calls rather than to human languages.

Languages have intricate internal structures, and formation of a language in a developing human brain plays its role in *compression and structuralization* of flows of information carried by the senses under the stimuli coming from "real things".[48]

Signs, objects, meanings, etc. are all about appearances, not about structures and/or structuralizing processes.

And *semiosis* per se for LANGUAGE is like *applications* for MATHEMATICS: it carries only a secondary and rather shallow structure. Yet, thinking in terms of graphs is instructive.

For instance, one may notice that "signs" and "things/meanings" appear on an equal footing in the semiosis model and ask:

Can one tell *who is who*, i.e. which vertices represent *signs* and which *things*, by the combinatorics of the graph Σ?

Also one observes by looking at graphs that a semiosis can support an interesting structure only if a typical sign $si \in Si$ has *multiple meanings* – the same name for different things, and, moreover, if most signs come in significantly numerous groups of *synonyms*.

[45]And they appeal to different groups of people: you hardly find (non-existential) "graph" in a semiotics text or "semiosis" in an article on graph theory.

[46]A mammoth hunter will have another idea of who is being naive.

[47]This is how a Homo sapience child approaches the kindergarten Ramsey, see 2.1.

[48]Superficially, this is like mathematics: there have been thousands of *different* stones that have crossed your field of vision (have been stored in you visual memory?) but the *same* word *stone* stands for all of them in your brain with probably only a few neurons occupied for this purpose.

(The multiplicity should not be excessive: even a monkey would figure our that there is something wrong if *all* things *th* are connected with *all si* by semiotic edges of Σ. A paradigm of a logically perfect language in the eyes of a vervet monkey would be a *one-to-one* correspondence between "signs" and "things", that is if each signs has a unique well defined meaning, and also every meaning is encoded by a single sign. But such a language will be no good for hunting animals as intelligent as mammoths.)

Also, our graph suggests how one may measure distances between different "things/meanings" in terms of signs associated to them.

Namely, given a thing $th \in Th$, one assigns to it the subset $S_{th} \subset Si$ of all signs *si* associated to *th* by an edge in Σ and define the *Hamming distance* between th_1 and th_2 in terms of the cardinalities of the corresponding sets and their intersections as follows.

$$dist_{Ham}(th_1, th_2) = card(S_{th_1}) + card(S_{th_2}) - card(S_{th_1} \cap S_{th_2}),$$

and similarly, one defines a distance on Si by representing all *si* by subsets $T_{si} \subset Th$.

Now one can approach the above "who is who" question by comparing the geometries of Th and Si with respect to these distances.

Are, for instance, the shapes of spaces $(Si, dist_{Ham})$ of signs tend to be round or oblong?

Do the spaces $(Th, dist_{Ham})$ of things look smooth or hairy?

Taking the geometry of a space like $(Si, dist_{Ham})$ seriously may strike you as silly. But we shall see next that this acquires significance for a class of graphs that are *intrinsically* associated to LANGUAGE with *no reference* to "things".

1.10 Ergo and Ego Worlds.

When one day, while Matata was away, Kanzi began competently using the lexigrams, becoming not only the first observed ape to have learned aspects of language naturalistically rather than through direct training. Within a short time, Kanzi had mastered the ten words that researchers had been struggling to teach his adoptive mother, and he has since learned more than two hundred more. When he hears a spoken word (through headphones, to filter out nonverbal clues), he points to the correct lexigram." (Adult Kanzi can flake Oldowan style cutting knives. He understands how to beat the arcade game Pac-Man.)

The obvious (for some and paradoxical for others) advantage of Kanzi over his mother was that he *did not understand* the teaching instructions, he *had not developed* the taste for (internal and external) emotional rewards and *had not learned* how to obey.

It may seem there was no essential stimulus, no reason for Kanzi to learn, but *self propelled (or free) structure learning* is the game babies play.

In fact, "ego-stimuli" may inhibit learning rather than stimulate it as the following story, shows (http://www.intropsych.com).

Seeing through this wall is an essential intellectual difficulty in modeling the ergobrain and ergosystems in general. Our ergobrains are entrenched into services of our ego-minds and it is hard to separate the "real ergo-staff" from the ego-imprints. We attempt to identify and erase these imprints by *every* imaginable means; the pertinacious resistance of our ego-minds, both of the reader and of the writer of these lines, makes this task difficult.

For example, it is hard to mentally switch from the common ideas of "reality" and its "meaning" – two stereotyped "survival ego-programs" installed into you by the evolutionary selection process, to the "reality" of the ergoworld and to the structural "meaning" of combinatorics of "networks of ergo-ideas" in the ergosystems.

(Our emphasis on the "ergo-reality" should not be confused with the negation of "physical reality" by solipsists and, more recently, by cosmological biocentrists.

In fact, solipsism, after centuries of unsuccessful attempts, has been eventually brought to the firm ground by Terry Pratchett in his "Soul Music" of *Discworld*:

"... horses won't walk backward voluntarily, because what they can not see does not exist."

This confirms what solipsists were predicting all along, Pratchett's presupposition that the equidae, i.e. perissodactyl ungulate mammals, of Discworld possess no recognizable visual organs on their hindquarters notwithstanding.)

But from ergobrain's own perspective, the "real world" emerges via the reversed arrow symbolizing the transformation from the signals the ergobrain generates to their "echos" – the responses to these signals. The ergobrain learns by analyzing/structuralizing the "information" carried by these echos – "patches of ripples" on the external flows of signals resulting from ergobrain's "actions", where some "echos" originate within the ergobrain itself.

?Myself? is the central reference point in the visible part of our mind as much as the sun is central for accounting for the observed planetary orbits. But as the centrality of the sun is secondary to such concepts as ?differential equation? and ?Hamiltonian system?, the centrality of ego is secondary to the (unknown to us) ?laws? running general ergosystems.

1.11 Ergo Insulating Wall

Such a representation is possible despite the fact that, a priori, we know less about our internal ergo than about life on Earth half a billion years ago: by the time you are about two years old, when your *personality* – call it *egomind* takes its essentially final shape, you loose a direct access to your *ergoself*.

The legacy of evolution keeps this "ergo-power" contained by the ego-protective wall: a hunter-gatherer whose ergobrain had overrun his/her pragmatic egomind did not survive long enough to pass on his/her genes.

Pragmatically oriented egomind was favored by selection they stay on guard of our survival and passing on our (and of close kin ego can be altruistic) genes

while "ergo" left to itself — a *purpose blind* structure building machine, would be incompatible with survival in a hunter-gatherer society where our present day psyche has evolved.

This "blindness" is inevitably correlated with simplicity/universality of ergo-processes generated by the brain, where this simplicity is what had made these processes evolutionary accessible.

This structure may appear as intricate and as difficult to understand as emergence and evolution of *live structures* on our planet. But we *conjecture* that the working of our internal "ergo", call it *ergobrain,*

can be represented by a class of abstract ergosystems
and studied on the basis of several simple principles.

Such a representation is possible despite the fact that, a priori, we know less about our internal ergo than about life on Earth half a billion years ago: by the time you are about two years old, when your *personality* – call it *egomind* takes its essentially final shape, you loose a direct access to your *ergoself.*

This may seem paradoxical, since egomind, that is what you perceive as your "self", is a direct product of your internal *ergo* and the essential mental processes in your mind, especially those concerning *learning*, remain almost 100% ergo during all your life. Yet, your adult *ego* is insulated from ergo by an almost impenetrable wall.

The evolutionary reason is simple: pragmatically oriented egomind was favored by selection while "ergo" left to itself — a *purpose blind* structure building machine, would be incompatible with survival in a hunter-gatherer society where our present day psyche has evolved.

This "blindness" is inevitably correlated with simplicity/universality of ergo-processes generated by the brain, where this simplicity is what had made these processes evolutionary accessible.

The most dramatic evidence for the existence of an unbelievably powerful survival indifferent mental machinery in our heads comes from the rare cases where the ego insulating wall has *leakages.* (People with such "leakages" may shine like intellectual supernovas. They had no chance for "survival" in older times; in today's civilized societies they may live but their fire is stifled by educational institutions.[49])

It seems likely that similar mental blocks, call them *ego-protective walls* are ubiques in humans as well as in animals (ego)-minds, where, e.g. the reward/punishment stimuli inhibit structure learning by a student. At the same time, similar "walls" in a teacher's ego-mind, which was evolutionary "designed" for "listen to me", "obey me", "please me" and "conform to me" rather than for promoting free structure learning, can nullify the best efforts of the kindest of teachers blind to this.

In fact it is amazing that any teaching is possible at all: the structure analyzing ability of a one year old chimpanzee is, conceivably, by an order of magnitude greater than that of the conscious mind of the brightest homo sapience.

"Zoo-psychologist Sarah Boysen found that two of her chimpanzee subjects, Sarah and Sheba, were both capable of learning such concepts as "more than" and "less than" easily.

Boysen then used chimpanzee's favorite gumdrops as stimuli. The chimpanzee was presented with two plates of gumdrops. One had more on it, the other had less. While the other chimpanzee watched, the chimpanzee being tested was asked to point to one of the plates. Whichever plate it pointed to was given to the other chimpanzee.

In this situation, the chimp doing the pointing should have learned to point to the plate with fewer gumdrops on it, in order to have that plate given to the

[49]Srinivasa Ramanujan (1887 – 1920) is, probably, the brightest such supernova in mathematics who had a rare luck of becoming visible.

other chimpanzee and in order to get the plate with more gumdrops for itself. Instead, the chimpanzee insisted on pointing to the plate that had more gumdrops, even though these went to the other chimpanzee. Chimpanzees seemed to know they were making a mistake but they could not stop themselves.

Then Boyson replaced gumdrops with plastic poker chips. Now the chimpanzees had no trouble with the task. They pointed to the plate with fewer poker chips on it. This meant the plate with fewer gumdrops went to the other chimp, and the chimp that did the pointing got the larger number of gumdrops."

It seems likely that similar mental blocks, call them *ego-protective walls* are ubiques in humans as well as in animals (ego)-minds, where, e.g. the reward/punishment stimuli inhibit structure learning by a student. At the same time, similar "walls" in a teacher's ego-mind, which was evolutionary "designed" for "listen to me", "obey me", "please me" and "conform to me" rather than for promoting free structure learning, can nullify the best efforts of the kindest of teachers blind to this.

In fact it is amazing that any teaching is possible at all: the structure analyzing ability of a one year old chimpanzee is, conceivably, by an order of magnitude greater than that of the conscious mind of the brightest homo sapience.

Seeing through this wall is an essential intellectual difficulty in modeling the ergobrain and ergosystems in general. Our ergobrains are entrenched into services of our ego-minds and it is hard to separate the "real ergo-staff" from the ego-imprints. We attempt to identify and erase these imprints by *every* imaginable means; the pertinacious resistance of our ego-minds, both of the reader and of the writer of these lines, makes this task difficult.

For example, it is hard to mentally switch from the common ideas of "reality" and its "meaning" – two stereotyped "survival ego-programs" installed into you by the evolutionary selection process, to the "reality" of the ergoworld and to the structural "meaning" of combinatorics of "networks of ergo-ideas" in the ergosystems.

(Our emphasis on the "ergo-reality" should not be confused with the negation of "physical reality" by solipsists and, more recently, by cosmological biocentrists.

In fact, solipsism, after centuries of unsuccessful attempts, has been eventually brought to the firm ground by Terry Pratchett in his "Soul Music" of *Discworld*:

"... horses won't walk backward voluntarily, because what they can not see does not exist."

This confirms what solipsists were predicting all along, Pratchett's presupposition that the equidae, i.e. perissodactyl ungulate mammals, of Discworld possess no recognizable visual organs on their hindquarters notwithstanding.)

The difficulty in modeling the ergoworld is compounded by the lack of a general mental mechanism for inventing new kind of structures: our own ergobrain is severely limited in its structure creative ability. The main source of ideas for designing "ergo-architectures" comes from mathematics.

The ego-wall also makes ergo-processes hard to study experimentally. For example, if a psychologist or a neurophysiologist give a mathematical task (above the multiplication table level) to a subject, the dominant difficulty of the subject is breaking through his/her ego-ergo wall; all the observer will see is a "hole in the wall" – the trace of this breaking process.

And when the subject crosses the wall and enters the internal ergoworld, he/she becomes unable to (ego)communicate, as every scientist, artist or mathematician knows. (Composing music and filling-in tax forms do not go along together.)

Besides, the ergo-building of any significant structure, e.g. by a child learning a language, a mathematician digesting a non-trivial idea or a scientist designing a new experiment is a long (mostly unconscious) process, taking weeks, months, years.

This is comparable with the time schedule (weeks) of adaptive immune responses, where "learning" depends on molecular mutation/selection mechanisms. Gerald Edelman [**?**] proposed a similar principle for learning by the (neuro)brain which, according to Edelman, is based on Darwinian kind of competition/selecton between groups of neurons.

The structure encountered by an invaded organism is a collection of many copies of an essentially *random* object with little correlations between different classes of objects : the profiles of exposed surfaces of antigens. The learning strategy of the immune system – production/mutation/selection of antibodies is well adapted for "understanding" this structure, where "learning" one antigen does not much help to "understand" another one.

But the problems faced and solved by the ergobrain are structurally quite different from selecting antibodies for efficient binding to antigens. The basic function of the brain, consists in finding *correlations/similarities* in flows of signals and factoring away the *redundancies*; this seems hard to achieve by "blind selection" on the realistic time scale. On the other hand, even if this were possible, this would not help us much in unraveling the structures of most ergo-learning processes, such as the language acquisition, for instance.)

A common scenario in the mathematical community is as follows. X gives a lecture, where a (preferably young) Y is present in the audience but who understands nothing of the lecture; every word of the lecture is forgotten next day. (This is what normally happens when you attend a mathematics lecture with new ideas.) A year later, Y writes an article essentially reproducing the subject matter of the lecture with full conviction that he/she has arrived at the idea by himself/herself.

Unsurprisingly, X is unhappy. He/she believes that Y could not arrive at the idea(s) by himself/herself, since Y has no inkling of how the idea came up to him/her, while X is well aware when, why and how he/she started developing the idea. (A similar "structure recall" is common in solving non-mathematical problems, such as "egg riddle" in 3.3, for example.)

The structural patterns of the ergobrain, although being of evolutionary origin, can not be accounted for by the naked survival/selection mechanism, but rather by inevitable constrains on possible ergosystem's architectures; these are, essentially, mathematical constrains.

Conversing in a natural language is, metaphorically [**?**], just another kind of game where the "*real* meaning of words" is derived from the structural patterns of the language similarly to (although differently in some essential respects from) the "*real* meaning of moves" in a chess game, where none of the two "realities" (directly) connect to the "real world". A characteristic feature of a "pure-ergo" system (unlike the biological $ergo + ego$) is that it *does not try to win* a game, "winning" is an ego-notion: the system will adjust to a (possibly weaker) player and will try to make the game as interesting (to itself) as it can. (This may be not very "interesting" to the second player, as, e.g. in the cat and mouse game.)

This example leads us to another concept, that of an *interesting structure* and/or of an *ergo-idea*: this is something an ergobrain recognizes as such and then starts learning this structure (idea) by incorporating it in its own structure. (A learning by a biological organism, also involves an acquisition of ego-ideas, but these carry little structure of their own and have no significant effect on the structural modifications of the ergobrain.)

We (usually) identify: *structure = interesting structure*. To close the circle, we declare that

the goal free structure learning is a structurally interesting process.

Eventually, the very circularity of these "definitions" (similar to Peano's axiomatic "definition" of numbers) will bring us toward constructions of specific models of ergosystems, relying on and starting from our mathematical/scientific (ergo)perception of

interesting, amusing, surprising, amazing and beautiful structures,

Unsurprisingly, X is unhappy. He/she believes that Y could not arrive at the idea(s) by himself/herself, since Y has no inkling of how the idea came up to him/her, while X is well aware when, why and how he/she started developing the idea. (A similar ?structure recall? is common in solving non-mathematical problems, such as ?egg riddle? in 3.3, for example.) The structural patterns of the ergobrain, although being of evolutionary origin, can not be accounted for by the naked survival/selection mechanism, but rather by inevitable constrains on possible ergosystem?s architectures; these are, essentially, mathematical constrains. The structure of the ergobrain is somehow encoded by the neurobrain, but this ?code? is by far more complicated (partly, due to its redundancy) than, say, the genetic code. Even if one is granted a detailed description of the physiological brain, (which is unrealistic anyway) along with faithful correlations of neuronal processes with the behavior of an organism one would have no means of learning much about the ergobrain. The problem seems harder than reconstructing the 3-dimensional architecture of a folded protein molecule in space on the basis of a genome sequence, without any idea of the 3-space and/or of the interaction rules between molecules. However, if we assume that the evolution is ?big rather than structurally smart? we conclude that the the ?logic? underlying/generating ergobrain (and ergosystems, in general) can not be too complicated and can be guessed on the basis of ergo-moods (surprised, amused, bored, etc.) that ergobrain induces in our conscious mind. These moods, being independent of the pragmatic content of the signals re- ceived by the ergobrain, serve as universal signatures/observable of ergo-states. They reflect one?s ?emotional? reactions to structural peculiarities in the flow of incoming signals, e.g. feelings of being amused, surprised, doubtful, etc. 36 Although these moods/signatures are ?visible? only at the ?intake? of ex- ternal signals by the ergobrain, we conjecture that similar signatures mark and guide the internal ergoprocesses as well. Besides signatures of ergo-moods, the functioning of the ergobrain is seen in:

2 Mathematics and its Limits.

2.1 What Kind of Mathematics?

The signals entering the ergobrain via vision, hearing and olfaction are "written" on certain physical/chemical backgrounds the structures and symmetries of which have been much studied by mathematicians.

1. The visual signals are carried by the four dimensional *space+time* continuum. Signals break the symmetry (and continuity) of the *space+time* but eventually, the ergobrain reconstructs this symmetry.

2. The auditory signals are carried by the two dimensional *time*+$frequency$ space. Ergobrain, unlike mathematicians, does not seem to care about the underlying (symplectic) symmetry of this space; it is concerned with the "information content" of these signals and with *correlations* and/or *redundancies* in flows of the signals.

3. The "letters of smells" are aromatic molecules entering our noses. The background space here has by far more symmetries than the the above two spaces. (Its symmetry group is something like the automorphism group of a *Lebesgue-Rokhlin measure space.*) This is, probably, the reason of why the olfactory perception depends on so many different receptors and why there is no human olfactory language.

4. There is no apparent uniform (symmetric) spacial background for somatosensory and touch (haptic) perceptions but their information carrier potential is comparable to that of vision and hearing.

5. Linguistic information entering the ergobrain does not much depend on the physical carrier of this information. This suggest a universal class of structures encoding this information; our main objective is describing these structures, which we call *syntactic ergo-structures.*

Such a structure is a *combinatorial* object X, a kind of a *network* made of finitely many, 10^7-10^9, "atomic units – ergo-ideas". This network structure generalizes/refines that of a *graph*, where some patterns are similar to those found in the mathematical theory of n*-categories.*

The combinatorial structure is intertwined with a geometric one; both an individual network X and the totality of these, say $\mathcal{X}$, carry natural distance-like geometries which are derived from combinatorics and, at the same time, are used for constructing combinatorial relations. This is essential for the *learning process* which is modeled by some transformation(s) $\mathcal{L}$ in the space $\mathcal{X}$; the resulting "educated learner" appears as an *attractive fixed point* X (or a class of points) under this transformation(s).

The (ergo)learning process (transformation) $\mathcal{L}$ starts from a space of signals and eventually compresses (folds) the *information* they carry by some *coclustering algorithms* into our X. When $\mathcal{L}$ applies to visual signals it recaptures the underlying (non-Lorentzian) symmetry of *space+time*; in all cases, ergo-learning creates certain *syntactic symmetries* – rudiments of what one finds in the mathematical *theory of groups.* (Our syntactic structures are more "elementary" and more "abstract" than those studied by mathematical linguists, where "elementary", "abstract", "fundamental", "rudimentary" are synonymous concepts from our (ergo) point of view.)

Building and identifying symmetries *within itself* serves as an essential guideline for an ergo-learner. These are created, seen from outside, by a *statistical* *analysis of signals*, but the ergo-learner itself can not count (unless being taught) and it knows nothing of probability and statistics. Yet *something like* statistics and probability is essential to an ergo-learner who must distinguish "significant/persistent" signals and ignore "accidental" ones.

(Certain rare signals may be more significant than frequent ones, e.g. a randomish sequence of, say, fifteen letters, such as "gifkumfinkotnid" which appears on three different pages of a book may impress you more than "representations" appearing six times or something like oooooooooooooooooooo coming between lines on twenty different pages.)

The above suggests what kind of mathematics is relevant for modeling ergo-learning; we briefly review some mathematical ideas in section 3, which, although *not directly applicable* to ergo-modeling, point toward avenues of thought one may pursue.

The reader may be surprised by our non-mentioning of "axioms for logical deduction", "Gödel theorem", "automata", "Turing machines", as well as of a logical/philospical analysis of such concepts as "thinking", "understanding", etc. Frankly, these concepts appear to me kind of "anti-mathematical", they do not fit into the folds of a geometer's ergobrain. (We shall explain this in 3.12 on a more technical level.) But they become more accessible when seen against a familiar not so logical background; we shall sketch several "real life" scenarios, which makes these concepts understandable for mathematicians and, we hope, to some non-mathematicians as well.

-

3 Structures within and next to Mathematics.

Mathematics is all about *amazing structures* clustered around *symmetries*: perfect symmetries, hidden symmetries, supersymmetries, partial symmetries, broken symmetries, generalized symmetries, linearized symmetries, stochastic symmetries. Two thirds of *core mathematics* (and of theoretical physics) would collapse if the symmetry axes had been removed.

But the peripheral branches of mathematics seen by an outsider are mainly made of useful formulas, difficult computations, efficient algorithms, logical axioms, reliable (or unreliable) statistics..., where the symmetry is diluted to the point of invisibility.

Mathematical concepts and ideas grow like flowers from tiny inconspicuous seeds if cultivated in a fertile soil. We shall demonstrate several instances of how this happens in geometry, algebra and combinatorics, where the *process* of growth, rather than the final result of it, is what will direct our ergo-modeling in the following chapters. Start with two illustrative examples.

Geometry is one and eternal shining in the mind of God.

JOHANNES KEPLER.

Apparently, ?number? is a difficult concept for the human ergobrain, albeit the neurobrain ? being a physical model runs on numbers (e.g. electric charges of neurons ?are? real numbers). We shall be careful in introducing ?numbers? to ergosystems.

We are no gods and our minds are not pure ergo. To build a mathematical frame for "ergo" we need to recognize what of our mathematics is ready to serve as "parts" of ergosystems, what should be rejected and what needs to be be made anew.[50] And "ergo-criteria" for these "ergo-parts" are exactly those we use everywhere in mathematics:

NATURALITY, UNIVERSALITY, LOGICAL PURITY and CHILDISH SIMPLICITY.

Universality of (many) learning programs in our ergobrains $\mathcal{EB}$ can be seen in the fact that we, humans, at least some of us, *enjoy and learn many* logically complicated games (and not only games). This suggests, for example, that a chess learning program in somebody's $\mathcal{EB}$ must be a specialization of a universal learning program for a rather generous concept of "universality".

On the other hand, why such programs should be simple? After all

The human brain is the most complicated object in the Universe. Isn't it?

But being mathematicians, we know that most general/universal theories are logically the simplest ones.[51] What is not simple is formulating/discovering such theories.

Also, as mathematicians we are ready to accept that we are hundred times stupider than the evolution is but we do not take it for the reason that evolution is able to make miracles, such as a logically complicated brain at birth. Believers into simplicity, we are compelled to seek our own solution to the *universal learning problem.*

As we aim at the very source of mathematics – ergobrain itself and try to develop a theory of ergosystems, purity and simplicity of the building blocks of such theory becomes essential. It is not logical rigor and technical details that are at stake – without clarity you miss diamonds – they do not shine in the fog of an ego-pervaded environment.

But our thinking is permeated by ego that makes hard for us to tell "true and interesting" from "important" and that makes the (ergo)right choices difficult. For instance, in the eyes of the egomind, *simple and concrete* is what you see in front of you; much of mathematics appears *abstract and difficult.* But this simplicity is deceptive and unsuitable for "ergo-purposes": what your eyes "see" is *not* simple – it is an outcome of an elaborate image building by your visual ergosystem that is, probably, more abstract and difficult than most of our mathematics.

Evolution of mathematical concepts in their convergence to clear shapes suggests how one may design ergosystems. Our mathematical diamonds have been polished and their edges sharpened – century after century, by scratching away layers of ego from their facets, especially for the last fifty years. Some of what came out of it may appear as "abstract nonsense" but, as Alexander Grothendieck points out,

The introduction of the cipher 0 or the group concept was general nonsense too, and mathematics was more or less stagnating for thousands of years because nobody was around to take such childish steps.

Yet, not all routes we explored had lead us to the promised land; understanding what and why did not work may be more instructive than celebrating

[50] Mathematics is the last born child of ergobrain and a *mathematician is an ergobrain's way of talking to itself* – as Niels Bohr would say.

[51] The simplicity of a universal idea, e.g. of *Gödel's incomleteness theorem*, may be obscured by plethora of technical details.

our successes.

3.1 Logic and Rigor.

Contrariwise, if it was so, it might be;
and if it were so, it would be;
but as it isn't, it ain't.
That's logic. LEWIS CARROLL

According to *logicism* of Frege, Dedekind, Russell and Whitehead mathematics is composed of atomic *laws of thought* dictated by formal logic and the rigor of formal logic is indispensable for making valid mathematical constructions and correct definitions.

Admittedly, logicians participated in dusting dark corners in the foundations of mathematics but... most mathematicians have no ear for formal logic and for logical rigor.[52] We are suspicious of "intuitive mathematical truth" and we do not trust *meta*mathematical rigor of formal logic.

(Logicians themselves are distrustful one of another. For example, Bertrand Russell, pointed out that Frege's *Basic Law V* was self-contradictory, while in Gödel's words,

[Russel's] *presentation ... so greatly lacking in formal precision in the foundations ... presents in this respect a considerable step backwards as compared with Frege.*

Russel's words *"Mathematics may be defined as the subject in which we never know what we are talking about, nor whether what we are saying is true"* apply to formal logic rather than to mathematics.)

Cleanness of things does not make them beautiful in the eyes of a mathematician. We care for logical pedantry as much as a poet does for preachings of grammarians.

Soundness of mathematics is certified by an ***unbelievably equilibrated** harmony* of its edifices rather than by the strictness of the construction safety rules. Criticism of insufficient rigor in mathematics by George Berkeley (1734) as well the idea of "redemption" of Leibniz' calculus by Abraham Robinson (1966) strike us as nothing but puny in the presence of the miraculous formula

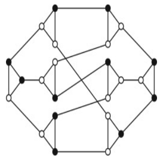
$$1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - ... = \frac{\pi}{4}$$

for $\pi = 3.14159265...$ being one half of the length of the unit circle. (Leibniz, 1682).[53]

Historically, the system of calculus rolling on fuzzy wheels of *infinity* and *infinitesimals*, has been the main intellectual force driving the development of mathematics and science for more than three hundred years. But just a step away from mathematics, volumes of philosophical speculations on the "true nature" of infinity remain on libraries shelves covered in dust year after year.

[52]We happily embrace *model theory, set theory, theory of algorithms* and other logical theories that became parts of mathematics.

[53]The achievement of Robinson from a working mathematician perspective was not so much in justification of Leibniz' idea of infinitesimals but rather in a vast and powerful extension of this idea.



(Unbelievably, as recently as at the beginning of the 20th century, Florian Cajori, then a leading historian of mathematics, hailed *The Analyst* – the treatise by George Berkeley who had lambasted "non-rigorous infinitesimals" – as *the most spectacular event of the* [18th] *century in the history of British mathematics.*

The landscape of the 18th century British mathematics was, indeed, so bleak that even *The Analyst* was noticeable. But there were, however, two English mathematicians who, unlike Berkeley, had left non-trivial imprints on the 18th century science –Thomas Bayes who suggested what is now called a *Bayesian* approach to empirical probability[54] and Edmond Halley famous for computing the orbit of *Halley's Comet.*[55])

We can not take seriously anything like $(a,b) \coloneqq \{\{a\},\{a,b\}\}$[56] but for some inexplicable reason this century old foundational dust finds its way to our textbooks under pretext of rigor as, e.g., in the following definition of a *graph G* as

an ordered [by whom?] *pair* $G = (V, E)$ *comprising a set* V *of vertices...*

We better keep clear of this "rigor".

Not everything in logic is collecting, cleaning and classifying morsels of common sense. In 1931, the *logician* Kurt Gödel defied everybody's intuition, including that of the *mathematician* David Hilbert who formulated the question a few yeas earlier, by *mathematically* proving that

Every ***formalization of mathematics*** *contains*
unprovable *propositions that can not but be regarded as being "**true**".*

Geometrically speaking,

the "body of mathematical truth" is DISCONNECTED.

(In fact, this "body" consists of *infinitely many* islands with *no bridges of deductive logic* joining them.[57])

Here "formalization of mathematics", denoted $\mathcal{MATH}$, means a "formal mathematical system or theory" – a language with a prescribed vocabulary and grammar rules. An essential property of such a $\mathcal{MATH}$ needed for the validity of Gödel's theorem is that $\mathcal{MATH}$ contains a sufficient vocabulary for speaking about languages $\mathcal{Y}$ regarded as mathematical objects. Basically, what one needs is the concept of a certain *mathematical property* to be *satisfied* by a given word

[54] Bayesian approach relies on continuous updating of conditional probabilities of events rather then on integrated frequencies; it is systematically used now-a-days in *machine learning*.

[55] Halley is the only short-period comet that is clearly visible from Earth when it returns to the inner solar system, approximately with 75 year intervals.

[56] This is the 1921 definition of an *ordered pair* by Kuratowski. To get "convinced" that this definition is worth making, you must accept logicians' appeal to metamathematical intuition.

[57] Udi Hrushovski was explaining to me several times (not that I have understood it) that this metaphor applies only to "bridges on a given level". Also he pointed out to me that what we call and prove here under the heading of *Gödel's theorem* is what logicians call *Tarski's "undefinability of truth" theorem.*

(a sentence if you wish) y in $\mathcal{Y}$ and/or to have a *proof* in $\mathcal{MATH}$. Then what $\mathcal{MATH}$ says about itself translates to Gödel's proof of the theorem.

Nothing special about $\mathcal{MATH}$ is needed for Gödel's theorem – it is valid for all "reasonable formal systems". One does not even have to know what a formal language is; all one needs is to spell out "reasonable" in general terms and apply *Cantor's Diagonal Argument* to some function F in two variables p and s, where this F says in effect that a certain "property" depending on p is satisfied or not by an s.

And to be "rigorous" one has to suffer through half a page of (inevitably?) boring definitions.

The vocabulary of a $\mathcal{MATH}$ must include the following.

• A set S, the members $s \in S$ of which are called *sentences in the the language of* $\mathcal{MATH}$.

• A set T called the *set of truth values for* $\mathcal{MATH}$. (In the "every day $\mathcal{MATH}$" this T consists of two elements **true** and **untrue** where meaningless sentences s are regarded as **untrue**.)

• A class $\mathcal{F}$ of T-valued functions $f(s)$ on S called *functions defined in* $\mathcal{MATH}$. (In "real math", such an f tells you whether a sentence s is true or untrue/meaningless.)

• A subset $P \subset S$ where sentences $p \in P$ are called *proofs*.

• A *reduction* map $R : p \mapsto f \in \mathcal{F}$ from P to $\mathcal{F}$ where the functions $f(s)$ in the image of R are called *provably defined in* $\mathcal{MATH}$. (This means that every proof p includes a "statement" of what it proves; this "statement" is called $R(p) \in \mathcal{F}$.)

Then Gödel's incompleteness theorem says that under the following assumptions (A) and (B)

the map R can not be onto:

there exist functions ***defined*** *in $\mathcal{MATH}$ that* ***can not be provably*** *defined.*

(A) The "P-diagonal" $F(p,p)$ of the T-valued function in two variables $F(p,s) = R(p)(s)$ admits an extension to a function on $S \supset P$, say $f_R(s)$, that is *defined in* $\mathcal{MATH}$.

(B) There exists a transformation $\tau : T \to T$, such that

(B_1) the composed functions $\tau \circ f : S \to T$ are *defined in* $\mathcal{MATH}$ *for all* $f \in \mathcal{F}$,

(B_2) τ has *no fixed point*: $\tau(t) \neq t$ for all $t \in T$.

(Properties (A) and (B_1) are satisfied by the "real world math" almost by definition, while (B_2) says that no sentence can be simultaneously **true** and **untrue**.)

Proof of Gödel's Theorem. By (A), the function $f_\circ(s) = \tau \circ f_R(s)$ is defined in $\mathcal{MATH}$; this $f_\circ(s)$ is different from the functions $f_p(s) = R(p)(s)$ for all p, since $f_p(s) \neq \tau \circ f_R(s) = \tau \circ R(p)(s)$ at $s = p$ because of (B_2).

Discussion. (a) Cantor's diagonal argument was designed for showoing that the set (space) of *all* functions $f : S \to T$ is *greater* than P for all $P \subset S$ and all T of cardinality at least two. This *greater* is strengthened and "quantified" in many geometric categories as follows.

No family $f_p = f_p(s)$ of functions on S contains *generic* $f = f(s)$.

This, applies, for instance, with several geometrically defined notions of genericity[58] for maps between Euclidean spaces where functions f may be continuous, smooth analytic or algebraic (and where genericity is accompanied by *transversality*).

On the other hand, explicitly described functions that one finds in "real life" (e.g. on Google) are, as we mentioned earlier, more scarce than, say, natural numbers n, partly, because descriptive (less so graphical) presentation of "interesting" functions occupies more space than that for numbers. We shall see similar patterns in the hierarchical organization of our ergosystems.

(b) The childish simplicity of the proof of Gödel's theorem[59] does not undermine its significance. *Metamathematics* is similar in this respect to other non-mathematical sciences where a mathematical argument is judged not by its difficulty but by its applicability to "real life". Nontriviality of Gödel's theorem resides in a possibility of a meaningful metamathematical interpretation of the above "provably defined".

In logical practice, the truth value set T usually (but not always) consists of two elements, say, *yes* and *no* with τ interchanging the two and, in Gödel's case, one takes $P = S$. Our functions $f(s)$ are associated with "properties" Π describable in the language of $\mathcal{MATH}$", with $f_\Pi(s)$, equal *yes* or *no*, depending upon whether Π is satisfied or not by s, where, in general, the truth value comes without being accompanied by a proof.

For example, a sentence s may describe an equation with Π saying "solvable", where an equation, is either solvable or not regardless of an availability of a proof of this in a given $\mathcal{MATH}$. (The certainty of this "either *yes* or *no*" is debatable even for *Diophantine* equations $f(x_1, ..., x_k) = 0$, i.e. where f is a *polynomial* with *integer* coefficients and where one speaks of *integer* solutions $(x_1, ..., x_k)$.)

By definition of P, a proof $p \in P$ that certifies correctness of the truth values $f_\Pi(s)$ at all s, "says" in particular, what is the property Π that this p proves; this information is extracted from p by the reduction map R.

But anything that can be called "rigor" is lost exactly where the things become interesting and nontrivial – at the interface between mathematics and "logical reality". For instance, a variation of Gödel's theorem may tell you that there exists a mathematical proposition that can be written, say, on 10 pages but the proof of which will need between $10^{10^{10}}$ and $1000^{1000^{1000}}$ pages. This is perfectly acceptable *within* mathematics but becomes non-sensical if you try to apply it to mathematics "embedded into the real world".

To see what makes us preoccupied with these "logical trifles", look closely at what stands behind the following *kindergarten Ramsey theorem*:

Given a group of six chess players, then necessarily,

[58]Geometry is non-essential here: concept of "genericity" belongs with mathematical logic. The universal logical power of "genericity" was *forcefully* demonstrated by Paul Cohen in his proof that the cardinality of "a generic subset" in continuum is *strictly* pinched between "countable" and "continuum".

[59]Originally, Gödel's theorem was stated for a certain formalization $\mathcal{ARITH}$ of arithmetic that was designed for talking about numbers rather than about languages; that necessitated a lengthy translation from the language of $\mathcal{ARITH}$ to the language in which one could formulate the theorem.

A transparant categorical rendition of Gödel's theorem is presented in "Conceptual Mathematics" by Lawvere and Schanuel [8]. This was pointed out to me by Misha Gavrilovich who also explained how the above proof may be seen as an adaptation of their argument.

either there are *three of them* where *every two* of these players once met over a chessboard,

or there are *three* such that *no game* was ever played between any two of them.

A child may *instantaneously* visualize a *graph* with green (played a game) and a red (never played a game) strings/sticks/*edges* between the pairs of these six people for *vertices*. (The child does not have to know graph theory.) Then the proof of the existence of a *monochromatic triangle* will come after a few minutes (hours?) thought.

Moreover, a mathematically inclined child will soon generalize this to the full Ramsey theorem:

*if the subsets of an infinite set X are colored either in green or in yellow, then, for every $k = 1, 2, 3, ...,$, this X contains an infinite k-**monochromatic** subset $Y = Y(k) \subset X$, i.e. where all k-element subsets are of the same color.*[60]

No existing computer program is anywhere close to doing this. The main difficulty is not finding proofs of mathematically stated Ramsey level theorems - these may be within the range of "symbol crunching" programs. It is automatic translation of the "real world" problems to mathematical language what remains beyond our reach. Probably, only a *universal* ergoprogram that would teach itself by reading lots of all kinds of texts will be able to achieve such translation.

LOGIC IN SCIENCE.

Mathematical rigor and logical certainty are absent not only from logical foundations of mathjematics but also from all natural sciences even from theoretical physiscs. Einstein puts it in words:

As far as the laws of mathematics refer to reality, they are not certain; and as far as they are certain, they do not refer to reality.

But "the physical level of rigor" is higher on certainty than the logical one, since reproducible experiments are more reliable than anybody's, be it Hilbert's, Einstein's or Gödel's, intuition.

I what follows, we shall be very choosy in in our terminology and in assigning basic concepts/operations to our learning systems, where nothing, no elementary structure, no matter how simple looking and "obvious", can be taken for granted.

3.2 What is Equality?

The starting point of finding a structure in a "flow of signals" is identifying "individual signals", such, for example, as phonemes, words and phrases in the flow of speech, or distinct "objects" in a visual field. Next, one has to decide which signals (e.g. words) are "equal" and which are not. We postpone the discussion of the "signal isolation problem" till ??? and indicate below instances of structures present in "equalities" (compare [**?**]).

Start with the familiar

$$1 + 1 = 2, \quad 2 + 2 = 4, \quad 3 + 3 = 6, \quad 4 + 4 = 8...$$

These look plain and boring to a casual eye, no individuality, no apparent structural beauty.

[60] Graphs correspond to $k = 2$, while 6 and 3 are equated with ∞ in kindergartens.

However, there is a significant structural difference between the first two: $1+1=2$, $2+2=4$ and the other equalities $n+n=2n$ for $n \geq 3$. Namely,

$1+1$ equals 2 in a unique way: the 2-point set has *only* 1 decomposition in two singleton;

$2+2$ equals 4 in three different ways: the 4-point set has 3 decompositions into two 2-point subsets.

In both cases the number d of decompositions is *strictly* less than the number being decomposed, i.e. $d=1<2$ and $d=3<4$.

But starting from $n=3$ the number of decompositions of a $2n$-set into two n-subsets equals $\frac{1}{2}\binom{2n}{n} = \frac{2n\cdot(2n-1)\cdot\ldots\cdot(n+1)}{n\cdot(n-1)\cdot\ldots\cdot 3} > n$ (e.g. $d=10>6$ for $2n=6$). In fact, apart from $1+1=_1 2$ and $2+2=_3 4$, the number d of decompositions of a finite set S of cardinality s into two subsets of given cardinalities satisfies $d \geq s$. (For example, $1+3=_4 4$, $2+3=_{10} 5$, $3+3=_{10} 6$, $2\times 3=_{60} 6$, etc.)

Let us concentrate on $2+2=_3 4$. The "triple of decompositions" yields a "canonical reduction" $4 \rightsquigarrow 3$. A classical implication of that, going back to Ferrari, Cardano, Lagrange, Abel and Galois, is the reduction of quartic equations to cubic ones via the homomorphism of the permutation groups, $\Pi(4) \to \Pi(3)$. Below is a fragment of the resulting (Ferrari) formula for a root x of the equation $x^4+bx^3+cx^2+dx+e=0$.

$$\ldots\sqrt{\ldots+\frac{\sqrt[3]{2}(c^2-3bd+12e)}{3\sqrt[3]{2c^3+\ldots-27b^2e-72ce+\sqrt{-4(c^2-3bd+12e)^3+(2c^3+\ldots+27b^2e-72ce)^2}}}}\ldots$$

(No such formula is possib le for equations of degrees ≥ 5 by Abel-Galois theory in concordance with the inequality $d \geq s$ for $s \geq 5$.)

A more recent amazing descendent of $2+2=_3 4$ is the world of geometric/topological structures on 4-manifolds (with the Lie algebras splitting relation $so(4)=su(2)\oplus su(2)$ as an intermediate, which says, in effect, that rotations in the 4-space are products of *commuting* 3-dimensional rotations).

The topology of n-manifolds, starting from Milnor's exotic 7-sphere of 1956, has been fully developed by the end of 1960s for $n \geq 5$, where (almost) all geometric problems were reduced to (generously understood) algebra with no (apparent) great mysteries left, with no (known) significant links with other fields and with no(?) futher perspective left open.

It was accepted that the dimensions 2 and 3 were special but it was believed that the high dimensional (surgery and alike) techniques would eventually conquer $n=4$. (This was confirmed in the topological category [**?**].)

The landscape has dramatically changed with the appearance of Donaldson's 1983 paper [**?**]. Not only the dimension four turned out to be quite different and richer in texture than $n \geq 5$, but it grew up full of vibrant connection with a new body of geometry, analysis and mathematical physics. After a quarter of a century, it shows no sign of aging and sprouts new green shoots every 5 - 10 years.

No ergosystem at the present moment would be able to arrive at the Donaldson theory by departing from $2+2=_3 4$, but it is possible that another humble mathematical something, if *properly* represented by an ergosystem, may turn into something unexpectedly interesting.

Egg Riddle. A man comes to a restaurant, orders to eat something slightly bizarre but innocuous – a boiled seagull egg, tries it and soon thereafter kills himself.

What is a possible story behind this macabre scene? Which questions, allowing only yes/no answers, would bring you closer to the solution ?

(We shall tell this at the end of this section where also we will explain what it has to do with the "equality problem".)

In my experience, only children ask questions which leads them straight to the point: a developed egomind biases you toward dozens of irrelevant questions. (Some of my friends solved it with only a half a dozen questions, but I am certain that they heard this puzzle before and forgot about it, since they performed as poorly as everybody else at similar problems. They themselves vehemently denied the fact.)

Let us return to "mathematical" equalities, now in a geometric context.

Given two copies Im_1 and Im_2 of the same image we regard them as equal but the equality $Im_1 = Im_2$ is more elaborately structured object than the above "$=_d$". In fact there are several kinds of equalities in this case.

Assume the two images lie on a flat table T.

● □ ● ● □ ● ● ● ● ● □ ● ● □ ●	● □ ● ● □ ● ● ● ● ● □ ● ● □ ●
□ ● ● ● ● □ ● ● □ ● ● ● ● □ ●	□ ● ● ● ● □ ● ● □ ● ● ● ● □ ●
● ● ● ● □ ● ● ● ● □ ● ● □ ● ●	● ● ● ● □ ● ● ● ● □ ● ● □ ● ●
● ● □ ● ● ● ● □ ● ● ● □ ● ● □	● ● □ ● ● ● ● □ ● ● ● □ ● ● □

The equality may be achieved by sliding the first image on the table till it merges with the second one. The result of such "slide", can be regarded as a rigid motion – isometry of the infinite (Euclidean) plane extending T; thus, "=" is implemented by an element *iso* of the isometry group $isom(\mathbb{R}^2)$ of the Euclidean plane $\mathbb{R}^2$ and "=" reads "$=_{iso}$".

The group element *iso*, in turn, can be realized by a variety of actual motions of Im_1 to Im_2 within T. Such a motion is a path *pth* in the group $isom(\mathbb{R}^2)$ issuing from the identity element of $isom(\mathbb{R}^2)$ (which keeps the plane $\mathbb{R}^2$ unmoved) to out *is*; this upgrades "$=_{iso}$" to "$=_{pth}$".

However, your eye, while establishing $Im_1 = Im_2$ does none of the above: it rather jumps back and forth between the images "checking " similarity between "corresponding patterns". This is another kind of a structured equality, $Im_1 =_{eye} Im_2$.

Deja Vu.

$$\sqrt{1+2\sqrt{1+3\sqrt{1+4\sqrt{1+5\sqrt{1+\cdots}}}}}.$$

This is the *same* expression as in the first Ramanujan formula from 1.7 and the "equality" between the two is established by an (ergo)mental mechanism which, apparently, involves the memory but the overall (neuro-brain) structure of which remains obscure.

What is perceptible, however, is "the deja vu click" inside your head. Such "clicks" have certain universality but also they may have variable flavors and intensities; also they may carry additional structures, e.g. be accompanied by two distinguishable soft low grunts, if you are a chimpanzee, for instance.

Conversely, if you expect such click and it does not come, you may experience another kind of click – a feeling of being perplexed/amazed upon encountering something new and unusual in a familiar surrounding, e.g. a particular smell, which, probably, can be quantified with something like conditional entropy of the smell signal reaching your brain. (Smell processing by the brain is not a simple matter as it may seem, not even for insects, such as honey bee with the nervous system of $< 10^6$ neurons. Mammals, even olfactory challenged Homo sapience, have tens of millions olfactory receptor neurons in their nasal cavities reaching $\approx 4 \cdot 10^9$ in a bloodhound's nose. [Olfaction, Wikipedia], [**?**].)

In sum, "equalities/nonequalities" are structural entities which are employed and simultaneously analyzed as well as modified by ergosystems in the course of learning. Our insight into these depends, almost exclusively, on our experience with "the equality structures" in mathematics.

3.3 Miracle of Numbers.

All the mathematical sciences are founded on relations between physical laws and laws of numbers.
JAMES CLERK MAXWELL.

The existence of Mathematics as we know it strikes one as improbable as emergence of Life on Earth. Nothing in the foundation of mathematics suggests such thing is possible, like nothing in the Earth chemistry suggests it can beget Life.

One may say that mathematics starts with *numbers*. We are so used to the idea that we forget how *incredible* properties of *real numbers* are. The seamless agreement of several different structures – *continuity, order, addition, multiplication, division* – embodied into this *single* concept is amazing.

Unbelievably perfect *symmetries* in geometry and physics – *Lie groups, Hilbert spaces, gauge theories...*–emerge in the world of numbers from the seed of the Pythagorean theorem. Mathematics and theoretical physics are the two facets of these symmetries that are both expressed in the essentially same mathematical language.

As Poincare says,

... without this language most of the intimate analogies of things would forever have remained unknown to us; and we would never have had knowledge of the internal harmony of the world, which is, as we shall see, the only true objective reality.

In the "harsh real world", away from pure mathematics and theoretical physics, the harmony of the full "symmetry spectrum" of numbers comes into play only rarely. It may even seem that there are several different kinds of numbers: some may be good for *ordering* objects according to their size and some may be used for *addition* of measured quantities. Using the all-powerful real numbers for limited purposed may strike you as wasteful and unnatural.

For example, *positive* numbers appear in classical physics as *masses* of bulks of matter while electric charges represent positive and negative numbers. The relevant *operation* with these numbers is *addition*, since mass and electric charge are naturally (nearly perfectly) additive: $(a, b) \mapsto a + b$ corresponds to bringing two physical objects together and making a single $(a + b)$-object out of the two corresponding to a and to b.

But there is no comparably simple implementation of, say, $a \mapsto 2a$ – one can not just copy or double a physical object. And writing $2a = a + b$ for $a = b$ does not help, since mutually equal macroscopic physical objects do not come by themselves in physics.

In contrast, doubling is seen everywhere in Life. All of us, most likely, descend from a polynucleotide molecule which had successfully doubled about four billion years ago. Organisms grow and propagate by doubling of cells. Evolution is driven by doublings of genomes and of significant segments of the whole genomes (not by the so called "small random variations").

A true numerical addition may be rarely (ever?) seen in *biology proper* but, for example, additivity of electric charges in neurons is essential in the function of the brain. This underlies most mathematical models of the neurobrain, even the crudest ones such as neural networks. But the ergobrain has little to do with additivity and linearity.[61]

The apparent simplicity of *real numbers* represented by points on an infinite straight line is as illusory as that of visual images of the "real world" in front of us. An accepted detailed exposition (due to Edmund Landau) of real numbers by *Dedekind cuts* (that relies on the order structure) takes about hundred pages. In his book *On Numbers and Games*, John Conway observes (and we trust him) that such an exposition needs another couple hundred pages to become complete.

To appreciate this "problem with numbers", try to "explain" real numbers to a computer, without ever saying "obviously" and not resorting to anything as artificial as decimal/binary expansions. Such an "explanation computer program" will go for pages and pages with a little bug on every second page.

We shall not attempt to incorporate the full theory of real numbers in all its glory into our ergosystems, but some "facets of numbers" will be of use. For example we shall allow an ergo-learner the ability of distinguishing frequent and rare events, such as it is seen in behaviour of a baby animal who learns not to fear *frequently* observed shapes.

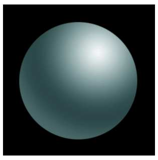
On the other hand, while describing and analyzing such systems we shall use real numbers as much as we want.

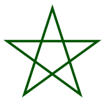
The shape of the heaven is of necessity spherical.
ARISTOTLE.

Numbers are not in your ergobrain but the idea of symmetry is in there. Much of it concerns the symmetries of our (Euclidean) 3-space, the essential ingredient of which – the group of the *(3-dimensional Lie) group* $O(3)$ of rotations of the Euclidean *round 2-sphere* within itself – has been facinating mathematicians and philosophers for millennia. And not only "the haven" but also your eyes and some of your skeletal joints that "talk" to the brain are by necessity spherical; hence, *rotationally symmetric.*

(The rotation group $O(2,1)$ of the non-Euclidean *hyperbolic plane*, that is logically more transparent than $O(3)$ as it can be represented by symmetries of a calender [SLE, §2.1], was discovered less than two centuries ago. This group along with $O(3)$ serves as as a building block for other *simple Lie groups* that are representatives of essential geometric symmetries.)

[61] "*Non-linear*" customary applies to systems that are set into the framework of numbers with their *addition structure* being arbitrarily and unnaturally contorted.





A plausible (ergo)brain's strategy for *learning space*, in particular, for reconstruction of spacial symmetries from the retinal images of moving objects, was suggested by Poincaré in §IV of *La science et l'hypothèse*, where Poincaré indicates what kind of mathematics may be involved in *learning space by our visual system*. An aspect of our "ergo-approach" is an attempt to spell out what Poincaré might have in mind.[62]

Our ergobrain is also sensitive to *arithmetic symmetries* that issue from prime numbers as is seen in the recurrence of the *magical pentagram figure* depicting the *finite (Galois) field* $\mathbb{Z}_5$ with the miraculous symmetry of $20(=5\cdot(5-1))$ (affine) transformations acting on it.

A fantastic vision, unimaginable to ancient mystics and to mediaeval occultists, emerges in the *Langlands correspondence* between arithmetic symmetries and the *Galois symmetries* of algebraic equations, where much of it is still in the clouds of conjectures. It is tantalizing to trace the route by which the ergobrain has arrived at comprehension of this kind of symmetries.

3.4 Big and Small.

Mathematicians treat all numbers on equal footing, be these

2, 3, 4,

or 20, 30, 40,

or 1 000, 10 000, 100 000, 1 000 000,

or 10^{10}, 10^{20}, 10^{30}, 10^{40},

or 10^{10^2}, $10^{10^{30}}$, $10^{10^{400}}$

But "democracy of numbers" breaks down in the "real world", be it the physical Universe or the human ergo-world.

For example, the grammars of some languages, e.g. of Russian, distinguish the numbers 2, 3 and 4, while 5, 6, ..., 20, 30, 40,..., 100, 200,...(but not, say 23, 101 and 202) are put, syntactically speaking to "the same basket as infinity.[63]

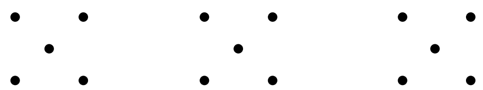
One instantaneously evaluates the cardinality of • • • •, one needs a fraction of a second to identify "unstructured five" • • • • •, it takes a couple of seconds for • • • • • • • (it is much faster if the symmetry is broken, for instance, as in • • •   • • • •) and it is impossible with

[62]A similar idea can be seen in Sturtevant's 1913 construction of the first *genetic map* as we explain in $4 of [4].

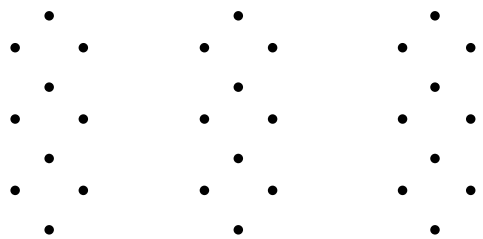
[63]Amusingly, there is also a chasm in essential properties between geometric spaces of dimensions 1,2,3,4, and those of dimension 5 and more.

• • • • • • • • • • • • • • •.

But a little structure helps:



And slightly larger numbers, such as



if perceived, then through a lens of mathematics.

Our intuition does not work anymore when it comes to thousands, millions, billions. Answer fast:

Do you have more hairs on your head (assuming you are not bald) than the number of people an Olympic stadium may contain?[64]

What is greater the number of bacteria living in your guts or the number of atoms in a bacterium?[65]

Below are ergo-relevant numbers.

• TIME. Hundred years contain < 3.2 billion seconds. With the rate three words per second you vocalize *less than ten billion* (10^{10}) words in the course of you life.

Ten billion garrulous individuals all together[66] will utter *at most*

$$10^{10} \times (3 \times 3.2 \cdot 10^7) \times 5 \cdot 10^9 < 5 \cdot 10^{27}$$

words until Sun turns into a *red giant* in about five billion years.

Speaking more realistically, humanity *can not* come up with more than 10^{12} -10^{18} *different ideas* — poems, theorems, computer programs, descriptions of particular numbers, etc.[67]

10^{15} years of possible duration of the Universe is made of less than $10^{46} = 10^{15} \times 3 \cdot 10^7 \times \frac{1}{3}10^{24}$ *jiffie-moments.*[68]

• BRAIN. The number of neurons in the human brain is estimated between ten and hundred billion neurons with hundreds synaptic connections per neuron, somewhere 10^{12}-10^{14} synapses all together.

This gives an idea on the volume of the memory stored in the brain, that is comparable to that on a computer hard disk of about 10^{12}-10^{13} bits.

The (short time) brain performance is limited by the *firing rates* of neurons – something about 100 times per second. Thus, say hundred million active

[64]Both numbers are about 100 000.

[65]There about 10^{11}-10^{14} atoms in bacteria and more than 10^{12}-10^{13} bacteria living in your body, mainly in your guts.

[66]The human population on Earth today is slightly above seven billion.

[67]LIFE on Earth, in the course of its $\approx 3.9 \cdot 10^9$ year history, has generated a comparable number of "ideas" and recorded them in DNA sequences of organisms inhabiting the planet.

[68]*Jiffy* $\approx 3 \cdot 10^{-24} s$ is the time needed for light to travels the proton-sized distance.

neurons can perform 10^{10} "elementary operations" per second[69] that is what an average computer does.[70]

• SPACE. A *glass of water* contains about 10^{25} molecules, the *planet Earth* is composed of about 10^{50} atoms and the *astronomically observable universe* contains, one estimates today, 10^{80} particles.[71]

Thus, there are (significantly) less than 10^{130} *classical* (as opposed to quantum) "events" within our space-time and this grossly overestimated number makes

an unquestionable bound of what will be ever achieved by any conceivable (non quantum) computational/thinking device of the size of the Universe.

But... there are at least $2^{10^{10}} > 10^{3\,000\,000\,000} >> 10^{130}$ possible "texts" that you, a humble 21st century human being, *can* (?) write in sequences s of 10^{10} bits on the hard disc of your tiny computer. Can't you?

How comes that only a negligible percentage, less than $\frac{1}{10^{10^9}}$ of possibilities, can be actualized?

Worse than that, it is *impossible* to pinpoint a *single instance* of non-realizable sequence s: indicating an s will make this very s actual.

It is far from clear whether such inconsistency between "can" and "will" admits a clean mathematical reformulation or this belongs with the *paradox of the heap.* Yet, there are a few purely mathematical theorems and open problems that address this issue, albeit not satisfactorily.

• The oldest is the so called *Scolem's paradox*, that is a theorem in the mathematical logic saying that *uncountably* many mathematical objects (sets) can be "adequately represented" by *countably* many "verbal descriptions".

• It is often (almost always?) quite difficult to *explicitly construct* a single mathematical object O that satisfies a certain (non-trivial!) property P, despite (because of?) a presence of a counting or similar argument showing that a "predominate majority" of objects O do satisfy P.

The most annoying open class of such problems concerns explicit construction of "simple" functions $f(n)$, $n = 1, 2, 3, ...$, evaluation of which needs *long computations.*

Ad hoc Example. Let $d_n(\pi)$ be defined as the 10^n-th digit of

$$\pi = 3.1415\ 92653\ 58979\ 32384\ 62643\ 38327\ 95...\ .$$

Here, one sees that $d_1(\pi) = 3$, $d_2(\pi) = 4$, $d_3(\pi) = 7$. As for today, *ten trillion* (10^{13}) digits of π were computed[72] with a use of versions of the

RAMANUJAN MYSTERIOUS FORMULA

$$\frac{1}{\pi} = \frac{2\sqrt{2}}{9801} \sum_{k=0}^{\infty} \frac{(4k)!(1103 + 26390k)}{(k!)^4 396^{4k}} = \frac{2\sqrt{2}}{9801} \left(1103 + \frac{24 \cdot 27493}{396^4} + ...\right).$$

[69] But the rate of learning is measured not in seconds but in hours, days, months, years. This is so, partly, because modification of the strength of synaptic connections is slow.

[70] The speed of modern *supercomputers* is measured in *petaflops* corresponding to 10^{15} (*floating point*) operations per second. This is achieved with particularly designed network architectures of processors that allow thousands (not millions as in the brain) operations performed *in parallel.*

[71] Archimedes evaluated the number of sand grains that would fill the Universe by $\approx 10^{60}$ where *exponential representation of numbers* was invented by him for this purpose.

[72] The computation took more than a year.

Thus, one knows the values $d_n(\pi)$ up to $n = 13$.

One can envisage a similar brut force computation of $d_n(\pi)$ up to $n = 16$ or even $n = 19$ but, probably, the sequence $d_1(\pi)$, $d_2(\pi)$, $d_1(\pi)$, ..., $d_{1000}(\pi)$ is not "written" anywhere in the space-time continuum of *our* Universe and the number $d_{1000}(\pi)$ will be never determined by any superhuman civilization.[73]

Beside such artifacts as $d_1(\pi), d_2(\pi), d_3(\pi), ...$ there are mathematically meaningful objects in the "real world" that are also beyond our present day computational prowess. For instance, it is unlikely that a genome sequence of a viable organism, e.g. of a photosynthesizing plant, can be (re)constructed with the only input/knowledge being "*fundamental laws of physics*" (if these exist) by a computation composed of 10^{25}-10^{30} "elementary steps".[74]

There is a "natural" class of "simply describable" functions that is called NP. The famous $P \neq NP$ *problem*/conjecture says that there are functions $f = f(n)$ from NP that *can not* be computed in ***p**olynomial time*, say, in at most $const_f \cdot n^3$ steps for some constant depending on f. The stark failure in solving this problem shows how limited our vision of the basic structure of *computations in mathematics* is.[75]

On the positive side, despite the overall lack of progress in understanding *"generic-versus-effective"* in the mathematical world, there were a few successes, e.g. specific constructions of graphs with "random, properties" e.g. of *expander graphs.*

3.5 Probabilities Old and New.

The true logic of this world is the calculus of probabilities.
JAMES CLERK MAXWELL.

The notion of a probability of a sentence is an entirely useless one, under any interpretation of this term. NAUM CHOMSKY.

Human languages carry imprints of the mathematical structure(s) of the ergobrain and, at the same time, learning a natural (and also a mathematical[76]) language is a basic instance of the universal learning process by the human ergobrain. We hardly can understand how this process works unless we have a fair idea of what LANGUAGE is. But it is hard to make a definition that would catch the *mathematical essence* of the idea of LANGUAGE.

But isn't a language, from a mathematical point of view, *just*

a set of strings of symbols from a given alphabet,

or, more generally,

a probability distribution on the set of such strings?

A linguist would dismiss such definitions with disgust, but if you are a mathematician these *effortlessly* come to your mind. Paradoxically, this is why we

[73]A miraculous advance in quantum computing or an equally miraculous discovery in mathematics of π can make it possible but this would hardly help, for instance, in deciding wether the number digitally represented by $0.d_1(\pi)d_2(\pi)d_3(\pi), ...,$ is transcendental.

[74]Nature managed doing this in something like 10^{50} steps.

[75]The concept of *"computational polynomiality"* makes little sense outside pure math: a mathematician may be happy with anything like $10^9 \cdot (10^6)^3$ but this is as bad (good?) as infinity when it comes to the "real life". Learning algorithms in the (ergo)brain must be effectively *log×linear,* rather than polynomial.

[76]*Mathematical language* for us is the language used for communication between mathematicians but not a mathematical language of formal logic.

would rather *reject* than accept them:

Mathematics is shaped by definitions of its fundamental concepts, but there is no recipe for making "true definitions". These do not come to one's mind easily, nor are they accepted by everybody readily.

For example, the idea of an *algebraic curve* that is a *geometric* representation of

solutions of a polynomial equation $P(x_1, x_2) = 0$ *in the* (x_1, x_2)*-plane*

by something like ⬭, originated in the work by Fermat and Descartes in 1630's and these curves have been studied in depth by generation after generation of mathematicians ever since.

But what is now seen as the simplest and the most natural definition of such a curve – the one suggested by Alexander Grothendieck in 1950s in the language of *schemes*, would appear absurd, if understood at all, to anybody a few decades earlier.

Defining "language" and/or "learning" is, non-surprisingly, more difficult than "algebraic curve", since the former have non-mathematical as well as purely mathematical sides to them. They are similar in this respect to the concept of *probability* that by now is a well established mathematical notion.

It is instructive to see how "*random*" crystallized to *"probability"*, what was gained and what was lost in the course of this "crystallization".

Also, we want to understand how much of "random" in languages in (ergo)learning process (including learning languages) is amenable to what Maxwell calls "the calculus of probabilities".

The concept of *chance* is centuries old as is witnessed by some passages in Aristotle (384– 322 BCE) and also in Talmud.[77] And Titus Lucretius (99 –55 BCE), a follower of Democritus, describes in his poem *De Rerum Natura* what is now called *Einstein-Smoluchowski stocahstic model* of Brownian motion[78].

But mathematics of "random" was originally linked to gambling rather than to science.

I of dice possess the science and in numbers thus am skilled

said Rituparna, a king of Ayodhya, after estimating the number of leaves on a tree upon examining a single twig. (This is from *Mahabharata*, about 5 000 years ago; also 5 000 years old dice were excavated at an archeological site in Iran.)

What attracts a mathematician to random dice tossing and what attracts a gambler are the two complementary facets of the *stochastic symmetry*.

Randomness *unravels and enhances* the *cubical symmetry* of dice (there are $3! \times 2^3 = 48$ symmetries/rotations of a cube) – this is what fascinates a mathematician.

But randomness also *breaks* symmetries: the only way for a donkey' ergobrain (and ours as well) to solve Bouridan's ass problem is to go random.[79]

[77]Our sketchy outline of the history of probability relies on [10] [2], [14], [7], [6], [15] with additional *References for Chronology of Probabilists and Statisticians* on Ming-Ying Leung's page, http://www.math.utep.edu/Faculty/mleung/mylprisem.htm

[78]This is the collective random movements of particles suspended in a liquid or a gas that should be rightly called *Ingenhousz' motion*.

[79]No deterministic algorithm can select one of the two points in the (empty) 3-space as it follows from the existence of the *Möbius strip*. And a general purpose robot that you can ask, for instance, *bring me **a** chair* (regardless of several available chairs being identical or not)

Emanation of the "miraculous decision power of random" intoxicates a gambler's ergo.[80]

The first(?) documented instance of the *calculus* of probabilities – "*measuring chance*" by a European[81] appears in a poem by Richard de Fournival (1200-1250) who lists the *numbers* of ways three dice can fall. (The symmetry group in the case of n dice has cardinality $n! \times (48)^n$ that is 664 552 for $n = 3$.)

Next, in a manuscript dated around 1400, an unknown author correctly solves an instance of *the problem of points*, i.e. of division of the stakes.

In 1494, the first(?) treatment of the problem of points appears in *print*[82] in Luca Paccioli's *Summa de Arithmetica, Geometria, Proportioni et Proportionalita.*[83]

Paccoli's solution was criticized/analized by Cardano in *Practica arithmetice et mensurandi singularis* of 1539 and later on by Tartaglia in *Trattato generale di numerie misure*, 1556.

About Cardano.

Gerolamo Cardano was the second after Vesalius most famous doctor in Europe. He suggested methods for teaching deaf-mutes and blind people, a treatment of syphilis and typhus fever. Besides, he contributed to mathematics, mechanics, hydrodynamics and geology. He wrote two encyclopedias of natural science, invented *Cardan shaft* used in the to-days cars and published a foundational book on algebra. He also wrote on gambling, philosophy, religion and music.

The first(?) systematic mathematical treatment of statistic in gambling appears in Cardano's *Liber de Ludo Aleae*, where he also discusses the psychology of gambling, that written in the mid 1500s, and published in 1663.

In a short treatise written between 1613 and 1623, Galileo, on somebody's request, effortlessly explains why upon tossing three dice the numbers (slightly) more often add up to 10 than to 9. Indeed, both

$$9 \underset{1}{=} 1+2+6 \underset{2}{=} 1+3+5 \underset{3}{=} 1+4+4 \underset{4}{=} 2+2+5 \underset{5}{=} 2+3+4 \underset{6}{=} 3+3+3$$

and

$$10 \overset{1}{=} 1+3+6 \overset{2}{=} 1+4+5 \overset{3}{=} 2+2+6 \overset{4}{=} 2+3+5 \overset{5}{=} 2+4+4 \overset{6}{=} 3+3+4$$

have six decompositions, but 10=3+ 3 +4=3+4+3=4+3+3 is thrice as likely as 9=3+3+3.

(If you smile at the naivety of people who had difficulties in solving such an elementary problem, answer, instantaneously,

What is the probability of having two girls in a family with two children where one of the them is known to be a girl?[84])

needs a "seed of randomness" in its software.

[80]In the same spirit, the ABSOLUTE ASYMMETRY of an individual random ± sequence of outcomes of coin tosses complements the ENORMOUS SYMMETRY of the whole space S of dyadic sequences that is acted upon by the compact Abelian group $\{-1,1\}^{\mathbb{N}}$ for $\mathbb{N} = \{1,2,3,4,5,...\}$ and by automorphisms of this group.

[81]Some "calculus of probabilities", can be, apparently, found in the *I Ching* written about 31 centuries ago.

[82]The first book printed with movable metal type was Gutenberg Bible of 1455.

[83]Paccioli became famous for the system of *double entry bookkeeping* described in this book.

[84]This would take half a second for Galileo – the answer is 1/3 ($\pm\varepsilon$).

Formulation of basic probabilistic concepts is usually attributed to Pascal and Fermat who discussed gambling problems in a few letters (1653-1654) and to Huygens who in his 1657 book *De Ratiociniis in Ludo Aleae* introduced the idea of *mathematical expectation*.

But the key result – *the Law of Large Numbers* (hinted at by Cardano) was proved by Jacob Bernoulli only in 1713.

This, along with the *Pythagorian theorem* and the *quadratic reciprocity law*[85] stands among the ten (±2) greatest mathematical theorems of all time. To appreciate its power look at the following example relevant to some (ergo)-learning algorithms.

Let X be a finite set, e.g. the set of numbers $1, 2, 3, ..., N$ and let Θ be a collection of (test) subsets $T \subset X$. Say that a subset $Y \subset X$ is Θ-*median* if the cardinalities of the intersections of Y with the members T of Θ satisfy

$$\frac{1}{3}card(T) \leq card(T \cap Y) \leq \frac{2}{3}card(T) \text{ for all } T \in \Theta.$$

A slightly refined version of the Law of Large Numbers implies that if Θ contains *at most* $2^{M/10}$ (test) subsets $T \subset X$, for $M = \min_{T \in \Theta} card(T)$, i.e. if

$$card(\Theta) \leq 2^{card(T)/10} \text{ for all } T \in \Theta,$$

then,

for "large" M, "most" subsets $Y \subset X$ with $card(Y) = \frac{1}{2}card(X)$ are Θ-median.

(If $card(X)$ happened to be odd, let $card(Y) = \frac{1}{2}card(X) + \frac{1}{2}$.)

In particular,

if $M \geq 10$ and $card(\Theta) \leq 2^{M/10}$ then X contains a Θ-median subset $Y \subset X$.

What is interesting is that even if a collection Θ is defined by "simple explicit rules", say in the case $X = \{1, 2, 3, ..., N\}$, there may be *no* "simple description" of any Θ-median subset Y, albeit we do know that such a Y does exist. (This is a characteristic instance of the poorly understood *generic-versus-effective* phenomenon mentioned in the previous section.)

Example. Let $X = X_N$ equal the set of integers $1, 2, ..., N$ and $\Theta = \Theta_M$ be the set of all arithmetic progressions T of length M in this X_N.

If $M \geq 1\,000$ and $N \leq 10^{20}$, then Θ-median subsets $Y \subset \{1, 2, ..., 10^N\}$ exist.

But exhibiting any single one of them, say for $M = 1000$ and $N = 10^{12}$ seems difficult.[86] And effective description of M-median subsets $Y \subset X = \{1, 2, ..., N\}$ becomes progressively harder for tricker, yet, explicitly described Θ.

"Continuous probability" was invented in 1733 by Buffon who thought of

[85]Let p, q be odd primes and $q^* = (-1)^{(q-1)/2}q$. Then $n^2 - p$ is divisible by q for *some* integer n if and only if $m^2 - q^*$ is divisible by p for *some* m.

[86]Conjecturally, if $N \geq 10^M$, then *no* Θ-*median* subset $Y \subset \{1, 2, ..., N\}$ exists for this $\Theta = \Theta_M$ (made of arithmetic progressions of length M), but this is known only for much larger N, e.g. for $N \geq 2^{2^{2^{2^N}}}$ by Gowers' refinement of *the Baudet-Schur-Van der Waerden-Szemeredi theorem.*

a needle of unit length (instead of dice) randomly thrown on the plane, where this plane was divided into parallel strips of unit width.

He proved that

the probability of crossing a line between two strips by the needle equals $2/\pi$ *for* $\pi = 3.14...$ *being one half the length of the unit circle*

About Georges-Louis Leclerc Buffon.

Besides opening the fields of *geometric probability* and *integral geometry*, Buffon also contributed to optics: lenses for lighthouses and concave mirrors of his design have been in use for two centuries afterwards.

But his major contribution was to what he called "natural history" – a development of a synthetic picture of Life on Earth, where he outlined many essential interactions between organisms and their environment, much of which is now goes under the heading of "biogeography".

Buffon emphasized the preeminence of biological reproduction barriers between different groups of organisms over the obvious geographical ones that suggested a definition of *species* that has withstood the attempts to "improve" it by later natural philosophers including some 20th century post Darwinian evolutionary thinkers.

Buffon was the first(?) who articulated the main premise of the evolutionary biology – the concept of the *common ancestor of all animals*, including humans.

Buffon's view on Nature and Life, expounded in his *Histoire naturelle, génèrale et particuliere* published between 1749 and 1789 in 36 volumes, became a common way of thinking among educated people in Europe for two centuries afterwards.

With the Buffon's needle, "random" merged with "analysis of continuum" and were empowered by "calculus of infinitesimals". This is what was hailed by Maxwell and exploited by generations of mathematicians and physists after Buffon.[87]

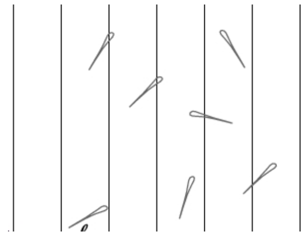
This calculus comes at a price: probability is a "full fledged number" with the addition/multiplication table behind it. But assigning a *precise specific* numerical value of probability to a "random event" in "real life", e.g. to a sentence in a language, is not always possible.

Apparently, the elegance and success of probabilistic models in mathematics and science (always?) depends on (often tacitly assumed and/or hidden) symmetry.

(A bacterium size speck of matter may contain, say, $N_{AT} = 10^{12}$ atoms and/or small molecules in it and the number N_{BA} of bacteria residing in your colon is also of order 10^{12}. If there are two possible states for everyone – be they atoms or bacteria – then the number of the *conceivable* states of the entire system, call it S, is the monstrous

$$M = M(S) \geq 2^{10^{12}} > 10^{3\,000\,000\,000}$$

[87]The brightest supernova in the 19th century sky of science, as it is seen from the position of the 21st century, was the 1866 article *Versuche über Pflanzen- Hybriden* by Gregory Mendel who derived the *existence of genes* – atoms of heredity by a statistical analysis of the results of his experiments with pea plants. The world remained blind to the light of this star for more than 30 years.



where its reciprocal

$$\frac{1}{M} < 0.\underbrace{000...000}_{3\,000\,000\,000}1$$

taken for the probability of S being in a particular state is too small for making any experimental/physical/biological sense.

However, the assignment of the $\frac{1}{M}$-probabilities to the states is justified and will lead to meaningful results IF, there is a symmetry that makes these tiny meaningless states "probabilistically equivalent", where the nature of such a symmetry, if it is present at all, will be vastly different in physics and in biology.[88]

On Symmetry in Randomness.

Essentiality of "equiprobable" was emphasized by Cardano and *parametrization* of random systems by "independent variables" has always been the main tenet of the probability theory. Most (all?) of the classical mathematical probability theory was grounded on *(quasi)invariant Haar(-like) measures* and the year 2000 was landmarked by the most recent triumph of "symmetric probability" – the discovery of (essentially) *conformally invariant* probability measures in spaces of planar curves (and curves in Riemann surfaces) parametrized by increments of *Brownian's processes* via the *Schram-Loewner evolution equation.*

But if there is not enough symmetry and one can not *postulate* equiprobability (and/or something of this kind such as *independence*) of certain "events", then the advance of the classical calculus stalls, be it mathematics, physics, biology, linguistic or gambling.

On Randomness in Languges.

Neither unrealistic smallness of probabilities, nor failure of "calculus with numbers" preclude a use of probability in the study of languages and of learning processes. And if you are too timid to contradict Chomsky, just read his "*under any interpretation of this term*" as *"under any interpretation of the term probability you can find in a 20th century textbook".*

Absence of numbers for probabilities in languages is unsurprising – numbers are not the primary objects in the ergoworld. Numbers are not there, but there is a visibly present *partial order* on "plausibilities" of different sentences in the language. This may look not much, but a *hierarchical use* of this order allows recovery of many linguistic structures as we shall see later on.

[88]It is not fully accidental that the numbers N_{AT} and N_{BA} are of the same order of magnitude. If atoms were much smaller or cells much bigger, e.g. if no functional cell with less than 10^{20} atoms (something slightly smaller than a Drosophila fly) were possible, then, most probably, LIFE, as we know it, could not have evolved in our short lived Universe with hardly 10^{80} atoms in it.

An essential problem with probability is a mathematical definition of "events" the probabilities of which are being measured.

The now-a-days canonized solution, suggested in 1933 by Kolmogorov in his *Grundbegriffe der Wahrscheinlichkeitsrechnung*, is essentially as follows.

Any kind of randomness in the world can be represented (modeled) geometrically by a subdomain Y in the unit square ■ *in the plane. You drop a points to* ■*, you count hitting Y for an* ***event*** *and define the probability of this event as $area(Y)$.*

However elegant this *set theoretic* frame is, (with ■ standing for a *universal probability measure space*) it must share the faith of André Weil's *universal domains* from his 1946 book *Foundations of Algebraic Geometry*. The set theoretic language introduced in mathematics by Georg Cantor that has wonderfully served us for almost 150 years is now being supplanted by a more versatile language of *categories* and *functors*. André Weil's varieties were superseded by Grothendieck's schemes, and Kolmogorov's definition will eventually go through a similar metamorphosis.

A particular path to follow is suggested by Boltzmann's way of thinking about statistical mechanics – his ideas invite a use of *non-standard analysis* as well as of a Grothendieck's style *category theoretic* language. (This streamlines Kolmogorov"s ■ in certain applications as we explain in [5].) But a mathematical interpretation of the idea of probability in *languages* and in *learning* needs a more radical deviation from (modification? generalization of?) this ■.

CARDANO, GALILEO, BUFFON. The very existence of these people challenges our vision on the range and spread of the human spirit. There is no apparent wall between the ergos and egos in the minds of these men.

Where are such people to-day? Why don't we see them anymore? Nobody in the last 200 years had a fraction of Cardano's intellectual intensity combined with his superlative survival instinct. Nobody since Buffon has made long lasting contributions to domains as far-distant one from another as pure mathematics and life sciences. What needs to be done to bring Galileos back to us?

4 Flows of Signals.

4.1 Seven Flows.

... think of some step that flows into the next one, and
the whole dance must have an integrated pattern.
FRED ASTAIRE

Incoming flows of signals can be divided according to the sensory receptors and pathways by which they enter the brain: *visual, auditory* and *somatosensory* where the two relevant aspect of the latter are *proprioception* – the body sense, and *tactile*, i.e. *touch perception.*

(Perception of temperature, pain as well *gustatory* and *olfactory* signals are are not ergo-relevant as being comparatively structurally shallow, at least in humans.)

But from an ergo-learner perspective, signals differ by how one learns their "meanings", how one interacts with them, how one arrives at understanding of their structures.

1. *Spoken language* depends on the auditory and sensory-motor systems; ears to listen and sensory-motor systems to generate speech. However, deaf-mute people speak in sign language and deafblind people communicate in tactile sign language.[89]

2. *Written language* (whenever it naturally exists) is likely to have a huge overlap with the spoken one in the human brain (of a habitual reader) but it also makes a world of its own. It is not inherently interactive, at least not so superficially[90], and it is not bound to the flow of time. Persistence of written literature is hard to reconcile with a naive selectionist's view on co-evolution of language and the brain.

3. *Mathematics.* Learning mathematics is an interactive process but it is hard to say exactly in what sense.

The images a mathematician generates in his/her mind are neither of Language nor do they belong with any particular "sensory department". Thinking mathematics is like driving an imaginary bicycle or performing/designing a dance with elaborate movements entirely in your head. (This may differ from person to person.)

4. *Languages of games.* We are able to enjoy and to learn a variety of mental and physical games. Probably, these are divided into several (about dozen?) classes depending on how they are incorporated into our erghobrains. Written language and mathematics may be particular classes of games.

5. *Music.* People gifted in music replay melodies in their minds and they can reproduce melodies vocally and/or with musical instruments; the rare few may generate new melodies. But melodies, unlike sentences in a Language, can not talk about themselves and there is no general context where one can formulate what human (unlike that of birds) music is and/or what should be regarded as "understanding of music".

(An avalanche of superlatives that a music lover pours on you when he/she speaks about music tells you something about endorphins release into his/her blood triggered by music but nothing about music related (ergo)structures in his/her brain.[91])

6. *Proprioceptive/somatosensory system.* Running over a rough unpredictable terrain is kind of talking to the road with the muscles in your body. This is much simpler than the ordinary language but is still beyond the ability of computers that control robots. Neither a present day robot is able to hand sew a button on your shirt.

7. *Vision.* At least half of the neocortex in humans is dedicated to vision, but this may be mainly due to the sheer volume of the information that is being processed and stored, rather than to the structural depth of visual images. And amazingly, vision impairment, even vision+hearing impairment, do not significantly affect human ergo. The ergo is robust and independent of particular

[89] Most amazingly, some deafblind people can understand spoken language by picking up the vibrations of the speaker's throat.

[90] Writing and reading is kind of talking to oneself.

[91] Recently, there was an attempt to understand what music does to one's brain: http://phenomena.nationalgeographic.com/2013/04/11/why-does-music-feel-so-good/ and http://www.zlab.mcgill.ca/home.php?1592876871.

sensory inputs.

Three flows among these: *Language, Mathematics, Music* have an essential feature in common: the receiver of such a flow F develops an ability, with no external reinforcement, to creatively generate a new flow F' in the class of F. (In the case of Mathematics and Music this happens rarely, but the miracles of this having happened in the brains of Mozart and Ramanujan outweighs any statistics.)

Modeling the transformation $F \mapsto F'$ is one of the key aspects in our picture of the universal learning problem. (Possibly, there are counterparts of F' for other incoming flows F, but they may be kind of *internal*.)

The most interesting object for the study among these is the learning mechanism of native languages by children that is, probably, similar to how mathematics is learned by mathematicians.

Of course, the structure of a most sophisticated mathematics we build in our minds is by far simpler than that of natural languages (not speak of the vision), but it is still quite interesting, while the corresponding learning process may be more accessible, due, besides its relative simplicity, to a great variance in people's abilities in learning mathematics[92] and a presence of criteria for assessing its understanding.

4.2 Words, Graphs, Categorization and Co-clustering.

There can be no isolated sign. Moreover, signs require at least two Quasi-minds.

CHARLES SANDERS PEIRCE.

Let us assume that a language we study admits a simple general definition of *word-unit* and where we possess a universal rule for identification of *word boundaries*. (In real life defining what is a word and devising an algorithm for identifying them in a flow of signals is by no means easy.)

Let us try to classify words according to their *functions* where two words w_1 and w_2 are regarded *functionally similar* if the other words with which they systematically "cooperate" are themselves tend to be similar.

The condition

w_1 *is similar to* w_2 *if coworkers of* w_1 *are often similar to coworkers of* w_2

may strike you as being circular, but this is easily taken care of by the formal definition below with the apparent circularity making co-clustering mathematically so nice.

What is more difficult is to define and/or identify *togetherness* of "doing something" for pairs (or larger groups) of words. But it is relatively easy to decide, without any reference to "meaning" or "function" whether two given words, say w_1 and w_2, *often* come *close together* or, on the country they come close relatively rarely.[93]

[92] Every sane person understands his/her mother tongue and has an adequate visual picture of the world. This uniformity makes understanding of these "understandings" as difficult as would be understanding motion in the world where all objects moved in the same way.

[93] This preassumes that we know what it means to be *"same"* for words positioned at *different* locations in flows of speech or in written texts.

This gives you what is called the *co-occurrence graph* on the set W of words,[94] where w_1 is joined with w_2 by an edge if the two often come close together, where, moreover, one may vary "often" (measured by a *frequency* an evaluation of which may need some care) and "close" (in some *positional distance*) and thus obtain a family of graphs depending on two parameters.

The remarkable fact is that such graphs, if they come from "real life", have huge redundancy in them – they are very far from anything that can be regarded as "random".

More specifically such a G, typically admits *approximate reductions* to certain much smaller graphs $\underline{G}$.

On terminology.

Division of "objects" into classes is called *categorization* in linguistic and in psychology, while doing this by means of a G is called *co-clustering* in linguistics and *bi-clustering* as well as *two mode clustering* in data mining and in bioinformatics where one says *clusters* rather than of "classes".

This kind of analysis, probably, has been used in other branches of science/statistics under different names that makes it hard to find out when and by whom this idea was originally introduced. (Not impossibly, this was understood and implicitly used by Aristotle.)

Humble Example of Bi-Clustering. Let W consist of letters (kind of) representing phonemes of the English language and let the edges in G represent those pairs of letters that *often* appear *next to each other*, where "often" for (w_1, w_2) signifies that the frequency of this pair is significantly higher than what one would expect from a random sequence of letters. That is

$$prob(w_1, w_2) \geq (1 + S) \cdot prob(w_1) \cdot prob(w_2)$$

in terms of probabilities, where $S > 0$ is a positive constant the specific value of which depends on what "significant" signifies.[95]

Since this typically happens when one of the letters is *vowel* and another one is *consonant*, this G (approximately) "reduces" to the two vertex graph •—• by dividing the vertex set W into two classes/clusters

W=**vo***wels* & **co***nsonants*

vo **co**
•—•

Let us emphasize that this partition of W does not depend on any a priori knowledge of the "nature" of letters, but only on the relative frequencies of letters and pairs of letters in texts; the idea of *meaning* we attribute to these classes comes along with the *names* we assign to them.

To define graph reduction in general, it is convenient to think of G as a $\{0,1\}$-function on the vertex set (of words) W of G, written as $G(w_1, w_2)$, where "reduction" is a representation of G as a composition, sometimes called *superposition*, of a surjective (i.e. onto) *reduction map* $R : W \to V$ for some set V, usually significantly smaller than W, and a $\{0,1\}$-function on V, say $\underline{G}(v_1, v_2)$,

$$G(w_1, w_2) = \underline{G}(R(w_1), R(w_2)).$$

[94] We assume here that words constitute *sets*.

[95] Being vague here poses no danger of lulling ourselves into a false sense of understanding, as it frequently happens to people carrying out speculative discussions with their intuition unaided by mathematics.

It is unrealistic to expect the existence of such "perfect reduction" if V is much smaller than W. All we may hope for is a *good approximation* of $G(w_1, w_2)$ by $\underline{G}(R(w_1), R(w_2))$ for some R and $\underline{G}$, where "approximation" (usually in this context) means that the above equality holds for a "significant majority" of the pairs $(w_1, w_2) \in W \times W$ that correspond to closeness in the l_1-*metric* in the space of real functions on the set $W \times W$.[96]

To get a rough idea of a possible magnitude and efficiency of such a reduction let the set W contain 200 000 of (generously understood) "words", including common *di-grams* (pairs of words). Then, in general, the description of G needs about

$(2 \cdot 10^5)^2 = 40\ 000\ 000\ 000$ – *forty billion* – bits of information.

This is a pretty big number, you can not learn that much during your lifetime of three billion seconds.

On the other hand, if you reduce W to a set V with, say 300 elements – classes of words – in it, then the pair $(R, \underline{G})$ can be encoded with only

$(\log_2 300) \cdot 200\ 000 + 300^2 < 2\ 000\ 000$ – *two million* bits.

This is more than *twenty thousand-fold* reduction of information!

It is the (ergo)-grammar, of course, not texts themselves written in our language, that may admit such incredible compression. (The maximal compression of texts is believed to be well below *ten*.) But this is exactly what ergo-learning is about: it is not remembering whatever enters you brain, but classifying, forgetting, compressing and further structuralizing the incoming "information".

Now there are three questions that beg for answers.

1. What do you gain by this compression of information if you need to know *all of* G to start with in order to construct R and $\underline{G}$?

2. Even if you are given G, say written down in your computer memory, how can you find a reduction in practice?

3. If there are several different "reasonably looking" reductions, how can you trust any one of them?

Ansver to 1. In order to find R and $\underline{G}$ you do not need to know all of G but only a part of it that carries somewhat more than 2 000 000 bits fn information. For example, a child who learns a language may recognize, in the course of first fifteen years of his/her life, with 20% of this time being exposed to the language (20% of 15 years make about hundred million seconds), ten million pairs of words (w_1, w_2) frequently (decided according to some criterion) coming together.

You extend $G(w_1, w_2) = 1$ at these pairs by zero everywhere else on $W \times W$ and search for a reduction for this incomplete version of G to a 300×300 function $\underline{G}$.

Answer to 2. In general, there is no simple and fast *co-clustering algorithm* for finding a reduction, but certain (some are not fully understood) features of human languages make such algorithms possible.[97]

For example, a presence of small body of *core words* that have exceptionally high frequencies makes naive iteration algorithms quite efficient.

[96]There is much to be done in order to make this more specific or, on the contrary, more general, e.g. by using other metrics on spaces of functions in *two* variables.

[97]Co-clustering programs must be present in the image processing systems of animals (all vertebrates?) and these kind of ready made algorithms in our brain had directed the path taken by the evolution of the "language instinct" in Homo sapience.

Answer to 3. The existence of a sufficiently sharp approximate reduction, i.e. with $\underline{G}(R(w_1), R(w_2))$ sufficiently close to $G(w_1, w_2)$, is rather exceptional and miracles do not happen twice: if there are, say, two such reductions $R_1 : W \to V_1$ and $R_2 : W \to V_1$, where the set V_2 has cardinality $card(V_2) \leq card(V_1)$, then, most likely R_2 equals a reduction of R_1.

This means, there exists a reduction $R_{12} : V_1 \to V_2$ of the graph $\underline{G}_1$ to $\underline{G}_2$, such that R_2 equals the composition of the two maps, $R_2 = R_{12} \circ R_1$, i.e. $R_2(w) = R_{12}(R_1(w))$, or at least R_2 is "quite close" to $R_{12} \circ R_1$.

All of the above being said, a doubt may linger in you mind.

Isn't this $\underline{G}(R(w_1), R(w_2))$ too simple to teach you anything substantial about learning and understanding?

Do you need mathematics to express the idea of two words being similar if they have similar surroundings in texts?

Prior to responding to this let us make it clear that the above kind of co-clustering is neither the final product of building a structure from "flows of words'" nor is it an "atomic unit" of such a structure.

One rather should picture it as a large molecule with simple, yet, non-trivial, internal architecture where this molecule, in turn, serves as a building block for more elaborate syntactic structures.

The simplicity of this "mathematical molecule" makes it quite versatile: one can modify it in many ways and adjust it to building a variety of different global structures.

For instance:

• One may apply co-clustering to functions in more than two variables (this is why we prefer "co-clustering" to "bi-clustering") where these functions may take values in more interesting sets than $\{0, 1\}$.

• Instead of a single reduction one may bring forth *diagrams* of several of them, such as the above $W \to V_1 \to V_2$, or something combinatorially more elaborate and interesting than that.

• On may extract more subtle and/or more substantial information about the structure of a language by looking closer at the geometry of the set of words W with respect to the metric (distance) induced from the space $\mathcal{F}$ of functions $f = f(w)$ on W, for the (tautological) imbedding of W to $\mathcal{F}$ defined as in the previous section for the semiosis graph:

a word $w_0 \in W$ is assigned $f_{w_0} \in \mathcal{F}$, such that $f_{w_0}(w) = G(w_0, w)$.

Here, unlike how it was with the semiosis graph, the geometry of $(W, dist_{\mathcal{F}})$ for the metric induced from the space $\mathcal{F}$ (with a suitable Hammnig-like metric on it) is significant.

For example, co-clustering can be seen in the light of this geometry as an ordinary (mono)clustering of $(W, dist_{\mathcal{F}})$ into "loosely connected pieces" with respect to $dist_{\mathcal{F}}$.

On the other hand, the geometry of $(W, dist_{\mathcal{F}})$ also suggests another classification of words $w \in W$, namely, according to the geometry of small balls around them in $(W, dist_{\mathcal{F}})$.[98]

[98]Different prepositions in English, e.g. *under* and *over*, may be accompanied by *different* kinds of words, say, nouns and/or verbs; yet the geometries/combinatorics of the balls around them in $(W, dist_{\mathcal{F}})$ look, nevertheless, quite *similar* that may serve as in indicator of the two belonging to the same class (cluster).

And this classification is not the end of the story – many structural features of "flows of signals" are encoded by (not-quite) *pseudogroups* of approximate isometries of spaces like $(W, dist_{\mathcal{F}})$.

One may continue indefinitely along these lines but one has to stop somewhere. Wings of imagination supplied by the power of mathematics can bring you beyond of whatever can be reached by a more pedestrian kind of thinking. But if you fly too high in the sky of math you may miss your destination down on Earth.

4.3 Similarity, Co-functionality, Reduction.

Let us make a short (and incomplete) list of four "logically (quasi)atomic constituents" of (ergo)operations applied to flows of signals, in particular of those we used for co-clustering. But we do not attempt at the present point to give precise definitions of these "atoms", to justify their reality, and/or to explain how one finds them in flows of signals.

1. Segmentation and Parsing. The first step in structuralizing flows of signals is identifying/isolating *units* in these flows, where the simplest (but not at all simple) process serving this purpose is *segmentation*: dividing a flow into non-overlapping "geometrically simple" parts.

These may be small and frequently appearing signals, such as phonemes, words and short phrases in the flow of speech or basic visual patterns such as *edges* and *T-junctions*. But these may be as long as sentences, internet pages, chapters in books or intrinsically coordinated visual images of such objects as animals, trees, forests, buildings, mountains.

Our formalism must apply to general "flows of discretely discernible units" where the learning consists in building structures out of "internal units" that may be similar or dissimilar to the units of incoming flows.

Naively, *unit* is anything that can be given a name and/or characterized in a few simple words, but... these words may be of very different kinds depending not only on the intrinsic properties of such a unit, but also on how it is being processed by a particular ergosystem, e.g. a human ergobrain.

Sentences, words, morphemes, graphemes (letters) are all units but they belong to different categories, where, "macro-units" such as sentences come as "parametric families of a kind" or as "formulas with free micro-unit variables in them".[99]

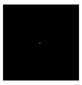
Similarly, elaborate paintings and simple figures both may regarded as units but they are unlikely to be filed by your visual ergosystem in the same "units-directory".

(Processing of linguistic and visual inputs by your (ergo)brain, probably, relies on natural parsing of incoming flows of signals followed by a combinatorial organization of the resulting "units".

But *proprioception sensory system*[100] and motor control of skeletal muscles may also depend on *continuity*, since the incoming signals may be not(?) naturally decomposable into "discrete units".)

[99]Straightforward attempts to make this precise may confine you to the Procrustean bed of the traditional mathematical/logical language.

[100]This is the perception of motion, of stresses and of position of parts of the body.



2. Similarity, Equivalence, Equality, Sameness. There are several *similarity relations* between units of languages/images where these relations may differ in kind and in strength.

For example, images may be similar in shape, size. color, subjects they depict, etc. while two sentences may be similar in the kind and style of words they employ, the idea they convey or in their syntax. The strongest similarities in texts *are letter-wise equalities* of *different* strings.

There is a discrepancy between how the concept of *equality* is treated in mathematics/logic and in in natural languages: we happily say:

2+3 equals 5

but:

5 equals 5

appears non very informative even to a most logically minded mathematician – these two *"equal"* are not mutually equal and the common language has no means to express this inequality. For example, if you try

5 is the the same as 5

this does not make it to look better. But this can be settled if we introduce an ergosystem in the picture, where equalities as well as weaker similarities result from certain *processes*, that are qualitatively different from how one arrives at *sameness.*[101] (We discuss some of this in our [SLE]-paper.)

On Composability of Similarities. Customary, one defines an *equivalence* as a *symmetric binary relations* on a set[102] S, denoted, say by $s_1 \sim s_2$, that satisfies the transitivity property:

$$s_1 \sim s_2 \ \& \ s_2 \sim s_3 \Rightarrow s_2 \sim s_3.$$

It is more convenient to depict equivalences (and similarities) of signals s in a category theoretic style by arrows with "names" attached to them, such as $s_1 \overset{f}{\leftrightsquigarrow} s_2$, where one think of such an arrow as an "*implementation of* ∼" by some "logical/computational process", e.g. by some co-clustering algorithm.

Then one may compose arrows

$$s_1 \overset{f}{\leftrightsquigarrow} s_2 \overset{g}{\leftrightsquigarrow} s_3 \text{ with the composition denoted } s_1 \overset{f\square g}{\leftrightsquigarrow} s_3.$$

This allows one, for instance, to say that

[101] The spirit of this is close to how different levels of "equivalence" are treated in *the n-category theory*.

[102] This definition does not cover equivalencies between theories and/or between categories since these are *are not* relations on *sets*.

the composition $f \circ g$ of two "strong similarities" f and g is itself a "weak similarity".

Also one can now speak of certain equivalencies f and g, e.g. one in color and another one in size, being *incomposable*.

3. Classification, Clustering, Reduction. Equivalence relations E on a set S go hand in hand with partitions of this S into the corresponding *equivalence classes $c \subset S$ of E,* where, in turn, such partitions are "*essentially the same*" as *reduction maps* $R = R_E$ from S *onto* sets C, and where, conversely, such a map R defines an equivalence relation $E = E_R$ by

$$s_1 \underset{E}{\sim} s_2 \text{ if and only if } R(s_1) = R(s_2).$$

However, implementations of binary relations $s_1 \underset{E}{\sim} s_2$ and of unary operations $R(s)$ are quite different from a working ergosystem point of view.[103] For instance, it is much harder to record $\approx N^2$ bits encoding an equivalence relation on a set S of cardinality N, than $\approx N \log N$ bits needed for defining $R(s)$, where N may be somewhere between 10^4 and 10^7. Because of this, *similarities* and *reductions* must be treated separately.

An essential feature of *reductions* from our perspective is *compression of information* and

"creation" of new units c from the original unites s, that are $c = R(s)$.

A more general and less cleanly defined class of operations is called *clustering* that is based on similarities that are not sharply defined and are not perfectly *transitive* unlike what is usually required of "equivalence".

The tautological map $R : s \mapsto c$ associated to a given clustering that assigns to each member s of S the cluster c in S that contains s (this R may be defined not for all s) is still called the *quotient map* or *reduction* from the original set S to the set C of clusters. The reduction that defines co-clustering is an instance of this.

Compression, Morphisms, Functors. Besides the above, there are reductions of quite different type that correspond to "non-local" *compression of texts with limited loss of information*, where one forgets non-essential in a text (or in a visual image) while preserving the significant structure/content of it; this is a hallmark of understanding.

It may happen, of course, that a text has little redundancy in it, such as a telephone directory, for instance. Then no significant reduction and no understanding of such text is possible.

In fact, "perfect texts" with no redundancy in them are indistinguishable from random sequences of symbols, while every meaningful text T admits many reductions, depicted by arrows, say $T \overset{r'}{\to} T'$, $T \overset{r''}{\to} T''$, where the bulk of the the process of understanding a text consists of a multi-branched cascade of such reductions.

An example of a significant commonly used reduction is making a *resume* or *summary* of a text. Also giving a *title* is an instance of a reduction – a *terminal reduction*: you can not reduce it any further.

If we agree/assume/observe that consecutive performance of reductions, say $T_1 \overset{r_{12}}{\to} T_2$ and $T_2 \overset{r_{23}}{\to} T_3$, make a reduction again, denoted $T_1 \overset{r_{13}}{\to} T_3$, also written

[103] This is discussed at length in the context of cognitive linguistics by George Lakoff in *"Fire Women and Dangerous Things"* where *classification* is called *categorization*.

as *composition*

$$r_{13} = r_{12} \circ r_{23},$$

then reductions between texts can be regarded as *morphisms*, of the **category** (in the mathematical sense) of *texts and reductions* where, strictly speaking the word *"reduction"* suggests these arrow r being *epimorphisms*, i.e. they add *no new information* to texts they apply.

It may be amusing to encode much (all?) information about a language $\mathcal{L}$ – *syntax, semantics, pragmatics,* in terms of such a category $\mathcal{R} = \mathcal{R}(\mathcal{L})$ of reductions in $\mathcal{L}$, with translations from one language to another, $\mathcal{L}_1 \rightsquigarrow \mathcal{L}_2$, being seen as *functors* between these categories; but we shall not try to force categories into languages at this point.

Reduction and Agglomeration of Similarities. There are circular relationships between similarities of different types and/or of different strengths. For instance two signals s_1 and s_2 that have equivalent or just strongly similar reductions may be regarded as weakly similar.

Conversely, if there are "many independent" weak similarity relations between s_1 and s_2 then s_1 and s_2 are strongly similar and possibly, equal. For instance if the numbers N_i and N_i' of the letters on the ith pages of two books B and B', say with 200 pages each, satisfy $N_i \neq 0$ and $|N_i - N_i'| \leq 2$ for somehow chosen hundred numbers i, then one can bet that B and B' are copies of the same book.

4. Co-functionality. Some units in a text T or in another kind of flow of signals form relatively tightly knit groups where we say that these units *perform a common function*.

A priori, co-functionality is not a binary relation (albeit we assumed so defining co-clustering in the previous section); it can be, however, made binary by give "names" to these *"functions"* and by regarding functions as new kind of units.

Then we say that *unit s performs function f* and depict this by a directed edge $s \longleftarrow f$. Alternatively, we depict f-co-functional units as being joined by f-colored edges $s_1 \underset{f}{\longleftrightarrow} s_2$.

4.4 Transformations of Incoming Flows.

Learning to Read by Learning to Speak.

The original form of signals carried by the above seven flows is different from what arrives at your sensory systems. For instance, visual images result from 2D *projections* of *three dimensional* patterns to the retina in your eyes; moreover, brain's analysis of these projections is coupled with the activity of the brain's motor system that controls movements of the eyes that continuously modify these projections.

Similarly, the flow of speech as it is being generated in one's mind is, according to the tenets of *generative grammar* has a tree-like structure that is then "packed" into single time line.

Reconstruction of F_{orig} from the flow F you receive is an essential and most difficult aspect of understanding the message carried by this F. For example, understanding a flow F of speech is coupled with one's ability to speak, i.e. to reconstruct/generate F_{orig}, or something close to it, in one's ergobrain.

The only aspect of this reconstruction we shall discuss is what can be expressed as an *annotation* to F.

For instance, upon receiving a flat image F on its screen (retina), an ergo learner $\mathcal{L}$ must correctly resolve depth in *interpositions/occlusions* and/or "guess" relative values of the third coordinates at essential points of F.

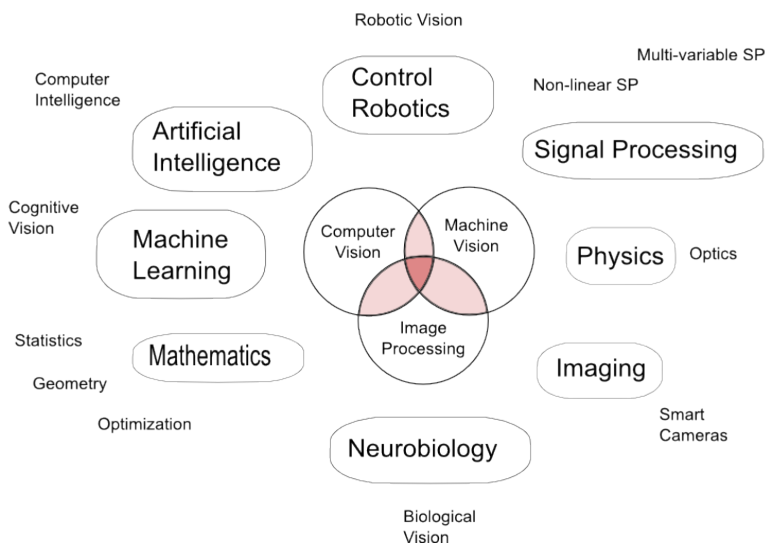
And the background tree structure in a (record of a) flow F of speech can be indicated with *parentheses* properly inserted into F. (An annotation may also include additional syntactic and/or semantic comments concerning particular words and sentences.)

Such annotations performed by a human ergobrain depend on an elaborate guesswork that is by no means simple or automatic and it is still poorly understood. And besides annotating flows of signals, the ergobrain augments them by something else.

For instance, formation of a visual image in one's mind depends on the activity of motor neurons involved in eye movements and "understanding" of these images depends on structural matching this activity with similar actions of these neurons in the past.

This active process of perception can be seen as a conversation or a kind of a game of the ergobrain with the environment. But such games, unlike anything like chess, are not easy to mathematically formalize.

4.5 Facets of Learning.



There are, in a rough outline, two actively pursued directions in the study of learning:

(1) The study of the neuro-physiological *mechanisms* of learning and computer design of (conjecturally) similar mechanisms.

(2) Description of various *specific* learning *problems*, a study of how are these are *solved* by humans and/or animals and an algorithmic approach to their *solution*.

To have just a glimpse of an idea of what goes on look at the corresponding pages in Wikipedia:

"learning", "memory", "motor learning", "language acquisition",
"mental representation" "outline of artificial intelligence",
"machine learning", "computer vision".

We do not attempt to contribute anything either to (1) or to (2) but we shall try to look at *learning* from a different perspective.

Our first objective is to identify a maximally general (quasi)mathematical concept reflecting essential features of learning processes.

Such a process must apply to an abstractly defined class of "incoming flows of signals", denote these by $[In]$, like those entering the brain via sensory receptor cells; then learning is seen as some kind of transformation applied to these flows. The results of such a transformation are twofold:

(A) The "visible" part of such a transformation is an outgoing flow of signals, we denote this flow by $[Out]$. The basic example of it is what goes from the brain to the muscles of the body.[104] This flow modulates an interaction – think of this as a "conversation" or a "game" – of the brain with the external world.

(B) What is invisible is incorporation of results of the $[In] \to [Out]$ transformation in the internal structure of the learning system. Building this (ergo) structure, call it $[ErgSt]$, constitutes the major part of the activity of the "brain who learns", but this process is invisible[105]; hopefully, we may guess how this $[ErgSt]$ looks like by imagining how a mathematician would proceed in making such a structure.

However, even the apparently easy problem of developing a language for speaking of incoming flows $[In]$ is by no means trivial since

on the one hand, we want to describe the incoming signals in mathematical terms with no reference to the external world where they originate from;

on the other hand, we want to keep track of the "real world meaning" of these flows.

We shall follow the usual recipe for solving this dilemma by resorting to *doublespeak*: we shall manipulate with $[In]$ as with an abstract mathematical entity but shall speak of it in metaphorical terms as if it were still residing in the "real world".

An essential issue in artificial learning, as we see it, is finding mathematical means for description of $[ErgSt]$. Without understanding this structure, attempts to reconstruct the transformation $[In] \to [Out]$ are like planning a trip to the Moon with no idea of rocket propulsion in your mind.

4.6 Frozen Flows.

We shall not attempt to approach any realistic learning problem and/or structuralization of general "flows of information" but we shall look at a model problem of learning an *imaginary non-human language*. We assume that one has a vast library of texts and may use a computer to analyses these but where

[104]What we know of the structure/message carried by $[Out]$ is mainly manifested by the (broadly understood) *behavior* of an organism.

[105]This is similar to *metabolism* versus *digestion*: the product of the latter is visible without being especially interesting. But the energy transfers and biochemical building processes in the cells are not discernible to the causal eye, but this is what we find fascinatingly interesting.

one has not even a whiff of knowledge of the semantic and the function of this language. It may be not even known beforehand whether this is a language or a digitalized record of music or of a visual image.

To see things in perspective, let us look at *freezing and putting on record* non-linguistic "flows of signals."

Visual and auditory signals in the outside world are "written" on the rigid space-time (x,t)-*background*. The stationary images "carved " on the x-space are most essential in human vision[106] while hearing is mostly associated with the time t-coordinate[107].

Since stationary landscapes are commonly seen in life, visual images are, psychologically speaking, the easiest to "freeze". And technology for doing this have been existing for, probably, more than 50 000 years.[108]

The idea of "frozen sound" is not so obvious: we often see snapshots of beautiful sceneries frozen in time but we hear no interesting stationary sounds. (But light is harder to understand than sound: the idea of sound waves goes back to Aristotle while the idea of light being of waves had to wait another two thousand years.)

The first sound recorder was, probably, *phonautograph* (1857) of Scott de Martinville, while the idea of inversion of this recording was suggested in 1877 by Charles Cros, and implemented in *phonograph* by Thomas Edison in 1878; this allowed historically the first recording of human voice.[109] But freezing and recording speech in writing goes more than 5000 years back.

No Ergo in the Nose.

Humans can distinguish about 10 000 different scents. We have about thousand of different kinds of *olfactory receptors* in our nose that are proteins[110] where each kind of proteins is coded by a particular gene.

But the internal library of smells has, apparently, no ergo in its architecture being organized simpler (this is seen not only by introspection) than how we remember visual images, sounds, words and ideas.

[106] Vision of many animals is more dx/dt dependent than ours. For instance, frogs seem to respond only to moving objects.

[107] People with fine hearing (often with impaired vision) distinguish some spacial features around them similarly to animals with acute hearing, such as owls, for instance. But the best are *echolocating animals* – bats, dolphins, porpoises, some whales and certain birds, e.g *South American oilbirds*.

[108] No known terrestrial or aquatic animal and no *randomly taken* (untrained) contemporary human is able adequately/artistically record visual images in paintings or otherwise. But this could have been different with Neanderthals and/or Cro-Magnon people.

[109] Some people and birds are good as remembering and imitating strings of sounds; also there are records of 9th century mechanical music playing devices invented by Banu Musa brothers.

[110] A single class of receptors may bind a range of odor molecules, and same kind of molecules may bind to several different receptors.

Scents, unlike images and sounds that can be written and rewritten on a variety of backgrounds, are supported by *specific to each of them* physical/chemical substances. No, rewriting, no simple digitalization, no universally organized library of smells comparable to that of images or sounds seem possible.

Since olfaction, unlike vision, does not depend on your muscles, it is disconnected from proprioception system and we have no means of (re)producing scents at will, albeit we we think we can recall them.

There aren't many clearly identifiable *universal smells* common to large groups of object. Non-surprisingly, languages (of urban populations?) have few specific names for smells — about ten in English:
musky, putrid, rotten, floral, fruity, citrus, vegetative, woody, herbaceous, spicy.
(There are slightly more smell-words in certain languages, for example there are about fifteen of them in the *Kapsiki of Cameroon.*)

Natural languages do not waste words for naming *individual* objects/properties but rather exercise the art of giving the same name to *many* different things, very much as mathematical theories do. There is no grammar of scents, no books in the language of odors, no "frozen" flows of olfactory perception.[111]

Frozen Assymetries.

Digital arrays on a magnetic tape have, physically speaking, little in common with flows of real sounds and/or images but recording preserves how spacial/temporal symmetries are broken by these flows.

"Ideal tapes" where signals are recorded make by themselves, without recordings on them, a rather "symmetric community":

a *directed tape* segment T_1 of length l_1 can be placed into a segment T_2 of length l_2 whenever $l_2 \geq l_1$, where every such direction preserving placement $\tau : T_1 \to T_2$ where τ is identified with a number $\tau \leq l_2 - l_1$ that is the distance between the left ends, t_2 of T_2 and t_1 of T_1 inside T_2.

|——τ—— |-----------------|——————|
t_1 t_2

Such placements, say $T_1 \to T_2$ and $T_2 \to T_3$, can be obviously composed with the result being a placement $T_1 \to T_3$; thus, placements make a (rather trivial)
category in mathematical sense, denoted $\mathcal{PT}$
that is is closely (and obviously) related to the additive group of real numbers.

Now, if T_1 and T_2 record something encoded by strings of symbols, then there are distinguished placements, called *matchings* $T_1 \to T_2$ that *agree* with what is recorded in there. For instance if T_1 with $length(T_1) = 5$ caries the "word" ⊡⊟⊞⊠⊟ then it admits exactly three (rather than 50) placements[112] into T_2 of $length(T_2) = 55$ with the following string in the four symbol-letters: ⊡, ⊟, ⊞, ⊠:
⊡⊟⊠⊡⊟⊞⊠⊟⊞⊠⊟⊡⊟⊡⊡⊟⊞⊡⊠⊠⊟⊡⊟⊞⊠⊟⊡⊞⊟⊞⊟⊡⊠⊡⊡⊟⊞⊠⊟⊟⊞⊞⊡⊡⊟⊡⊟⊠⊟⊡⊠⊟⊡⊡⊟
written on it.[113]

---

[111]Perfumes do not count.

[112]In the fully randomly equidistributed case the probability of this "three" would be $(50/4^5)^3 \approx 0.008$ that is less than 1%, but *meaningful* flows of signals are NEVER stochastically equidistributed.

[113]Searching for ⊡⊟⊞⊠⊟ in ⊡⊟⊠⊡⊟⊞⊠⊟⊞⊠... requires a concentrated effort on your part but the "equivalent" problem of locating the familiar pattern *pasta* (stored in your long-term memory) in
patpastastapappaspttapastapsasaptppastaassppapataptappa,
is solved by your visual system semiautomatically with rapid, 20-50ms, saccadic movements

Matching records constitute a *subcategory* of $\mathcal{PT}$, denoted $\mathcal{MR}$. This $\mathcal{MR}$ carries the same information as the records themselves, and it has the advantages of being insensitive to the nature and to the size of your "alphabet".

However, "alphabets" often carry significant structures of their own; such a structure may be lost[114] unless it is incorporated into the category theoretic setting.

Besides, perfect matching is too restrictive when it comes to real life signals where one should deal with "approximate matchings" and other (sharp and approximate) category theoretic arrow-morphisms, such as *similarities* and/or *reductions* in languages.

Such "approximate morphisms" are *not always composable* and the category theoretic language must be augmented with "certainty weights" assigned to compositions of arrows.

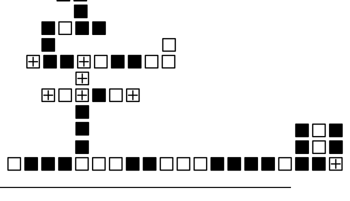
On Alignment of Sequences. Approximate matching is mathematically similar to how bioinformaticians compare sequences of DNA, RNA and proteins where these sequences are structurally different from what one encounters, say, in "flows" of spoken and/or written human languages. Also, one practices *multiple alignments* of k-tuples of sequences, for $k > 2$ that are kind of "integral curves" of alignment (similarity) "equations" in Cartesian products $T_1 \times T_2 \times T_3 \times ... \times T_k$.[115]

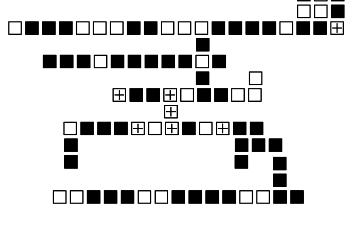
Maximal matching in vision.

A category theoretic style "matching records" description also applies to visual images "engraved" on spacial+temporal domains S that may be taken of dimensions $d = 2, 3, 4$.

If $d > 1$ it is impractical to "match records" for all kind of domains S, since there are two many different shapes of them but one can keep track of *partial matchings* between S_1 and S_2 e.g. of *maximal connected* subsets, say $S_1' \subset S_1$ and $S_2' \subset S_2$, where the records match by a map from some class of transformations with a given precision threshold and where one concentrates only on sufficiently informative/representative maximal matching pairs S_1' and S_2' rather independently of the ambient $S_1 \supset S_1$ and $S_2 \supset S_2$.

Pattern matching, is, probably, the most essential class of operations performed by our visual ergo-system (look at the two □-■ figures below) but the mathematical (probably, neuronal as well) mechanisms implementing these matchings remain unknown.

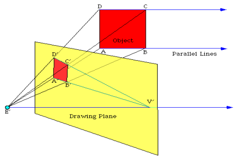




of the eyes.

[114]Often structures in alphabets, e.g. the options of upper/lower cases in many languages, can be eventually reconstructed from the global structures of texts.

[115]It is questionable if human ergo unaided by mathematical+biological knowledge could master such alignments.



Ergo-irrelevant(?) *Definitions of Geometric Transformations.* If $d > 1$, the counterpart of the above $\mathcal{PT}$ may be taken by a category $\mathcal{PS}$ for domains S in the Euclidean space $\mathbb{R}^d$ and some class of geometric "placements" that are transformations (maps) $P : S_1 \to S_2$, where relevant classes of placements are associated with the following transformations of Euclidean spaces.

$\bullet_{tra}$ *translations* of $\mathbb{R}^d$, where $x \mapsto x + p$ for all $x \in \mathbb{R}^d$ and some $p \in \mathbb{R}^d$ (which brings us back to $\mathcal{PT}$ for $d = 1$);

$\bullet_{iso}$ *isometries* that are *translations* & *rotations*;

$\bullet_{sim}$ *similarity transformations* that are *translations* & *rotations* & *homotheties*, where the latter are maps $x = (x_1, x_2, ..., x_d) \mapsto \lambda x = (\lambda x_1, \lambda x_2, ..., \lambda x_d)$ for some number λ (sometimes assumed > 0);

$\bullet_{aff}$ *affine transformations* where translations & rotations may be additionally composed with *anisotropic scaling maps* $(x_1, x_2, ...x_d) \mapsto (\lambda_1 x_1, \lambda_2 x_2 ... \lambda_d x_d)$;

$\bullet_{pro}$ *projective transformations* between domains $S_1, S_2 \subset \mathbb{R}^d$ that, by definition, send *straight linear segments* from S_1 to such segments in S_2.

Affine transformations of $\mathbb{R}^d$ are instances of these and, although it is counterintuitive, there are lots of *non-affine projective* transformations, that are, however, *never defined on all* of $\mathbb{R}^d$.

Non-affine projective equivalence is not (?) detectable by the human visual system, but projective transformations are used to create perspective in drawings. In fact, such a transformation for $d = 2$ can be seen as a *radial projection map* between planar regions S_1 and S_2 in the 3-space.

Namely, if p is a point in $\mathbb{R}^3$ away from the planes containing these regions then the radial projection, say $Pr_p : S_1 \to S_2$, is defined, provided the straight lines between p and all $s_1 \in S_1$ *meet* S_2, say at

$$s_2 = s_2(s_1, p) =_{def} Pr_p(s_1).$$

Notice that such a projection is an affine (in fact, similarity) transformation, if S_1 and S_2 lie in *parallel* planes.

All these transformations, even parallel translations of $\mathbb{R}^d$ for $d > 2$ unlike translation of $\mathbb{R} = \mathbb{R}^1$, do not admit natural discretizations and the groups of these transformations are non-commutative.

It remains unclear how the human visual system developes internal models of these groups, where the coupling with the somatosensory and motor systems is, probably, essential, at least for the isometry group, as it was suggested by Poincaré (see section 2.4).

Libraries of Moves.

There are no *external* libraries storing sequences of positions, velocities and accelerations/stresses of parts of your body along with memories of impulses sent by motor neurons to your skeletal muscles in these positions, but each of us has such a library in the brain.[116]

Structural organization of this information kept by your proprioception system, e.g. associated with locomotion is different from how strings of words are stored inside and outside our brains, where each "proprioceptive word-move"[117] comes from a multidimensional space. (For instance, fifteen finger joints on your hand contribute twenty active degrees of freedom to the positional part of this space.)

Only members of a tiny *sample subset* $\mathcal{S}$ of the full "space of motions" $\mathcal{M}$ of your body can be experienced during your lifetime and stored in your brain, with an essential part of the "neuronal motor memory" being accounted for by the temporal/sequential organization of $S \in \mathcal{S}$.

Apparently, this *discrete* sample set $\mathcal{S}$ of motion mathematically naturally extrapolates to a *continuous hierarchically organized* "manifold" $\Sigma \subset \mathcal{M}$ decomposable into simple(?) low dimensional blocks, where the geometric structure of this block structure of Σ admits experimental/observational study. However, a mathematical (ergo) model of Σ is indispensable for understanding this structure and/or for design of agile robots.

The tactile systems involved in handling objects and perproprioception, both coupled with the motor control system, are the most "active" of your perception systems[118] but all perceptions are far from being entirely passive.

For instance, formation of visual images in you brain depends on elaborate muscular movements of the eye that explores its field of view. (E.g. generation of visual images in dreams is associated with these movements.) If this exploration is interfered with, you visual perception suffers.

(If a lecturer blocks your field of vision on the screen and displays one word after another rather than showing longish strings of words, these words do not integrate into meaningful sentences in your mind. Many of us have suffered through such lectures.

But, amazingly, one can comprehend spoken language by apparently passive process of listening, although it is hard to listen to somebody else speaking without opening one's own mouth every so often.)

5 Ergo Thinking, Ergo Learning and Ergo Behaviour.

5.1 Memory, and Prediction.

Are we Smarter than Apes? Imagine you are subject to a psychology study by an alien who brought you from earth, kept without food for 12 h and put naked into a lighted cubical $3m \times 3m \times 3m$ with perfectly smooth floor, walls and ceiling. There is a cylindrical 1m long stick 1cm thick in the room and a

[116] The nature and mechanisms of human and animal memory are still shrouded in cloaks of mystery, but we know that even the long-term memory is fluid rather than frozen.

[117] These "words" are kind of moves in the game your body plays with the surrounding space.

[118] Auditory systems of echolocating animals are (at least) as "active" and as elaborate as perproprioception. Designing an (ergo)system with echolocating + auditory abilities of a bat, not to speak of a bat's flying agility, remains a robotist's dream.

banana attached to the ceiling. Also there is a string hanging from banana with a loop at the hight 2m.

What shall you do to demonstrate your "sapience", knowing that if you fail you will be treated no better than how we human treat animals. (See the answer at the end of 3.2.)

5.2 Ergo Moves of a Bug on a Leaf.

Think of a bug crawling on a leaf or of your eye inspecting a *green* spot on, say, a *brown* background.

We assume (being unjust to bugs) that all the bug can directly perceive in its environment are two "letters" a_1 and a_2 – the colors (textures if you wish) of its positions on the leaf, where the bug has no idea of color (or texture) but it can distinguish, say, green a_1-locations from the brown a_2.

The four "words" our bug (eye) creates/observes on contemplating the meaning of its two consecutive positions are a_1a_1, a_1a_2, a_2a_1 and a_2a_2. But can the bug tell a_2a_1 from a_1a_2? Which of these words are similar and which are not? Is a_1a_1 is "more similar" to a_1a_2 than to a_2a_2?

These questions are, essentially, group theoretic ones. (Tell this to a bug, and it would as much surprised as Monsieur Jourdain upon learning he spoke prose.)

There are two distinguished (commuting involutive) transformations acting on the set of these four words:

1.*Alphabetic symmetry*: Switching the colors, $a_1a_1 \leftrightarrow a_2a_2$ and $a_1a_2 \leftrightarrow a_2a_1$.

2. *Positional symmetry*: Interchanging the orders of the letters, $a_1a_2 \leftrightarrow a_2a_1$, with no apparent action on a_1a_1 and a_2a_2.

(Recall, that a transformation T is an involution if $T \circ T(x) = x$ for all x, and T_1 commutes with T_2 if $T_1 \circ T_2 = T_2 \circ T_1$, where $T_1 \circ T_2$ denotes the composition: $T_1 \circ T_2(x) =_{def} T_1(T_2(x))$.)

(If you tried to explain this to a bug, it would be much annoyed since:

1. Traversing a leaf is an *art* rooted in a *holistic intuition* of insects which can not be reduced to the abstract nonsense of algebra.

2. Insects have been managing pretty well for the last 300 million years, while the adepts of abstract algebra can not boast a comparably venerable past, no do they have much future on the evolutionary time scale.

3. The group theory is *not useful* in serious matters such as eating a tasty leaf, for a example.

Amusingly, some mathematicians also insist on the superiority of "concrete", "pragmatic", "artistic" and "holistic" approaches to their beloved science. This, however, is not supported by the historical evidence: "artful mathematical pieces" are eventually become incorporated into "abstract" general theories, or else, they become extinct. Thus, for example, the *Wilf-Zeilberger algorithm* relocated exquisite flowers of *hypergeometric identities* form the rain forest of artistic imagination to the dry land of formal reasoning.)

The above transformations do not essentially change the "meaning" of the words: a green square on a brown background is identical, for most purposes, to a brown square on a green background. (I wonder, how fast this "equality" is recognized by different kinds of animals.) Also, the "essence" of a_1a_2 is the

change of colors as you go from one location to the other, rather than the colors themselves.

(There is yet another involutive transformation: changing the color of the first letter : $a_1a_1 \leftrightarrow a_2a_1$, $a_1a_2 \leftrightarrow a_2a_2$, $a_2a_1 \leftrightarrow a_1a_1$ and $a_2a_2 \leftrightarrow a_1a_2$. This involution together with the positional one generate a non-commutative group with 8 elements in it, called the *wreath product* $\mathbb{Z}_2 \wr \mathbb{Z}_2$; the role this group plays in the life of insects remains obscure.)

Alphabet of Bug's Moves. Our bug (or eye) has a certain repertoire of moves but it knows as little about them as it knows about colors. Imagine, each move being represented by a letter, a "button" the bug may press. As a result, the bug (eye) sees the color of the location this move/button brings it to.

The basic knowledge the bug possesses is the equality/non-equality relation between two moves. Since the bug does not know its own position, the *equality of two moves* from bug's perspective – pressing the *same button* – implies that

the two spacial moves go in the *same direction* (relative to bug's orientation) and at the *same* distance, *regardless of the location* of the bug at the moment it presses the button.

(The eye, unlike the bug "knows" its position s and, in order to repeat a move, it needs to "forget" s. Besides the eye has several independent arrays of buttons corresponding to different modes of eye movements, some of which are rather random.)

This may seem not much to start with; amazingly, this is exactly what is needed for reconstruction of the *affine structure* of the Euclidean plane: this structure is uniquely determined by distinguishing particular triples of points x_1, x_2, x_3, namely those which lie on a line with x_2 being halfway from x_1 to x_3.

Moreover, if the bug can "count" the number of repetitions of identical moves, it can evaluate distances and, thus, reconstruct the full Euclidean (metric) structure of the space, 2-plane in the present case.

Which buttons has the bug (eye) to press in order to efficiently explore the leaf and learn something about its meaning – the shape of the leaf?

The bug (eye) feels good at the beginning being able to predict that the color usually does not change as the bug (eye) makes small moves. (The eye, unlike the bug, can make fast large moves.) But then it becomes bored at this repetitiveness of signals, untill it hits upon the edge of the leaf. Then the bug (eye) becomes amazed at the unexpected change of colors and it will try to press the buttons which keep it at the edge.

(Real bugs, as everybody had a chance to observe, spend unproportionally long time at the edges of leaves. The same applies to the human eyes.)

In order to keep at the edge, the bug (this is more realistic in the case of the eye) needs to remember its several earlier moves/buttons. If those kept it at the edge in the past, then repeating them is the best bet to work so in future. (This does work if the edge is sufficiently smooth on the bug's scale.)

Thus, the bug (eye) learns the art of to navigation along the edge, where it enjoys twice predictive power of what it had inside or outside the leaf: the bug knows which color it will see if it pushes the "left" or the "right" buttons assuming such buttons are available to the bug. (The correspondence "left"$\leftrightarrow$"right" adds yet another involution to the bug's world symmetry group.)

Amazingly, this tiny gain in predictive power, which make the edge interesting for the bug, goes along with a tremendous information compression: the information a priory needed to encode a leaf is proportional to its area, say $A\cdot N^2$ bits on the N^2-pixel screen (where A is the relative number of the pixels inside the leaf) while the edge of the leaf, a curve of length l, can be encoded by $const\cdot l\cdot N\cdot \log N$ bits (and less if the edge is sufficiently smooth). Unsurprisingly, edge detection is built into our visual system.

(The distribution of colors near the edge has a greater entropy than inside or outside the leaf but this is not the only thing which guides bugs. For example, one can have a distribution of color spots with essentially constant entropy across the edge of the leaf but where some pattern of this distribution changes at the edge, which may be hard to describe in terms of the entropy.)

Eventually the bug (eye) becomes bored traveling along the edge, but then it comes across something new and interesting again, the tip of the leaf or the T-shaped junction at the stem of the leaf. It stays there longer plying with suitable buttons and remembering which sequences of pressing them were most interesting.

When the bug start traveling again, possibly on another leaf, it would try doing what was bringing him before to interesting places and, upon hitting such a place, it will experience the "deja vu" signal – yet another letter/word in bug's language.

We have emphasized the similarities between the eye and the bug movements but there are (at least) two essential differences.

1. The eye moves much faster than the bug does on the neurological time scale.

2. The eye can repeat each (relatively large) "press the button" move only a couple of times within its visual field.

On the other hand, besides "repeat", there is another *distinguished* move available to the eye, namely, "reverse". This suggests that the (approximate) back and forth movements of the eye will appear unproportionally often. (Everybody's guess is that the eye would employ such moves for comparing similar images, but I have not check if this has been recorded in the literature.)

Geometric incarnations of "reverse" are more amusing than the affine spaces associated with "repeat". These are known in geometry as *Riemannian sym-*

metric spaces X (see 3.6) where each "move" issuing from *every* point x in X is "geometrically equivalent" to the reverse move from the same x, i.e., for every x in X, there is an involutive isometry I_x of X the fixed point set of which equals x (i.e. $I_x(y) = y \Leftrightarrow y = x$).

The simplest instances of such X are the Euclidean, hyperbolic and spherical spaces, e.g. the 2-sphere of the visual field of the eye but there are other spaces with more elaborate and beautiful symmetric geometries (see 3.6.)

But the problems faced by out bug are less transparent than evaluating a metric in a *given* space. Indeed, pretend that you – a mathematician – are such a bug sitting at a keyboard of buttons about which you know nothing at all. When you press a button, either nothing happens (the color did not change) or there is a blip (indicating the change of the color).

Can you match these buttons with moves on the plane and the blips with crossing the boundaries of monochromatic domains?

What is the fastest strategy of pressing buttons for reconstruction the shape of a domain?

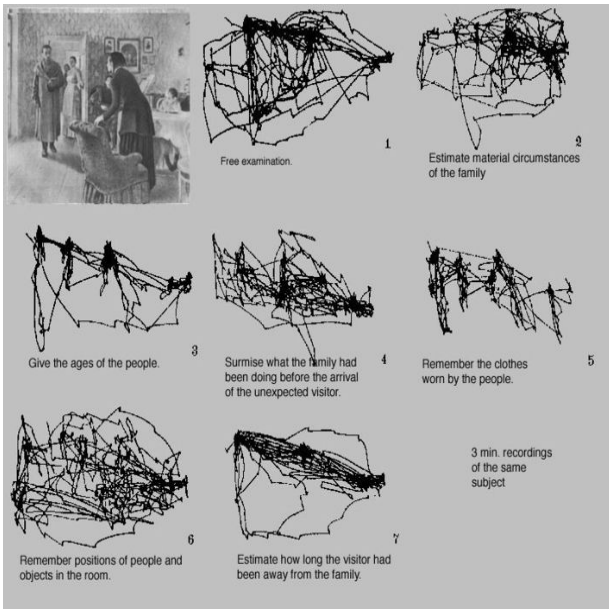
The answer depends, of course, on the available moves and the shapes of the domains you investigate: you need a rich (but not confusingly rich) repertoire of moves and the domains must be not too wild.

What you have to do is to create a language, with the letters being your buttons(+blips), such that the geometric properties of (domains in) the plane would be expressible in this language of sequences of (pressing on) buttons marked by blips. If in the course of your experiments with pressing the buttons, you observe that these properties (encoded by your language) are satisfied with the overwhelming probability, you know you got it right.

But what the bug (eye) has to do is more difficult, since, apparently, there is *no a priori idea of spacial geometry* in bug's brain. Bug's (eye's) geometry is, essentially, the grammar of the "button language". Thus, bug's brain (and an ergobrain in general) can not use a strategy tailored for a particular case, but may rely only on some *universal* rules, as the real bugs (eyes), we believe, do. The success depends on the relative simplicity/universality of the plane geometry, more specifically on the group(s) of symmetries of the plane. (This symmetry is broken by "colored" domains in it, and, amusingly, *breaking* the symmetry makes it perceptible to an "observer" – the bug or the eye at the keyboard.)

Apparently, the bug is able to make an adequate picture of the world, because the mathematical universality of bug's strategies matches the universal mathematical properties of the world.
is the unexpected switch from $\bullet$ to $\square$: your eye stays disproportionally long time focused at this point. Also, your pay most attention to the ends of words and it usually doesn't mcuh mttaer in waht oredr the ltteers in a wrod are. Similarly, your eye spends more time focused at the *edges* of images and bugs love crawling along the edges of leaves.

“Eye movements reflect the human thought processes; so the observer's thought may be followed to some extent from records of eye movements. ...The observer's attention is frequently drawn to elements which do not give important information but which, in his opinion, may do so. Often an observer will focus his attention on elements that are unusual in the particular circumstances, unfamiliar, incomprehensible, and so on."'(Yarbus, taken from Eye Tracking in



Wikipedia, also see [?])

Thus various degrees of “interest/amusement” and “ surprise/amazement” are rough indicators of a relation of the structure of your egobrain to the informational structure presented in the flow of signals. (If you are bored by a sequence of letters, this may be because you are not familiar with the language or, on the contrary, if you were obliged to memorize this sequence as a child at your school lessons.)

Several months old babies start playing with: "PA PA BA BA". Their auditory system records these "PA PA BA BA" with an *amazing* correlation between the sounds and the somatosensory and tactual perceptions. Deaf children start "PA PA BA BA" at the same age but stop doing it sooner: there is less to be amazed with. Eventually, babies get bored with this and start producing more interesting/meaningful sequences of sounds, unless taught that BLA-BLA-BLA is an acceptable adult speech.

Similarly, it is more interesting to run than to walk, to walk than to stand and to stand than to lie: the structure of brain’s strategies for keeping your from falling down is the most intricate when you run and the correlation between the visual and somatosensory perceptions is most amazing for the brain in this case.

Our main assumption, in agreement with [?], is that an ergobrain comes to "understand" the world by “trying to maximize” its “predictive power” but what it exactly predicts at every stage depends on what structure has been already built. (If predictability is understood broadly enough, on all levels of underlying structure, then there is no discordance with Rene Thom’s "*Prédire n’est pas Expliquer*".)

In order to maximize anything, one needs some freedom of choice, e.g. your eye needs a possibility to run along lines/pages or, in a chess game, you can choose from a certain repertoire of moves. When this repertoire becomes constrained, the ergobrain feels frustrated. This is not visible from the outside, (contrary, say, to the feeling of surprise which is visible on one’s face): one can only rely on introspection.

Many of us have experienced an uncomfortable feeling when a lecturer shows slides line by line, preventing you from seeing the whole page. (Such lectures are meant for the heads of Turing machines, rather than for those of humans.) One undergoes a similar feeling of frustration when studying a book or an article (often, alas, written by a mathematician) where the author is purposefully putting a horse blinder on reader's mind's eye. And the poetry by deafblind people tell us how much one should value one's freedom to learn.

("Freedom" for an ergobrain does not mean just a possibility to generate any kind of signals it "wants", but rather to have "interesting" environmental responses to these signals. Thus, for example, a bug crawling on a homogeneous surface of an imaginary infinite leaf has zero freedom: no matter where it goes it learns nothing new. But the presence of a structural feature, e.g. of an accessible edge of the leaf, significantly adds to bug's freedom.)

The "amusing, "amazing","bored"and other ergo-mood signatures seem little to start with, but we shall see that not so little can be reconstructed from these, if one studies the ergobrain by "ergomeans". (Ergobrain's out-coming "Eureca"s are not specific enough to be structurally informative. On the other hand, some short time scale perception phenomena, e.g. optical illusions, may carry a non-trivial ergo-message.)

This is not so easy as it looks since it is hard to tell which concepts and ideas are of ergo- and which are of of ego-origin. For example our egomind believes that mathematics is something "abstract and difficult" while the objects we see in front of our eyes are "concrete and simple". However, the ideas of these objects are created by a complicated process of the image building by your visual ergosystem, the result of which is something abstract and artificial from the point of view of your ergobrain, while mathematics is very similar to what ergobrain was doing all its life.

The input of the visual system amounts, roughly, to the set of samples of a distribution of something like a probability measure on the set of subsets of the light receptors in your retina. This "set of samples" is at least as "abstract" as the invariance of the Euclidean 3-space geometry under the orthogonal group $O(3)$, while the reconstruction *set-of-samples* $\rightsquigarrow$ $O(3)$-*invariance* by your ergobrain is a mathematical endeavor.

(The mathematics of building/identifying the $O(3)$-symmetry of the visual perception field is similar to but more complicated than how Alfred Sturtevant reconstructed in 1913 *linearity of the gene arrangements* on the basis of distributions of phenotype linkages long before the advent of the molecular biology and discovery of DNA [?].)

The difficulty of "abstract" mathematics is, apparently, due to the protective wall separating ego from ergo

5.3 Signals, Learning and Ergo-Structures

. We want to understand the process(es) of learning, e.g. of *mother tongue* or of a *mathematical theory*, in the context of what we call *ergostructures*. Such structures, as we see them, are present in the depth of the human (some animal?) minds, in natural languages, in logical/combinatorial organizations of branches of mathematics and, in a less mature form, in biological systems –

from metabolic and regulatory networks in living cells up to, possibly, ecological networks.

Learning from this perspective is

a dynamical process of building the *internal ergostructure* of an $\mathcal{L}$ from the *raw structures* in the incoming flows $\mathcal{FS}$ of signals, where $\mathcal{FS}$ may or may not itself contain an ergostructure or its ingredients.

???????????? Such an $\mathcal{L}$ interacting with a flow of signals is similar to a a photosynthesizing plant growing in a stream of photons of light or to an amoeba navigating in a see of chemical nutrients and/or of smaller microbes: $\mathcal{L}$ recognizes and selects what is *interesting* for itself in such a flow and uses it for *building* its own structure.

This analogy is not fully far fetched. There is no *significant* difference between human activities and those by amoebas and even by bacteria, well,... on the GRAND SCALE. Say, the probability of finding first 10^9 digit of $e = 2.718...$. "written" at some location u of a universe $\mathcal{U}$ increases by a factor $> 10^{100}$, if you find a bacterium kind machine feeding on a source of almost amorphous *free energy* at a point u' within a few billion light years from u. ???????????

Besides, what enters the brain mainly comes from plants, animals, humans and human artifacts – it would be little to learn for our nose, ears and even your eyes if not for Life around us.

This, however, does not apply to your *somatosensory* input: much of it comes from non-biological external sources. The somatosensory system is also exceptional in several other respects: it is short range and, most significantly, *proprioception* – your "body/muscle sense" is almost fully *interactive*. This is because the brain's output is mostly directed toward the muscles in the body and to feel your body you have to move it. (Besides muscles, the brain sends signals to the endocrine system and also it "talks to itself" but the conscious control over these brain activities is limited.)

And the integrated picture of the world that remains stable under transformation by Euclidean isometries of space and scaling transformations – the crowning creation of your visual+somatorsensory system[119], has little to do with surrounding life, but nearly all interesting visual sensations are of the biological origin, except, maybe, for the hypnotizing charm of wandering water streams and the irresistible beauty of cloud shapes sliding overhead.

It is virtually impossible to say something of substance on *mind, thinking, intellegence,* without a resort to metaphors that serve to create an illusion of understanding and to hide inconsistency of the ideas they purport to convey.[120]

A seductively attractive instance of this is the following sentence.

Our thoughts are determined by the laws of chemistry, not those of logic, since they are *result of the chemical processes in the brain.*[121]

Neither this sentence nor its negation make any sense because the concepts "thoughts", "determination", "laws of ....", "chemistry", "logic" belong to dif-

[119] An essential part of "somatorsensory" in understanding geometric space is *proprioception* – the sense of your body that is coupled with the *motor system*; the tactile/touch feeling is also involved but to a lesser(?) extent.

[120] In poetry, unlike how it is in science, we joyfully welcome illusions created by the beauty of metaphors.

[121] This is a misquote of John Haldane's *: If my opinions are the result of the chemical processes..."*

ferent categories.[122] "Pheromones paved" highways made by ants demonstrate the shortcomings of this kind of alchemy of words.

To supply substance to *learning, thinking, understanding*, etc., one has to to find mathematical counterparts of these concepts. These for us are certain unknowns x the essential properties P of which we try to guess with x satisfying P thought of as equations $P(x)$. Only when such an equation is "written down" one may proceed to search for its solutions, that is designing a leaning system $\mathcal{L}$ implementing x.

Eventually, we have to express this in truly mathematical terms, but we resort to a metaphoric language for a while, since we do not want to narrow our field of vision and miss the target with prematurely precise definitions.

(Being non-precise is tolerable in mathematics, since the ambience of mathematical *structures*, allows, within certain limits, a non-rigorous yet productive discourse with concepts that are not immediately clearly defined.)

Understanding and modeling human thinking processes is dissimilar to other problems in science and engineering but the following example may be instructive.

How to get to the Moon.

People might have been wondering about this for millennia but the following *formulation(s)* of the question became possible only with development of physical science and mathematics in the last three centuries.

What are *all conceivably possible orbits/trajectories* for bodies traveling between the Earth and the Moon?

What kind of *propulsion mechanism(s)* can bring you to such an orbit?

It is unthinkable to plan a trip to the Moon prior to forming adequate concepts of *equations* mathematically describing *properties* of these "orbits"[123] and of "mechanisms".[124]

Similarly, without a proper *reformulation* of the problem and *a mathematical description* of essential, partly conjectural, properties of

thinking/learning/understanding systems,

it is unimaginable to make a radical advance in the study of "thinking mechanisms"; also a "blind design" of a functional model of a "thinking system" appears unrealistic.

5.4 meaning and understanding

5.5 Teaching and Grading.

A universal language learning problem *PRO* is supposed to models a mind of child and it needs only a minimal help from a "teacher", such as ordering texts

[122]If you are a mathematician who feels shaky about "laws", "chemistry", etc, imagine a (non-poetic) statement about "small category" in a philosophical tractate (say, about Aristotle) with no hint at the appropriate context of (Grothendieck-stile) mathematics where this concept belong with.

[123]Orbits are described by *differential equations* of motion in a graviton field.

[124]A basic aspect of a relevant "mechanism" is described by the *Tsiolkovsky ideal rocket equation.*

according to their complexity[125] and allowing PRO a flexible access to texts.

On the other hand, evaluation of the *quality of understanding* by PRO is harder (albeit much easier than designing a learning program itself), since no one has a clear idea what *understanding* is.

Our formal approach is guided, in part, by how it goes in physics, where an unimaginably *high level of understanding* is reflected in the *predictive power* of mathematically formulated *natural laws* that encapsulate enormously *compressed data.*

This lies in a category quite different from what we call *"knowledge"*.

For instance, ancient hunters *knew* more of how planets wander in skies than most modern people do. But understanding of this wandering depends on "compression" of this knowledge by setting it into the slender frame of mathematically formulated laws of motion.[126]

Similarly, understanding languages depends on compression of structural redundancies[127] in flows of linguistic signals. albeit this compression is not as substantial as in physics.

Besides "sheer knowledge", *understanding* should be separated from *adaptation.* For instance, an experienced rodent (or a human for this matter) competently navigates in its social environment. But only metaphorically, one may say that the rodent (or human) *"understands"* this environment.

With the above in mind, we indicate the following two mutually linked attributes of what we accept as "*understanding*".

[1] *Structural compression of "information".*

[2] *Power of prediction.*

These **[1]** and **[2]** can be quantified in a variety of ways. For example, one may speak of the degree of compression versus the "percentage" of structure lost in the course of compression, while the essential characteristics of a prediction is *specificity* versus *frequency of success* .

This kind of quantification may be used for *partially ordering "levels of understandings"* that may suggest tests for evaluating progress achieved by a learning program PRO in these terms.

Another attribute of *"understanding"* that is easy to test but hard to quantify is as follows.

[3] *Ability to acquire knowledge.*

For instance, a program PRO minimally proficient in English, would "know", upon browsing through Encyclopedia Britannica, that cows eat grass and cats eat mice.[128]

Also, the following can be seen as a hallmark of understanding.

[125] One may also equip PRO with an ability, similar to that possessed by children up to the age 2-3, to resist a "bad teacher" by rejecting environmental signals that are detrimental for the learning. (This ability deteriorates with age as one has to adapt to the environment in order to survive.)

[126] Ancients astronomers came to *understand* periodicity of planetary motions and were able to make rather accurate predictions.

[127] One can not much compress the "useful information" without loosing this "information" but if we can "decode" the structure of redundancy it can be encoded more efficiently.

[128] Properly responding to *"Do black cats eat fresh mice?"* instead of plain *"Do cats eat mice?"* would need a study of a more representative corpus of English than Encyclopedia Britannica by PRO.

[4] *Ability to ask questions.*

(Those whose business is UNDERSTANDING – scientists and young children – excel in asking questions.)

Besides the ability to understand, learning programs PRO may be graded according to their "internal characteristics", such as the *volume of the memory* a PRO has to use, the *number of elementary operations* and *the time* needed for it to make, for instance, a particular prediction.

At some point letter on , we shall comprise a list of specific criteria a learning/understanding program must satisfy.

5.6 Understanding Structures and the Structure of Understanding.

If there was a parrot which could answer every question,
I should say at once that it was a thinking being.
DIDEROT, PENSEES PHILOSOPHIQUES, 1746.

But...

It never happens that it [an automaton] *arranges*
its speech in various ways, in order to reply appropriately
to everything that may be said in its presence,
as even the lowest type of man can do.
DESCARTES, DISCOURSE ON METHOD, 1637.

Is Descarte justified in his belief that no machine can pass what is now-a-days called *Turing Test*, i.e. *to reply appropriately to everything that may be said in its presence*?

Does passing such a test certify one as a THINKING BEING who UNDERSTANDS what is being said, as Didereot maintains?

What does it mean to UNDERSTAND, say a language or any other flow $\mathcal{FS}$ of signals?

Diderot indicates a possible answer:

the continuity of ideas, the connection between propositions,
and the links of the argument that one must judge if a creature thinks.

In general terms, UNDERSTANDING includes:

$[\bullet]_U$ *a certain mathematical (logical?)* ***structure*** *U in the understander's mind/brain/program;*

$[\bullet]_{[IU]}$ *a* ***process*** *IU of implementation of U by an ergosystem representing "an understander";*

$[\bullet]_{[RU]}$ *the* ***result*** *RU of such implementation, $RU = [IU](\mathcal{FS})$, where $[IU]$ is seen as a transformation applied to flows of signals.*

The catch is that nobody has a clear idea of what kind of structure it could be; this precludes any speculation on how and where such a structure can be implemented. Besides such a structure is by no means unique but rather different U are organized as a structural community that can be partly described in category theoretic terms.

An essential feature, one may say the *signature* of a U, is its space/time characteristics: U is much smaller in the volume content than the totality of the flow $\mathcal{FS}$ it "understands" and application of $[IU]$ to $\mathcal{FS}$ is much faster than achieving U.

It takes, probably, $\approx l \log l$ elementary steps for learning $\mathcal{FS}$ of length l that translates to months or years when it comes to learning a language or a mathematical theory.[129] But when learning is completed, it takes a few seconds to realize, for instance, that a certain string of symbols in the language of your $\mathcal{FS}$ is completely meaningless.

On the other hand, the space/volume occupied by an understanding program U is a few orders of magnitude greater than a learner's program $\mathcal{L}$, where such a program is *universally* (independently of the total number of signals from $\mathcal{FS}$ received/inspected by a learner) bounded by something like 10^6 bits. Picturesquely,

$$\overset{\star}{\mathcal{L}} \qquad \underset{U}{\circledast} \equiv \blacksquare - \blacksquare - \blacksquare - \blacksquare - \blacksquare - \blacksquare - \blacksquare -$$

where $\circledast$ represents the "core understanding" – a few thousand page "dictionary+grammar" of $\mathcal{FS}$ that is augmented by several (tens, hundreds or thousands depending on $\mathcal{FS}$) "volumes" $\blacksquare$ of loosely (imagine RAM on your commuter) organized "knowledge", while the available $\mathcal{FS}$ itself may number in tens or even hundreds of millions of nearly unrelated units – volumes, internet pages, images memorized by your visual system, etc:

... ●●●●●●●●●●●●●●●●●●●●●●●●
●●●●●●●●●●●●●●●●●●●●●●●...

(We do not know for sure if *understanding* is a formalizable concept, since the only convincing argument in favor of this would be designing a functional *thinking machine/program*, while the only conceivable NO might come from an incredible discovery of a hitherto unknown fundamental property of the live matter of the brain.

But impossibility of resolving the UNDERSTANDING and the *thinking machine* problems by speculative reasoning does not abate our urge to make the world know what our *gut feeling* tells us about these issues.[130]

Amusingly, the *gut feeling* itself, at least the one residing in dog's guts, unlike the ideas propagated from human guts to human minds, was experimentally substantiated by A. N. Drury, H. Florey[131] and M. E. Florey in their study of *The Vascular Reactions of the Colonic Mucosa of the Dog to Fright*, 1929.)

When we say "UNDERSTANDING" we mean *understanding structural entities* where such an *understanding* is seen as a structural entity in its own right that admits a non-trivial mathematical model/description.

We conjecture that most (all?) structures we encounter in life, such as

[129] The true measure of time, call it *ergo-time*, should be multi-(two?)-dimensional, since it must reflect parallelism in programs modeling learning and other mental processes.

[130] An attempt to explain the reason for the incessant flow of publicized expressions of YES and of NO opinions on this subject matter is made in section 6.5 of our SLE-paper.

[131] If there is a single person in the human history responsible for saving nearly hundred million lives – this is *Howard Florey* whose titanic efforts had brought penicillin to the therapeutic use by mid-40s.

natural languages, mathematical theories, etc. are *understandable*[132] and we search for mathematics that can describe this *understanding*.

Answers to the following questions, let these be only approximate ones, may serve to narrow the range of this search.

QUESTION 1. What are essential (expected? desired?) features/architectures of mathematical models of structural understanding?

QUESTION 2. If such a model exists should it be essentially unique? In particular, are the hypothetical structures of understanding, say of a language and of chess must necessarily be closely resembling one another?

QUESTION 3. How elaborate such a model need to be and, accordingly, how long should be a computer program implementing such a model?

QUESTION 4. What is an expected time required for finding such a model and writing down the corresponding program?

QUESTION 5. What percentage of this time may be delegated to machine (ergo)learning with a given level of supervision?

QUESTION 6. How much the supervision of such learning can be automated?

QUESTION 7. What are criteria/tests for performance of "I UNDERSTAND" programs?[133]

QUESTION 8. Can Turing-like tests be performed with *algorithmically* designed questions that would trick a computer program to give senseless answers?

QUESTION 9. Are there simple rules for detecting senseless answers?

QUESTION 10. Can the human learning (teaching?) experience be of use for designing clever learning algorithms?

QUESTION 11. Does ergo logic help answering the above questions?

Discussion. When we say "mathematical structure" we do not have in mind any particular branch of the continuously growing and mostly hidden from us enormous tree that is called MATHEMATICS. Whatever a relevant branch can be its structure, probably, is quite elaborated; very likely this branch has not grown yet on this tree.

But sometimes things are simple, e.g. for the vervet monkey *"Alarm Call Language"* that matches a few (four?) word-signs – their alarm calls, with object-events, that are particular predators – leopards, eagles, pythons, baboons.

However, no monkey would think that a mathematical *one-to-one correspondence*, call it ACL, between two 4-element sets *understands* the meaning of the alarm calls even if this ACL is implemented by a monkey shaped robot that properly reacts to predators by correctly emitting the corresponding calls and, thus, passes the vervet monkey Turing test.

Why then do Decartes and Diderot, not to speak of Turing himself, attach such significance to the Turing test?

Is there an *essential* difference between the correspondence *"questions"* $\longrightarrow$ *"correct answers"* and the ACL correspondence?

The answer is:

Yes, there is an essential difference, an ENORMOUS DIFFERENCE.

[132]Overoptimistic? Yet, in line with the remark *"... mystery of the world is its comprehensibility"* by Einstein.

[133]Designer's own ability to pass a test is a poor criterion for designing such a test.

Operating with tiny sets, e.g. composed of four uterings – alarm calls of vervet monkeys and with correspondences between such sets needs no structure in these sets. But one CAN NOT manipulate human uterings and even less so longish strings of utterings and/or written texts in a structureless way.

It is tacitly assumed by scientifically minded people — Decartes, Diderot, Turing... , that the above correspondence "$\longrightarrow$" must be compatible with the essential structure(s) of the human language, call it $\mathcal{HL}$, used in a particular Turing test, where the basic (but not the only) structure in $\mathcal{HL}$ is that of an *exponential/power set*:

an uttering/sentence, say in thirty words, in a language with dictionary D is seen as a member of the HUGE *power set*

$$D^{30} = \underbrace{D \times D \times D \times ... \times D}_{30}.$$

Such structurality is indispensable for an implementation of a "thinking automaton" and/or the program running it in a *realistic* space time model [134] that *necessarily excludes*, for instance, "imaginary programs" containing in their memories lists of more than, say of N^{15}, sentences with number N being comparable with the cardinality $card(D)$ of the dictionary.[135]

("Large sets", be they finite or infinite, have no independent existence of their own, but only as carriers of structures in them, similarly how the space-time in physics makes no sense without energy-matter in it. This is not reflected, however, in the set theoretic notation that may mislead a novice. For instance, it is rarely stated in elementary textbooks that the "correspondence" $x \mapsto y$ in the "definition" of a *real variable function* $y = f(x)$ is only a metaphor and that a function $f(x)$, if it claims the right to exist in mathematics, must *"respect" some structure* in the set of real numbers.[136])

5.7 Meaning of Meaning.

Everything we call meaningful is made of things that cannot be regarded as meaningful.

... "meaning" is... a word which we must learn to use correctly.

NIELS BOHR MISQUOTED[137]

Meanings of words are determined to a large extent by their distributional patterns.

[134]The property of being physically realistic is often missing in philosophical discourses on artificial intellegence.

[135]This very sentence: "Such structurality is... of the dictionary" contains forty words with roughly half of then being nouns, verbs and adjectives. By varying these, one "can" generate more than $1000^{20} = 10^{60}$ grammatical sentences. Can one evaluate the number of meaningful ones among them? Would you expect *thousand* of them or, rather, something closer to *ten thousand*? Is it conceivable that "weakly meaningful" sentences number in 10^6, or there are more than 10^{10}, or even greater than 10^{18} of them?

[136]There is no accepted definition of *"function"* that would separate the *wheat*: $\sin x$, $\arctan x$, $\sqrt{x}$, $\log x$, Rieman's $\zeta(x)$, Dirac's $\delta(x)$,..., from the *chaff*, such as the characteristic function of the subset of rational numbers.

[137]In the original, one has "real" instead of "meaningful" and "reality" instead of "meaning".

This "*meaning*" of Harris is quite different from the common usage of the word "*meaning*" that invariably refers to "the real world" with "meaningful" being almost synonymous to what is advantages for preservation and propagation of (observable features encoded by) your genes. (The speakers of the word are usually blissfully unaware of this and they are getting unhappy with such interpretation of meanings of their actions.)

The former is a *structural meaning* the full extent of which may be discerned only in the dynamics of the learning processes in humans, while the latter – the concept *pragmatic meaning*, is shared by all living organisms, at least by all animals from insects on. This idea of meaning – the commandment to survive – was firmly installed in our brain hardware by the evolutionary selection several hundred million years before anything resembling humans came to existence.

A possible way to look beyond the survival oriented mode of thinking is to turn your mind toward something like chess, something that does not (contrary to what Freudists say) carry a significantly pronounced imprint of the evolutionary success of your forefathers.

But even if you manage to switch your mind from ego- to ergo-mode, you may remain skeptical about (ergo)chess telling you something nontrivial about learning languages and understanding their meanings.

Superficially (this is similar but different to what was was suggested by Wittgenstein), one may approach a dialog in a natural language as a chess-like game that suggests an idea of (ergo) meaning: the *meaning* of an uttering U is derived similarly to that of the meaning of a position P in chess: the latter is determined by the combinatorial arrangement of P within the ergostructure $\mathcal{CHESS}_{ergo}$ of "all" ergo-interesting chess positions/games while the former is similarly determined by its location in the architecture of $\mathcal{TONGUE}_{ergo}$ of a language.

More generally, we want to entertain the following idea.

The meanings assigned by ergostructures (e.g. by our ergobrains) to signals are ***entirely*** *established by patterns of combinatorial arrangements and of statistical distributions of "units of signals", be they words, tunes, shapes or other kinds of "units".*
Understanding is a ***structurally organized*** *ensemble of these patterns in a human/animal ergobrain or in a more general ergosystem.*

But even leaving aside the lack of precision in all these "pattern", "arrangement", etc. one may put forward several objections to this idea.

The most obvious one is that words, and signals in general, are "just names" for objects in the "real world"; the "true meaning" resides in this world. But from the brain perspective, the only "reality" is the interaction and/or communication of the brain with incoming flows of signals. The "real word" is an abstraction, a model invented by the brain, a conjectural "external invisible something" that is responsible for these flows. Only this "brain's reality" and its meaning may admit a mathematical description and be eventually tested on a computer.[138]

[138]We do not want to break free from the *real world*, but from the hypnosis of the words EXISTENCE/NON-EXISTENCE coming along with it.

(There are many different answers to the questions "What is meaning?", "What is understanding?" offered by linguists, psychologists and philosophers.[139] We, on the other hand, do not suggest such an answer, since we judge our understanding of the relevant ergo-structures as immature. The expression "*structurally organized ensemble*" is not intended as a definition, but rather as an indication of a possible language where the concept of *understanding* can be productively discussed.)

Another objection may be that learning chess and understanding its meaning, unlike learning native languages by children, depends on specific verbal instructions by a teacher.

However, certain children, albeit rarely – this was said about Paul Morphy, Jose Raul Capablanca, Mikhail Tal and Joshua Waitzkindo – learn chess by observing how adults play. And as for supernovas, it would be foolish to rejects this evidence as "statistically insignificant".

More serious problems that are harder to dismiss are the following.

(∘) The structures $\mathcal{TONGUE}_{ergo}$ of natural languages are *qualitatively* different from $\mathcal{CHESS}_{ergo}$ in several respects.

Unlike how it is with chess, the rules of languages are non-deterministic, they are not explicitly given to us and many of them remain unknown. Languages are bent under the load of (ego)pragmatics and distorted by how their syntactic tree-like structures are packed into 1-dimensional strings.

Self and Time. The most interesting feature of natural languages – *self-referentiality* of their (ergo)syntax (e.g. expressed by *pronouns* and/or by certain *subordinate clauses)* that allows languages to *meaningfully* "speak" about themselves.

This is present in most condensed form in *anaphoras* such as in

X thinks he is a good chess player,

and related features common to all human languages are seen in *deixis*, such as in

but I am afraid you may be disappointed by the naivety of his moves,

along with various forms of *grammatical aspects* linked to the *idea of time.*

(It is hard to say how much of *time in the mind* is necessitated by the time dynamics of the neurobrain, what had been installed by the evolution and what comes with flows of incoming signals. And it is unclear if *time* is an essential structural component of the ergobrain and if should it be a necessary ingredient of universal learning programs.)

None of these have counterparts in chess[140] or in any other non-linguistic structure, e.g. in music. Yet, self-referentiality is seen in mathematics on its borders with a natural language, e.g. in *Gödel's incompleteness theorem.*

(∘∘) The internal combinatorics of $\mathcal{TONGUE}_{ergo}$ may be insufficient for the *full* reconstruction of the structure of the corresponding language.

For example, linguistic signals a child receives are normally accompanied, not necessarily synchronously, by what come via all his/her sensory systems,

[139] References can be found on the corresponding pages of Wikipedia.

[140] Does the "meaning" of the following sentence reside in the game being played or in the conjunction of syntactic self-referentiality loops in there?

I thought I understood why X's white knight was placed on a1 square but his next move caught me by surprise.

mainly *visual* and/or *somatosensory* signals – feeling of touch, heat, pain, sense of the position of the body parts, as well as *olfactory* and *gustatory* perceived signals.

The full structure of $\mathcal{TONGUE}_{ergo}$ and/or the meaning of an individual word may depend on (ergo)combinatorics of $\mathcal{VISION}_{ergo}$ coupled with $\mathcal{TONGUE}_{ergo}$ not on $\mathcal{TONGUE}_{ergo}$ alone.

$\mathcal{VISION}_{ergo}$ is vast and voluminous – more than half of the primate (including human) cortex is dedicated to vision, but the depth of structure of "visual" within $\mathcal{TONGUE}_{ergo}$ seems limited, as it is witnessed by the ability of deafblind people to learn to speak by essentially relying on their *tactile* sensory system that is feeling of touch.

The role of *proprioception* (your body/muscle sense) and the *motor control system* in learning (and understanding?) language is more substantial than that of vision, since production of speech is set in motion by firing motor neurons that activate muscles involved in speech production – *laryngeal muscles, tongue muscles* and hordes of other muscles (hand/arms muscles of mute people); thus, an essential part of human linguistic memory is the memory of sequential organization of these firings.

(Proprioception, unlike vision, hearing and olfaction, has no independent structural existence outside your body; also it is almost 100% interactive – you do not feel much your muscles unless you start using them. The internal structure of proprioception is quite sophisticated, but, probably, it is by no means "discretized/digitalized" being far remote from what we see in language. It is hard to evaluate how much of language may exist independently of $\mathcal{PROPRIOCEPTION}_{ergo}$(+$\mathcal{TACTILE}_{ergo}$[?]) coupled with the motor control system, since a significant disfunction of these systems at early age makes one unable to communicate.)

The above notwithstanding, (ergo)programs (as we see them) for learning chess and a language, and accordingly, the corresponding ideas of *meaning* and *understanding* have much in common.

To imagine what kinds of programs these may be, think of an *ergo-entity*, call it $\mathcal{EE}$, from another Universe to whom you want to communicate the idea/meaning of chess and with whom you want to play the game.

A preliminary step may be deciding whether $\mathcal{EE}$ is a *thinking* entity; this may be easy if $\mathcal{EE}$ possesses an ergobrain similar to ours, which is likely if ergo is universal.

For example, let $\mathcal{EE}$ have a mentality of a six-year-old Cro-Magnon child, where this "child" is separated from you by a wall and where the only means of communication between the two of you is by tapping on this wall.

Could you decide if the taps that come to you ears are produced by a possessor of an ergobrain – more versatile than yours if you are significantly older than six, or from a woodpecker?

If you happens to be also six year old, the two of you will develop a common tap language-game and enjoy *meaningfully* communicating by it, but possessors of two mature human minds separated by a wall will do no better than two adult woodpeckers.

To be a good teacher of chess (or of anything else for this matter), you put yourself into $\mathcal{EE}$'s shoes and think of what and how yourself could learn from (static) records of games and how much a benevolent and dynamic chess teacher would help. You soon realize that this learning/teaching is hard to limit to chess as it is already seen at the initial stage of learning.

Even the first (*ergo-trivial*) step – learning the *rules* of moves of pieces on the board will be virtually insurmountable in isolation, since these rules can not be guessed on the basis of a non-exhaustive list of examples, say, thousand samples, unless, besides ergo, you have a simple and adequate representation of the geometry of the chess board in your head.

If your are blind to the symmetries of the chessboard, the number of possible moves by the white rook ♖ in the presence of the white king ♔, that you must learn (in $64 \cdot 63$ positions), is $> 64 \cdot 63 \cdot 13 > 50\ 000$. "Understanding" space with its symmetries, be this "understanding" preprogramed or acquired by *a learning process of spacial structre(s)*, is a necessary prerequisite not only for learning chess but also for communication/absorbtion of the rough idea of chess.[141]

But if you have no ergo counterparts to such concepts as "*some piece on a certain line*"[142] in your head, you'll need to be shown the admissible moves of the rook in *all*($> 10^{45}$) possible chess positions.

And the more you think about it the clearer it becomes that the only realistic way to design a chess learning/understanding program goes via some general/universal mathematical theory equally applicable to learning chess and learning languages.

5.8 Sixteen Rules of Ergo-Learner.

The general guidelines/principles suggested by *ergo-logic* for designing universal learning algorithms can be summarised as follows.

1. Flows of signals coming from the external world carry certain structures "diluted" in them.

Learning is a process of extracting these structures and incorporating them into learner's own *internal structure*.

2. The essential learning algorithms are *universal* and they indiscriminately apply to all kind of signals.[143]

3. Universality is incompatible with any a priori idea of "reality" – there is no mental picture of what we call "real world" in the"mind" of the learner.

The only *meaning* the learner assigns to "messages" coming from outside is what can be expressed in terms of (essentially combinatorial) *structures* that are recognised and/or constructed by the learner in the process of incorporating these "messages" in learner's internal structure.

---

[141] The geometry of the board can be reconstructed from a moderate list of sample chess games with *Poincaré's-Sturtivant space learning algorithms* (see §4 in [4]), but these algorithms are slow.

[142] Such "abstractions" are probably acquired by the visual ergo-system of a child well before to something as "concrete" as *white knight in a particular position on a chessboard*, for example.

[143] The learner's behaviour, that is learner's interaction/conversation with incoming signals, also depends on the learner's internal structure that has been already built at a given point in time. In particular, a prolonged exposure to a particular class of signals makes learner's behaviour more specialised (more efficient?) while learner's ability to absorb and digest different kinds if signals declines.

4. Universality also implies that the actions of the learner – building internal structures and generating signals, both within itself and/or released outside,[144] *are not governed by goals* expressible in terms of the external world.

The learning is driven by learner's *"curiosity" and "interest"* in structural patterns the learner recognises in the incoming flows of signals and in the learner's delight in the logical/combinatorial beauty of the structures the learner extracts from these flows and the structures the learner builds.

Essential ingredients of the learning process are as follows.

5. The learner discriminates between familiar signals and novelties and tries to match new signals with those recorded in its memory.

6. The learner tries to *structurally extrapolate* the signals already recorded in its memory in order *to predict* the signals that are expected to come.

7. Besides the signals coming from the external world, the learner perceives, records and treats some *signals internally generates* by the learner itself.

8. The learner tends to *repeatedly imitate signals* being received, including some signals that come from within itself.

(The repetitiveness of their basic operations allows a description of learning processes as *orbits under some transformation in the space of internal structures of a learner.* The learning program that implements this transformation must be quite simple and the learning process must be robust. Eventually, "orbits of learning" stabilise as they approach approximately fixed points.)

9. The learner tends *to simplify signals* it tries to imitate.

10. The learner systematically makes guesses and "jumps to conclusions" by *making general rules* on the basis of regularities it sees in signals.

11. When the learner finds out that a rule is sometimes violated, the learner does not reject the rule but rather adds *an exception.*

12. The learner tends to use *statistically significant* signals for building its internal structure as well as for making predictions. But sometimes, the learner assigns significance to certain exceptionally rare signals and use them as essential structural units within itself.[145]

13. The learner probabilistic reasoning in uncertain environment is *yes-maybe-no* logic.

We impose the following restrictions on the abilities of our intended learner programs that are similar to those the human brain has.

14. The learner *does not accept unstructured sets* with more than four-five items in them; upon encountering such a set the learner invariably assigns a certain structure to it.[146]

15. The learner has *no built-in ability of sequential counting* beyond 4 (maybe 3); we postulate that $5 = \infty$ for the learner.

In particular, the learner is not able to produce or perceive five consecutive iteration of the same process, unless this becomes a *routine* delegated from *cerebral cortex* to *spinal cord.*[147]

[144] These are the "actions" the human brain is engaged in.

[145] It is the rare words in texts that are significant, not the most frequent ones.

[146] Partition of stars in the sky into constellations is an instance of this.

[147] Never mind *the kid that fought his dad that bought the car that struck the bike that hit the truck that brought the horse that kicked the dog that chased the cat that caught the rat that ate the bread.*

Our main conjecture is that *universal learning algorithms exist* and, moreover, their formalised descriptions are quite simple.

The time complexity of such an algorithms must be at most log-linear (with no large constant attached) and the performance of an "educated/competent program" must be no worse than logarithmic.

In fact, the essential features of (ergo)learning as we know it, make sense only on a roughly "human" time/space scale: such a learning may apply to flows of signals that carry 10^9 – 10^{15} bits of information all-together and one hardly can go much beyond this.[148]

Universality and Doublethink. If one expects an analysis of *a flow of signals*, e.g. of a collection of texts in some language $\mathcal{L}$ to be anywhere close to the TRUTH, and if one wants to design an algorithm for learning $\mathcal{L}$, one must, following ergo-logic, *disregard* all one a priori knows about this $\mathcal{L}$, *forget* this is a language, *reject* the idea of meaning associated to it

But the only way to evaluate the soundness of your design prior to a computer simulation of it, is to compare its performance to that of the corresponding algorithms in a human head.

5.9 Discretization, Classification, Interaction.

The process of learning mainly consists in *structuralizing* the incoming flows of signals by identifying redundancies in these flows and representing "compressed flows" of these signals in a structurally efficient way.

It is a fundamentally unresolved problem in psychology to identify *mathematical classes of structures* that would *model mental structures* built by human brains that assimilate incoming "flows of signals".

We do not know what, specifically, these structures are but their three ingredients are visible.

1. *Discretization and Formation of Units.* The structures built in the course of learning are *assumed to be discrete*, i.e. composed of *distinct units.*

There are several mechanisms, call them *discretizers*, responsible for formation of these units.

Initially, such units are obtained by "meaningful segmentation" of incoming flows of signals, such, for instance, as division of texts into *words* and *phrases* and/or identifying/isolating/separating "individual futures" in a visual field.

Eventually, everything deserving a name becomes a unit.

2. *Classification and Reduction of Units.* This is called *categorisation* by linguists, it is often implemented by *clusterization* algorithms and depicted by arrows

$$u \mapsto v = class(u)$$

called *reductions.* There are several distinct mechanisms/algorithms of classifications, call them *classifiers*, that run in parallel.

[148]The universal learning systems themselves, e.g. those residing behind our skulls, have no built-in ideas of *meaning*, of *time*, of *space*, of *numbers*. But any speculation on natural or artificially designed "intelligent" systems strikes one as *meaningless*, if *spacial* and *temporal* parameters of possible implementations of such systems are *not specified* and set within realistic *numerical* bounds.

Some classifiers find and/or establish *similarity relations* between different units. For instance, words are divided according to their grammatical functions, such as the traditional (and contraversial) division into eight "parts of speech".

Classes, as they are being formed, are incorporated as *units* in the learner's structure.

3. *Connections between Units.* Some units, be they incoming or internal, have non-trivial *connections* between them, also regarded as *relations* and/or *interactions.* These are found, identified and enregistered by several algorithms, called *connectors.*

Predominant number of connections link *pairs* of units – these are depicted by coloured edges between units where the colors represent (the names of) the corresponding connectors.

A significant part of connectors search for *similarities* between different units.

A quite different groups of connectors is occupied with finding pairs (also triples and possibly, quadruples) of units that *perform together* certain functions. This "togetherness" is manifested by systematic co-appearance of the corresponding units.

Another instance of a connection/relation is that between a unit and the class assigned to it by some classifier.

After singling out these three different, yet mutually interdependent, processes, we must design algorithms implementing them where these algorithm must be universal as well as simple.

Then, granted these algorithms perform as they should, we shall be able(?) to decide if there is *some unknown "else"* within human mind crucially involved in the "learning to understand" process that is fundamentally different from formation of units, their classification and their combinatorial organisation according to their connections and interactions.

The fundamental difficulty we face here appears when we attempt to structuralize not only incoming flows of signals, but also those *created and circulating within* learning system itself, where these "internal flows" are not, at least not apparently, grounded on any structure similar to what underlies "true flows": the linear (temporal or spacial) order between signals.

The data obtained in this regard by neurophysiologists and psychologists do not tell us, at least not directly, how to proceed – we take our cues from what mathematics has to offer.

But when selecting mathematically natural algorithms, we keep in mind possibilities and limitations of their (potential) realisation by the brain: such an algorithms can not have many (say, more than 5) *consecutive operations* on each round (unit) of computation (that, roughly, corresponds to what we routinely do on 1 second time scale); yet, allowing several thousand operations running in parallel.[149]

5.10 minimization of information/entropy

sciense, virusis, vertical stick by children ???

[149]This parallelism is the "technical reason" why our basic mental (ergo)processes are inaccessible to our sequentially structured conscious minds.

5.11 Information/Prediction Profile.

The Maximal Prediction idea of Schmidthuber-Oudeyer-Kaplan-Hafne is central in our thinking on ergosystems but we emphasize "structure" instead of "behaviour", with *degree of predictability* being seen as a part of the structure of flows of signals within and without an ergosystem.

This "degree" is defined as a function in three (groups of) variables: that are

the system $\mathcal{L}$ itself and two fragment F_1 and F_2 in the flow of signals $\mathcal{FS}$, where

$\mathcal{L}$ predicts "something of F_2" on the basis of its knowledge of F_1.

This "something" refers to the result of some *reduction* procedure applied to F_2, where such a reduction may be suggested by $\mathcal{L}$ itself or by another ergosystem, e.g. by a human ergobrain.

An instance of that would be predicting a *class of a word* W in a text S on the basis of several preceding words or classes of such words. Such a class may be either syntactic, such as *part of speech*: verb/noun/..., or semantic, e.g. referring to vision, hearing, motion, an animal, an inanimate object or something else.

And "degree of predictability" of a class of a word W derived from correlations of this class with words that *follow as well as precede* W is also structurally informative.

In fact, the *proper direction*, that is "follow" versus "proceed" relation, is *not intrinsic* for (a record of) a flow of speech. But, possibly, it can be reconstructed via some *universal* feature of the "predictability (information) profile" of such a flow *common to all* languages[150], similarly (but not quite) to how the arrow of time is derived from evolution of macroscopic observables of large physical ensembles

Besides the structure in $\mathcal{S}$, this degree also tells you how competent the ergo learner $\mathcal{L}$ is, where one can also judge the ability of $\mathcal{L}$ to learn by evaluating how much this competence increases with extra information about $\mathcal{S}$ getting available to $\mathcal{L}$.

The idea of "*interesting*", that is the feature of a structure that excites "curiosity" of an $\mathcal{L}$, can be best grasped by looking at the extreme instances of *uninteresting* flows of signals – the constant ones:

●●

○○

There is (almost) nothing to predict here, nothing to learn, there is no substance is this flow for building your internal ergostructure. (If you were deprived of freedom to learn by being confined to an infinite flat plane with no single distinguished feature on it, you will be soon mentally dead; boredom cripples and kills – literally, not metaphorically.)

And random *stochastically constant* sequences do not look significantly more interesting.

○●○●●○○●●○●○○○●●○○●●●○●●●○○●●○○○○●●●○●○○●●○●○●○●●●●○●○

[150] Phonetics of a recorded speech suggests an easy solution but it would be more interesting to do it with a deeper levels of the linguistic structures. In English, for instantce, the correlation of a *short* word W_1 with a neighboring W_2 is stronger if W_2 follows W_1 rather than proceeds it; but this may be not so in other languages.

This appears "non-interesting" because one loses control over incoming signals: ours ergo idea of "interesting" is suspended in balance between *maximal novelty* of what comes and *being in control* of what happens.

(Pure randomness looks boringly uneventful to your eye but your *vestibular* and the *proprioceptive/somatosensory systems*[151] would enjoy propelling your body through a rugged terrain with occasional random jumps from one rocky stone to another making the trip enjoyably dangerous.[152])

6 Languages, Libraries, Dictionaries.

6.1 Learning a Language.

We want to implement the process of *learning a language* $\mathcal{L}$ by an *orbit* of the *universal learning program PRO* that acts on the *linguistic space of* $\mathcal{L}$ and where this orbit must eventually converge to "*I understand* $\mathcal{L}$" state/program.

The principle existence of such a program PRO is demonstrated by the linguistic performance of the brain of (almost) every child born on Earth that receives flows of *electro-chemical signals* some of which come from linguistic sources and the "meaning" of which the child's brain learns to "understand".[153]

Another, closer to our experience scenario is that of a visitor from another Universe[154] who attempts to "understand" what is written in some human "library", e.g. on the English pages of internet.

In either case, the process of what we call "understanding" is interpreted as assembling *an (ergo)dictionary* D – a kind of "concentrated extract" of the combinatorial structure(s) that are present (but not immediately visible) in flows/arrays of linguistic signals.

The grammar of a language makes a part of D where the structural position of this grammar in D is supposed to imitate how it is (conjecturally) organised in the human mind.

A particular dictionary $D = D(L) = D_{PRO}(L)$ is obtained from a collection of texts in some language, called a *library* L, according to some universal (functorial?) process/program PRO that drastically reduces the size of L and, at the same time, endows what remains with a *combinatorial structure* – a kind of "network of ideas", that is similar to but more elaborated than the structure of a *(partially directed) graph*.

This D can be thought of as (a record of) *understanding* of the underlying language by the learner behind PRO. This understanding, call it U_t, is time dependent, with D being an *approximate fixed point* of the learning process

$$U_{t_1} \underset{PRO}{\rightsquigarrow} U_{t_2},\ t_2 > t_1,$$

[151] These sensory systems tell you what the current (absolute and relative) positions, velocities and accelerations of your body and of its parts are, with most accelerations being perceived via stresses in your skeletal muscles.

[152] *Irrationality* is a hallmark of humanity. Only exceptionally, grown-up *non-human* animals are able to derive pleasure from doing something that carries no survival/reproduction value tag attached to it.

[153] Bridging linguistic signals to non non-linguistic ones is an essential but not indispensable ingredient of "understanding Language" as it is witnessed by the linguistic proficiency of deafblind people.

[154] No imaginable Universe appears as dissimilar to ours as what the brain "sees" in the electrochemical world where the brain lives.

where, a priori, PRO can be applied to "understandings" U that were not necessarily built by this PRO.

The essential problem here is finding a *uniform/universal* representation that can be implemented as a *coordinatization* of "the space of understandings" U where a simple minded program PRO could act by consecutively adjusting "coordinates" $u_1, u_2, ...$ of $U$and where this space would accommodate incoming loosely structured flows of signals encoded by libraries as well as rigidly organised dictionary structures.[155]

Among relevant concepts and building blocks of an "understanding dictionary" and processes for assembling them we envisage the following.

Short range correlations,[156] *segmentation and identification/formation of units*

in flows of linguistic signals.

Memory, information and prediction on different levels of structure.

Similarities, equalities, contextual classification, cofunctionality and coclustering.

Local and non-local, links and hyperlinks.

Tags, annotations, reduction, classification, coordinatization.

Structuralization and compression of redundancies.[157]

Ability and tendency for repetition and imitation.

Fast recognition of known, unknown, frequent, significant, improbable, nonsensical.

Evaluation of degree of "playfulness" or "metaphoricity" of words, phrases and sentences.[158]

Recognition of self-referentiality.[159]

Evaluation of parameters of ability/quality of predictions:

speed, precision, specificity, rate of success, the volume of the memory and the numbers of parallel and sequential "elementary operations" employed, etc.

A program that would imitate a human conversing in a natural language and that is seen as "realistic" from the ergo-perspective must be within 10^9-10^{12} bits in length. If such a program would fool somebody like Diderot, then its level of structurality must necessarily be comparable to that of the human ergobrain and one would be justified in saying that this program *understands* what is being said.[160]

---

[155] We know that such programs are fully operational in the brains of 2-4 year old children.

[156] Relative frequencies of "events" are essential for learning a language but such concepts as "probability", "corellation", "entropy", can not be applied to languages, without reservation.

[157] The essence of understanding is not so much extracting "useful information" but rather understanding the structure of redundancies in texts. Non redundant texts, such telephone directories, for instance, do not offer much of what is worth UNDERSTANDING.

[158] Playfulness is the first manifestation of what we call "ergo" in humans and certain animals.

[159] Omnipresent self-referentiality, along with "playfulness", distinguishes languages from other flows of signals. The simplest instance of this is seen in *noun↔pronoun* linkages.

[160] Beware of ELIZA type programs that respond to everything you say by: "You are right, it is very profound what you say. You must be very intelligent".

6.2 From Libraries to Dictionaries.

The essential role of a dictionary D of a language $\mathcal{L}$, be it fully or partially assembled, is for reading and understanding texts in this language. From some moment on, the process of learning becomes an *elaborate interaction* of D with texts in $\mathcal{L}$ that is guided by *the simple* core program in PRO.

Let us limit the concept of *language* to that of *library* – a collection of written or spoken texts – recorded strings of words pictured above as ... ●●● ... The length of this may be as small as 10^6-10^7 "words in strings", something (implicitly) kept in the memory of a youngster or as large as 10^{12}-10^{13} words comprising what was ever recorded in the English language.

Even if such a library, call it $\mathcal{LIB}$, is disconnected from non-linguistic flows of signals and, being "frozen", it is not apparently interactive – this is very much unlike how it is with "true language", an ergo learner (e.g. a human child) that is run by a *universal* program would build a certain understanding U of such a library by formally manipulating strings of symbols comprising $\mathcal{LIB}$.

We think of the implementation IU of U as augmentation of texts-strings from $\mathcal{LIB}$ by their structural annotations that would include identification of *linguistic units* in texts, their *functional associations* as well as corresponding *reductions* and *clusters* of these units.

It is unrealistic at this point to develop a clear idea of the structure of U in an ergo-learner's mind/program but it is easier to think of what we call *learner's dictionary* $\mathcal{LD}(\mathcal{LIB})$, that is a kind of a "concentrated extract" of $\mathcal{LIB}$ that contains basically the same "information" as U and where $\mathcal{LD}$ by itself stands for such "extraction/concentration" process/opearation(s). But unlike U such an $\mathcal{LD}(\mathcal{LIB})$, may be written in the language of $\mathcal{LIB}$. Presence of an $\mathcal{LD}(\mathcal{LIB})$ would significantly facilitate building understanding U of $\mathcal{LIB}$ by an ergo-learner, since most of this understanding is encrypted in $\mathcal{LD}(\mathcal{LIB})$.

To get an idea, imagine yourself in a position of such a learner with $\mathcal{LIB}$ written by thinking entities that are far culturally removed from us but who have their ergobrains organized similarly to ours.[161]

Alternatively, think of $\mathcal{LIB}$ as the totality of internet pages in English and try to comprise $\mathcal{LD}(\mathcal{LIB})$ for a use by an ergo-learner, with a six year old Crog-Magnon child in mind, where $\mathcal{LD}$ must be constructed of some *universal operations* that should apply not only to $\mathcal{LIB}$ but to many other "flows of signals", such as libraries and /or records of speech of all kinds of human languages, collections of images, series of mathematical theorems and theories, lists of chess games and chess problems, etc.

Doing this, say for your native language, is mach harder than it seems: we have as little insight into the *ergo*-structure[162] of our mother tongues as fish have in the singularity structure of solutions of *Navier-Stokes equations* for motion of liquid.[163] Yet we do see, albeit rarely, such insights in textbooks written by

[161]One does not need another physical Universe for this – just think of a possible language of blind aquatic echolocating creatures.

[162]It is hard to give definition of this "ergo" but non-ergo examples are plentiful, e.g. *cats* defined as "carnivorous feline mammals" and *nouns* as "members of a class of words that typically can be combined with determiners...". Ask a Cro-Magnon child what he/she thinks of these definitions.

[163]Don't be arrogant: human mathematicians and physicists do not understand it either.

some people, such as Albert Sidney Hornby[164] and Gilbert Taggart.[165]

But unlike to how it is usually done, the a grammar of a language encoded in our $\mathcal{LD}$-dictionary must be fully expressible in structure terms of $\mathcal{LIB}$. For instance, an "explanation" of distinction between *"she reads"* and *"she is reading"* should depend entirely on combinatorial positions of such strings/phrases in $\mathcal{LIB}$, rather than on the concept of *aspect* that expresses how an action relates to the flow of time. The sole perspective on the time structure within $\mathcal{LIB}$ that we admit, is an interpretation of combinatorics of certain "grammatical forms" employed in $\mathcal{LIB}$.

Designing algorithms for making such "dictionaries" that would convey the essential grammar and semantic rules of the corresponding languages will by no means solve the fundamental ergo-learning problem, but thinking about such $\mathcal{LD}$ brings us a step closer to this goal.[166]

7 Linguistic Flows and their Structures.

7.1 Features of Linguistic Signals.

The essential attributes of "verbal signals" be they transmitted and/or perceived auditory, visually (sign languages) or via tactile channels (in deaf-blind communication) are as follows.

(1) *Fast Language Specific Clustering.* Formally/physically *different* signals, e.g. sounds, are perceived/recognised as *identical* verbal units, e.g. phonemes, words, phrases, where this is achieved within *half-second* time intervals.

The clusterization of *phonemes* (and, probably, of other, including non-auditory, basic verbal units) depends on a particular language and the mechanism of learning these clusters by children (that deteriorates with age) is poorly (if at all) understood.

Yet, abstractly speaking, this is the easiest of our problems as it is witnessed by the efficiency of (non-contextual?) speech recognition algorithms.

(2) *Formalized Division into Units.* Flows of speech are *systematically* divided (albeit non-perfectly) into *(semi) autonomous units*, where the basic ones are what we call *"words"*.

This division, that is sharper than that of signals coming from "natural sources", is based in a significant extent on *universal* principles of *segmentation* that are applicable to all kind of signals where the markers separating "segments" are associated with *pronounced minima* of the *stochastic prediction profiles* (see section ???) of signal flows, where determination of such a profile depends on structural patterns characteristic for a particular flow.

(3) *Medium and Long Range Structure Correlations.* There are more "levels of structure" in languages than in other flows of signals. This is seen, in part, in a presence of non-local "correlations" between different fragments in texts.

[164] The most famous is *The Advanced Learner's Dictionary of Current English by A.S. Hornby, E.V. Gatenby, H. Wakefield.* Its second edition, Oxford 1963, contains $\approx 2 \cdot 10^4$ word-entries and $\approx 1.5 \cdot 10^5$ sample phrases selected with unprecedented perspicacity.

[165] Taggart wrote several textbooks on French language published in Quebec.

[166] See [3] for a a superficial discussion of possible " combinatorial shapes" of $\mathcal{LD}$.

For instance, if a sentence starts with "*There are more ... in ...* " one may rightly expect "*than in* " coming next with abnormally high probability.[167]

And if "*Jack*" appears on every second page in a book and "*his eyes sparkled again*" than, you bet, "*Jack's eyes sparkled*" on the previous page.

(4) *Verbal Reduction of non-Linguistic Signals.* Many different non-verbal signals, corresponding to objects, events, features or actions may be encoded by the same word.[168] For instance, hundreds small furry felines that have ever crossed your field of vision reduce to a single "cat".

Non-verbal signals are many while their word-names are few. The use of a language replaces the bulk of the "raw memory" in the brain by a network of "*understand*" links between individual items in this memory. This is why small children visibly enjoy the process of the verbal classification/unification of "natural signals" from the "external world" as they learn to identically name different objects.[169]

(5) *Imitation, Repetition and Generation of Lingustic Signals.* Humans, especially children, have the ability to reproduce linguistic signals σ they receive, including those emitted by themselves, where, to be exact, *not signals* σ *themselves* are generated but members σ' of the *same class/cluster* as σ and where the choice of a particular *classification rule* is a most essential matter. (We shall return to this later on.)

One can hardly analyse languages without being able to generate them[170], where the language generative mechanisms – called *generative grammars* – result from the repetitive nature of imprecise imitation.

(6) *Many Levels of Self-Referentiality.* No other flow of signals, and/or human medium of communications have the propensity of self-reference that is characteristic of Language. The ergo-structures of languages contain multiple reflections of their own "selves", their internals "egos", such as

noun-pronouns pairs, allusions to previously said/written items,
summaries of texts, titles of books, tables of content, etc.

Understanding a language is unthinkable without ability of generation and interpretation of self-referential patterns in this language.

(7) *Pervasive Usage of Metaphors.* Metaphors you find in dictionaries are kind of frozen reflections of their precursors in multiple coloured mirrors of Language (where such a precursor may not exit anymore). But many metaphors are ephemeral; they appear once and never come again.

7.2 Grammars and Dictionaries.

Making a dictionary involves several *interlinked* tasks where a starting point is

Annotation &Parsing, that is identification and classification of ***textual units*** that are persistent and/or significant fragments in short strings s (say,

[167]Try: *there are more * in* on Google.
[168]This may be contrasted with the existence of *synonymous* words, but the multiplicities and significance of the latter are incomparable to the power of the verbal reduction.
[169]Children of this age are close to being ideal ergo-learners – the strive to learn and to understand is the main drive of ergo-systems.
[170]Neuronal signal generation mechanisms play an essential role also in vision: much of what you "really see" is conjured by your own brain, but the details of this process are inaccessible to us.

up to 50-100 letter-signs) as well as attaching ***tags*** or *names* to some of these fragments.

Tagging may be visualised as colouring certain fragments in texts, where these fragments and the corresponding colours may overlap. Or, one may represent an annotation by several texts written in parallel with the original one, where the number of different color-words is supposed to be small, a few hundred (thousand?) at most, with a primitive "grammar" that is a combinatorial structure organising them.[171]

One may think of such annotations as being written in strings positioned on several *levels*[172] over the original strings s, where the new tag-strings on the level l are written in the tag-words specific to this level and where the number of such l-tags (at least) exponentially fast decays with l.

An ergo-dictionary is obtained by several consecutive *reductions* or *factorizations* applied to a library of annotated texts where the resulting combinatorial structure of the dictionary is quite different from that of (annontated or not) texts.

We describe below the basic combinatorial/categorical structure of texts and libraries and briefly indicate how parsing is performed.

7.3 Categories and Diagrams of Presyntactic and Syntactic Insertions .

Deep linguistic structures display some *approximate category theoretic* features, e.g. *abridgements* may be seen as *semantic epimorphisms*, or as *functors* of a kind rather than mere "morphisms".

Then translations from one language to another come as functors between categories (*2-categories* if abridgements are regarded as functors) of languages, where the category theoretic formalism should be relaxed to accommodate imprecision and ambiguity of linguistic transformations.

But we shall be concerned at this point with the following more apparent combinatorial category-like structure that is universally seen in all kind of "flows of signals".

Let a *library* L in, say English, language $\mathcal{L}$ be represented by a collection of tapes with strings s of symbols, e.g. letters or words, written on them, where many different tapes may carry "identical" or better to say *isomorphic* strings with the notation $s_1 \simeq s_2$, with the equality notation $s_1 = s_2$ reserved to same strings in the same location on the same tape.

Let arrows $s_1 \hookrightarrow s_2$ correspond to *presyntactic insertions* between strings, i.e. where such an arrow associates a substring $s_1' \subset s_2$ to s_1, where $s_1' \simeq s_1$.

We assume our strings are relatively short, no more than 10-20 words of length: this is sufficient for describing any "library" since every 10 words long string *uniquely* (with negligibly rare exceptions) extends (if at all) to longer strings, since the total number of strings in any language is well below 100^{10} <<

[171] An annotation may include references to *non-linguistic* signals but this would contribute to one's *knowledge* rather than to one's understanding.

[172] These levels l can be regarded as numbers $0, 1, 2, 3, ...$ with $l = 0$ corresponding to strings in the original text, where the number of level is small, something between 3 and 5. But, as we shall see letter on, these levels are organised according to a structure that is not quite linear order.

n^{10} for n being the number of symbol-words in a language.[173] As for L one might thing of something with the number N of words in it in the range 10^6-10^{12}.

The resulting ***category*** $\mathcal{C}_{\hookrightarrow} = \mathcal{C}_{\hookrightarrow}(L)$ *carries the full information about* L.

DISCUSSION.

[+] Invariance. $\mathcal{C}_{\hookrightarrow}$ is invariant under the changes of "alphabets" – names of the symbols.

[++] Universality and Robustness The categorical description of languages satisfies the most essential ergo-requirement that is UNIVERSALITY.

For instance, *spoken languages* can be similarly described in categorical terms, where, unlike written languages the arrows must correspond to *approximate* insertion relations between auditory or visual patterns.

In fact, allowing *approximate* presyntactic insertions with *sequence alignments* (with a margin of error 5-10%) in place of syntactic isomorphisms between strings would enhance the *robustness* of categorical descriptions of *written* languages as well.

[∗] Non-locality. The $\mathcal{C}_{\hookrightarrow}$-description of libraries depends on comparison between strings that may be positioned mutually far away from each other in texts.

[∗∗] Long Term Memory. This comparison between strings, depends on the presence of a structurally organised, albeit in a simple way, *memory* within the learning program.[174]

REDUNDANCY AND EXCESSIVE LOCAL COMPLEXITY OF $\mathcal{C}_{\hookrightarrow}$.

[–] The full category $\mathcal{C}_{\hookrightarrow}(L)$ contains many "insignificant" arrows, e.g. insertions of *single letters* into ten word sentences and arrows between "nonlinguistic" strings, such as "tic stri".

This can be corrected by

allowing only **TEXTUAL UNITS** *for objects in* $\mathcal{C}_{\hookrightarrow}$.

and by

selecting a ***representative subdiagram*** $\mathcal{D}_{\hookrightarrow} \subset \mathcal{C}_{\hookrightarrow}$.

Such a diagram $\mathcal{D}_{\hookrightarrow}$ (that is a *network* of directed *arrow-edges* between *strings for vertices*) *must generate* (most of?) $\mathcal{C}_{\hookrightarrow}$ as a monoid, and also it must be "small", e.g. being a *minimal* subdiagram generating $\mathcal{C}_{\hookrightarrow}$.

(There is no apparent natural or canonical choice of $\mathcal{D}_{\hookrightarrow} \subset \mathcal{C}_{\hookrightarrow}$, but it may depend on the order in which the learner encounters texts in the library.)

[–+] Pruning and Structuralizing $\mathcal{D}_{\hookrightarrow}$. No matter how you choose $\mathcal{D}_{\hookrightarrow}$ it has *too many arrows* issuing from certain (relatively short) strings s, where the number of such arrows grows with the size of a library. Thus, in order to comply with the principles of ergo-logic, our learning algorithms must automatically reorganise $\mathcal{D}_{\hookrightarrow}$ in order to correct for this excessive branching. This may be achieved, as we shall see later on, by operations of *reduction*[175] applied to (the sets of) strings and arrows.

[173] Never mind the saying: "*there are infinitely many possible sentences in a natural language*".

[174] Conceivably, this organisation corresponds to how languages are perceived by their principal learners – 1- 4 year old children, where the $\mathcal{C}_{\hookrightarrow}$-categorical organisation of memory is the "ground level" of what we call "understanding" of $\mathcal{L}$.

[175] This is also may be called *clusterization, classification,categorisation, factorization.*

Categories $\mathcal{C}^{\hookrightarrow\downarrow}$ and Diagrams $\mathcal{D}^{\hookrightarrow\downarrow}$ of Annotated Texts.

If the texts in a library are annotated with tag-strings s' that are written on several level over original strings s, then the category with "horizontal" arrows $s'_1 \to s'_2$ is augmented by the "vertical" *position arrows* $s'' \downarrow s'$ saying that s'' lies over s', where such "mixed categories" and their representative subdiagrams are denoted by $\mathcal{C}^{\hookrightarrow\downarrow}$ and $\mathcal{D}^{\hookrightarrow\downarrow}$.

The presence of vertical arrows serves two purposes.

[1] Vertical arrows significantly *increase the connectivity* of diagrams since *a bound on the number* of tag-words on the *high levels* of annotations yields the existence of *many* horizontal (syntactic insertion) arrows between strings on these levels that were not present on the lower levels.

[2] And

the notion of a representative diagram $\mathcal{D}^{\hookrightarrow\downarrow}$
is modified in the presence of vertical arrows

by replacing many horizontal arrows issuing from lower levels strings in an annotated text by the corresponding arrows on the higher levels where the "low level information" is encoded by (inverted) vertical arrows. Thus one (partly) compensates for the excessive branching of $\mathcal{D}_{\hookrightarrow}$.

Remarks. (a) *Reconstructing* Arrow of Time. Linguistic strings are directed by *the Arrow of Time*. The category $\mathcal{C}_{\hookrightarrow}$ is unaware of this arrow but, probably, the time direction in strings can be reconstructed by some rule *universally* applicable to *all* human languages. Possibly, a predominantly *backward orientations* of self-references in texts may serve for such a rule.

(b) *Structures in Symbols.* In our categorical description (the alphabet of) the basic symbols, say letters, carry no internal structure of their own. But in reality letters in alphabets are non-trivially structured in agreement with one of the ergo-logic principles that allows no unstructured set of objects with more than three-four members in it. I am not certain what one should do about it.

(c) *Dimension of Vision.* Visual signals[176] For instance the visual are customary recorded on 2-*dimensional* backgrounds such as on photographs and/or eye retina, where the extra dimensions of depth and of time (in moving pictures) carry only auxiliary information. The morphisms $s_1 \hookrightarrow s_2$ here correspond to similarities between visual patterns s_1 and subpatterns in s_2.

But, probably, a significant part of visual perception is 1-*dimensional* being implemented/encoded by the neurobiology of *saccadic eye movements.* This suggests unified algorithms for learning to see and for learning to speak.

[176] There is a demarcation line separating visual structures of *Life* – plants, animals, humans, human artefacts, from those of *non-Life* – stretches of water, rocks, mountains. These two classes of images are, possibly, treated differently by the visual system.

7.4 Fragmentation, Segmentation and Formation of Units.

Certain fragments of incoming signals[177] e.g. particular strings of letters such as *"words"*,[178] or some distinguished regions in visual fields, such as "*perceived objects* or "things"[179] qualify as *textual units*.

One can hardly give a comprehensive definition of such a unit, or a *signal-unit* in general, but one may indicate the following essential feature common to most units.

Probability of encountering a unit u among a multitude of other signals in the same category as u (here "category" means *class*) is *significantly greater than the product of probabilities* of *"disjoint parts of u"*.

For instance the word "probability" that has 11 letters in it may, a priori, appear only once or twice in a library with billion books (<< 26^{11} letters) in it,[180]; but in reality it appears millon-fold more often than that.

This does not work quite so nicely for short words: scrabble dictionaries offer ≈1000 three-letter English words and ≈4000 four-letter words where many of them, e.g. *qat* (an African plant) or (to) *scry* (to practice crystal gazing) come rarely, but the improbable frequency of such a word may be seen in appearance of several copies of it in a single volume, or even on the same page.

The abnormal frequency alone, however, does not define units: the string "obabili" appears at least as often as the full "probability"; thus, one has to augment the "definition" of a unit by the following

completeness/maximality condition: If a string s is a unit, then larger strings $s' \supsetneqq s$ are *significantly* less probable than s.

Segments and Boundaries. Fragmenting texts into units is naturally *coupled* with the process of *segmentation* that is introduction of *division points* that make boundaries of string-units in texts.[181]

Determination when the position d in a string S between two letters may be taken for a division point depends on the strings s "to the left" and "to the right" from d in S, where such a string, say s_{left}, being a unit is an essential indication for d being a division point.

But it may also happen that there is no such clear cut units next to d in S but there is a 20 letter string S' somewhere else in the library that contains isomorphic copies of five letter strings to the left and to the right from d and

[177]We temporarily ignore overlaps between fragments such as *"hard to see"* and *"to see it"* in the unit-phrase *"hard to see it"*. (*"Hard to"* makes a perfect "unitary uttering"; yet this is a weaker unit than *"hard to see"*.)

[178]A textual unit may be "disconnected", e.g. it may consist of two (more?) strings separated by other strings in a text. This happens, for instance, to separable prefixes in German that are moved to the end of the sentences. Also this **is** *not exceptionally rarely* **seen** in English.

[179]The rigid concept of *object-unit* will be later combined with classification/reduction and applied to "things" that come in many shapes such as words with flexible morphological forms, the human body, or to something inherently random such as an image of a tree with multiple small branches. When out eye looks at such a tree, our mind, conjecturally, sees (something like) a *branch/shape distribution law* rather than the sample of such a distribution implemented by an individual tree.

[180]The number of different books in the world is estimated at about 100 million. This seems to imply 5-50 million authors, half of whom must be among 7 billion people who live today on Earth – lots of writers around us!

[181]Boundaries of the so called "words" are marked in most written languages by white spaces while phrases and sentences are pinched between division punctuation signs. But we pretend being oblivious to this for the moment.

such that the corresponding d' is recognisable as a division point in S'. Then we may accept d as a division point in S and to use this for identifying previously unseen units in S.

The coupled Fragmentation + Segmentation is a multistage process each step of which is a part of the learning transformation PRO on a certain space of pairs $(Frag, Seg)$ that will be incorporated into the full "understanding space" later on.

This process must comply with *"please, no numbers" principle*: the program PRO we want to implement must function similarly to an infant's brain that, unlike an extraterrestrial scientist, has very limited ability of counting and of manipulating large numbers (e.g. frequencies) as well as small ones such (e.g. probabilities).

This is achieved, as we shall explain later on, by consecutive "internal fragmentation" of the process PRO itself into a network of simple processors/directories where, they all, individually, perform (almost identical) "baby operations" with the global result emerging via communication between these processors.

Syntatactic from Presyntactic. Eventually, we isolate strings (sometimes pairs of strings) that are serve as *textual units* and also we identify *significant insertions* between them that we call *syntactic insertions.*

Linguistic 2-Spaces $\mathcal{P}_{\hookrightarrow} = \mathcal{P}_{\hookrightarrow}(L)$ and $\mathcal{P}^{\hookrightarrow\downarrow}$.

Let us represent strings from a given library L by line segments of lengths equal the numbers of letters in them. Attach rectangular 2-cells to the disjoint union of all these strings, where these "rectangles" are Cartesian products $s \times [0,1]$, with s being some strings/segments of length ≥ 5 letters each and where the attachment maps are syntactic insertions from the segments $s \times 0$ and $s \times 1$ to some string segments S_0 and S_1 such that the images are *maximal mutually isomorphic* (i.e. composed of the same letters) substrings in S_0 and S_1.[182]

In fact, it is more instructive to use the maps corresponding *not to all* syntactic insertions in the category $\mathcal{C}_{\hookrightarrow} = \mathcal{C}_{\hookrightarrow}(L)$ but only to those from a *minimal* diagram $\mathcal{D}_{\hookrightarrow} \subset \mathcal{C}_{\hookrightarrow}$, that *generates* all morphisms from $\mathcal{C}_{\hookrightarrow}$ on strings of length≥ 5.

Then the resulting 2-dimensional cubical (rectangular) polyhedron $\mathcal{P}_{\hookrightarrow} = \mathcal{P}_{\hookrightarrow}(L)$ adequately encodes the library L and if L is sufficiently large, this $\mathcal{P}_{\hookrightarrow}$ carries all structure knowledge of the corresponding language $\mathcal{L}$ with segmentation into basic units – words and short phrases made visible.

If one deals with the category $\mathcal{C}^{\hookrightarrow\downarrow}$ corresponding to an *annotated* library, or with a subdiagram $\mathcal{D}^{\hookrightarrow\downarrow} \subset \mathcal{C}^{\hookrightarrow\downarrow}$, then one attaches "vertical" rectangles along with "horizontal" ones, where the horizontal rectanglers are associated to the arrows $s'_1 \hookrightarrow s'_2$ and the vertical ones to the arrows $s'' \downarrow s'$.

Branching Entropy

Extensions of a *string-unit*, s, e.g. of a word, by short units t following next after s in a library L define a *probability measure* on these t for

$$p_{\vec{s}}(t) = p_{L,\vec{s}}(t) = N_L(st)/N_L(s),$$

[182] Our ad hoc bound *length* ≥ 5 serves to eliminate/minimise the role of "meaninglessly isomorphic" substrings, (e.g. of individual letters) where the same purpose may be implemented by a natural constrain on strings and gluing maps as we shall see later on.

where $N_L(s)$ and $N_L(st)$ denote the numbers of occurrences of the strings s and of st respectively in L.

The collection of numbers $\{p_s(t)\}$ indexed by t serves as an indicator of variability of usage of s in the texts in the library L, where it seem reasonable to use not all t but only a collection T of unit-strings (words) t corresponding to roughly 10 (it may be something between 3 and 50, I guess, that need be determined experimentally) *largest numbers* among $p_{\vec{s}}(t)$.

The standard invariant of the probability space $\{p_{\vec{s}}(t)\}$ that reflects variability of p and regarded as an invariant of s is the *(one step forward) entropy*

$$\vec{e}nt(s;L) = -\sum_{t\in T} p_{\vec{s}}(t)\log p_{\vec{s}}(t).$$

Similarly, one defines $\overleftarrow{e}nt(s;L)$ via left extensions ts of s as well the corresponding invariants reflecting relative frequencies of "double extensions" of s that are t_1t_2s, t_1st_2 and st_1t_2.

Probably, such entropies (this is definitely true for more elaborate invariants of this kind) will be quite different for the strings *"birds-fly"* and *"pigs-fly*[183] while *"bats-fly"* will be close to *"birds-fly"* in this respect.

Classification of Words and Partitions into Sentences. Segmentation of texts into strings with more than 2-3 words in them is impossible without preliminary syntactic classification of *basic units* – words and short phrases. But when such classification is performed and the number n of basic units u – this n is about 10^5-10^6 in English – is reduced to much smaller number $\underline{n}$ of classes $\underline{u}$, realistically with $10 \le \underline{n} \le 30$. Then a library with N basic units in it would allow one to reconstruct the rule of formation of strings of length about $\log_{\underline{n}} N$. For instance, if we classify with $\underline{n} = 20$, then a modest library with 10^9-10^{10} basic units in it[184] gives an access to 6-8 basic unit long strings, for $\log_{20} 1.3\cdot 10^9 \approx 7$ that may allow an automatic discrimination between *admissible* and *nonsensical* strings up to, maybe, 12 words in length. Then *generation* of *meaningful strings* becomes a purely mathematical problem.

Gross Contextual Segmentation. In the spoken language, utterings are divided according to when, where and who is speaking to whom, while texts in written languages are organised into paragraphs, pages, books, topics, libraries with a similar arrangement of pages on the web.

These partition structure are essential for making a statistical analysis of languages; conversely, texts can be classified/partitioned according to relative frequencies of short range structural patterns. e.g. basic units, present in them. vspace 2mm

Non-Textual Syntatctic[185] *Units.*[186] Languages, unlike non-linguistic arrays of signals, "contains" units that are *not fragments of texts.* For instance, the groups of words, such as

[183]The two strings have comparable frequencies on Goggle

[184]There are about 100 basic units on a page, 10^4-10^5 of such units make a book, a 10 000 book library comprises $\approx 3\cdot 10^9$ units, while the world wide web may contains up to 10^{12} basic units of the English language.

[185]The word "syntactic" is understood in the present article as "characteristic of languages".

[186]*"Unit"* can be "defined" as *"everything worth giving it a name"*; it remans, in order to implement this "definition", to design a program that would understand what "worth" is.

{yes, no, maybe }, {we, us, our}, {big, large, huge}, {smelly, tasty, crunchy}. are kind of "outlines" of such units. We shall explain later on automatic processes for locating of these and other "higher order" units, in texts and properly incorporating them into dictionaries.

7.5 Similarity, Coclustering, Classification and Coordinatization.

Organizing multi-branched (hierarchical) classifications of units is essential for developing "understanding dictionaries". where the resulting classes become "higher order" units.

Even the basic units – *the words* come up as *equivalence classes of strings containing these* words, rather than as the mere "spell-strings". For instance, the two collections of strings

[**bats-eat**]: *bat-with-flapping-..., bats-from-..., bats-are-present-..., vampire-bat, bats-catch-..., innoculation-of-bats, bat-captured...*

[**bat-hits**]: *training-bat, used-bats-on-sale, made their own bats, increase-your-bat-..., throws-his-bat, ...-bats-per-game, raised-a-bat...*

represent two different "bat" class-words.[187]

Our goal is formulating a *universal* classification rule(s) *a priori, applicable to all kinds of strings* (and, desirably to differently structured signals) that would essentially agree with the above division of "*bats*" into two classes.[188]

Classifications are often (but not always) achieved by means of *similarity* and/or *equivalence* relations R that, besides *similarity* and *equivalence*, reflect the ideas of

"sameness", "identity", "equality", isomorphism", "analogy", "closeness", "resemblance",

where such relations R are regarded as *higher order units* and are themselves subjects to further classification.

And not all similarities lead to what we call "classification", essentially, because the "equivalence axiom" $A \sim A$ is unacceptable in ergo-logic. (If the space in your head is filled with such staff you are brain-dead.)

In fact, similarity concepts S that are applicable only to small groups of objects, such as what brings together

{sweet, bitter, salty, sower, tangy},

and that do not *meaningfully*[189] extend to majority of words, *unite* their respective S-similar members rather than *divide* non-similar ones into classes.

Another kind of groups of words having much in common that may or may not be regarded as true classes are those of *morphological word forms* such as

[187]Non-existence of the string "*bats eat and hit*" shows how far apart the two classes are but ambiguous strings such as *"hit by a flying bat"* effectuate "linguistic bridges" between the two classes.

[188]There are more – about a dozen – different class-words spelled "bat", that are, essentially, subclasses of [**bat-hits**].

[189]The common idea of "meaning" is inapplicable *within* ergo logic but *ergo-meaningless* formalism, such as $A = A$, is non-acceptable either.

$$\{works,\ worked,\ working\}$$

or

$$\{white,\ whiteness,\ whiten.\}$$

On the other hand, traditional parts of speech: *verb, noun, adjective*,..., etc. represent typical *classes of words*; also division of words into "common" and "rare" is essential despite being ambiguous.

Certain type of classes (*categories*, as they are called by linguists) are often hard to extract from the depth of a language, but interesting and often unexpected results may be achieved by applying *universal classification schemes* to languages.

Basic Example: Coclusterization principle.[190] Two units u_1 and u_2 are regarded as *similar* and/or brought to the same class/cluster if they *similarly interact* with units v_1 and v_2 in-so-far as v_1 is *similar* to v_2.

To see how this seemingly circular "definition" works let, for instance, u and v be words that are regarded as *"interacting"* if v *often*[191] *goes next after* u. If we have about 100 000 words to work with and "often" means *"at least ten times"*, then such an "interaction" is described by a $U \times V$ matrix R with $\{yes, no\}$ entries of size $10^5 \times 10^5$, and a reliable evaluation of these entries needs a library of about $10^{11} = 10 \cdot 10^{10}$ words in it.[192]

But it may happen, and it does often (albeit approximately) happen in "real life", that this huge matrix is (approximately) determined by something mach smaller, say by a 300×300 matrix, where you need only 90 000< 10^5 entries to fill in and for which a a 10^6-10^7-word checking would be sufficient.

Namely, think of R as a $\{yes, no\}$ valued function in two variables, $R = R(u,v)$, and *conjecture* that

• there are *reduction maps* $\underline{f} : U \to \underline{U}$ and $\underline{g} : V \to \underline{V}$ where the cardinalities of the sets $\underline{U}$ and $\underline{V}$ are ≤ 300;

• there exists a $\{yes, no\}$ valued *reduced relation (function)* $\underline{R} = \underline{R}(\underline{u}, \underline{v})$, such that

$$R(u,v) = \underline{R}(\underline{u}, \underline{v}) \text{ for } \underline{u} = \underline{f}(u) \text{ and } \underline{v} = \underline{g}(v).$$

Observe, that the existence of $\underline{R}$, $\underline{f}$ and $\underline{g}$, a priori, is *extremely unlikely* even if $R(u,v)$ is only approximately equal $\underline{R}(\underline{u}, \underline{v})$ and if such (even approximate) $\underline{R}$, $\underline{f}$ and $\underline{g}$ are found they would be *unique with a overwhelming probability.*

Also notice that description of R by means of $\underline{R}$ needs only

$$90\ 000 + (2 \log_2 300) \cdot 10^5 < 2 \cdot 10^6 \text{ bits}$$

instead of the original 10^{10} bits.

The above may be generalised and developed in a variety of directions, made more precise, more detailed, more specific (we shall do this later on). But the following question remains:

What can one do, say, with functions $R(u_1, u_2, u_3)$, where each u-variables may take $\approx 10^6$ values (that makes 10^{18} u-triples) and where no kind of $\underline{R}$-reduction is available?

[190] It is hard to trace the origin of this idea without knowing how it was christened at birth.

[191] This is, purposefully, formulated ambiguisly. But within the range of possible "often" there is a "coclusterization stability region" that allows a resolution of this ambiguity.

[192] If you check one word per second - eight hours a day - five days a week, it will take more than 10 000 years to go through such a library.

Our pessimistic answer is "*In general, nothing*": the human mind will be unable to understand the structure of such an R, unless... a miracle happens: somebody discovers a hidden regularity in such an R something like what we call "a law" in physics.

Biclustering for Words ×Contexts. We reserve the word "biclustering" to the case of functions R in *two* variables as in the above $U \times V$, where we want to look now at another kind of example where $u \in U$ are *words* while $v \in V$ are *books* and the function R encodes presence/absence of a u in v. Here biclusterization serves to classify books by topics according to their "key words" while the words themselves become classified by configurations of subjects, such as: chemistry of plants, animal foods, etc.

Such a clusterization may be also applied, besides $R(u,v)$, to the entropy function $\vec{e}nt(u;v)$ defined in the previous section and it may go along with that for pairs of words (u,u') according to their *systematic appearance in the same book* and/or with *tri-clusterizatuion* for $R(u,u',v)$ encoding the *insertions of words u and u' in the book v*. The resulting three classifications may be different and they must be incorporated into different facets of the "I understand" program/structure.

Iterative Clustering. Coclusterization may be computationally hard in general, but the following simple algorithm for biclustering words illustrates a possible resolution of this problem.

Evaluate the above $R(u,v)$ for v taken from the set of 100 most common words in English. Observe that this will need a modest library only with 10^8 words ($\approx$ 1000 books) in it.

Represent the function $R(u,v)$ by a map, say R_* from words u to $\{yes, no\}$ functions $r(v)$ for

$$u \overset{R_*}{\mapsto} r(v) = R(u,v).$$

Now, let us endow our space of 2-valued functions on the 100-element set by some metric, where the simplest one is *the Hamming metric* and clusterize words u using the induced metric on them. (Assume, some unambiguous clusterization exists.) Take this for the preliminary classification/reduction of words u.

Suppose, you thus divided words into 3000 classes $\underline{u}$. Select 300 most common $\underline{v}$ among $\underline{u}$ and apply the above procedure to $\underline{u}$ (that would need less than 100 books), Then, after possibly doing it yet another time, you will arrive at the desired classification.

Trees and Coordinates. These are two most common classification structures in life , and, typically classifies contain both trees and coordinate in them that are presented as **1** and **2** below.

1: Classification as a Tree. This is may be seen as a sequence of partitions of units u into smaller and smaller classes, where the rule defining the $Part_i$-class of a unit u depends $Part_{i-1}$ of u.

A linguistically rather artificial instance of that, is classification/positioning of words in alphabetically organised dictionaries.

More significant example is where $Part_1$ divides words into the two classes:

A. Class of content words: {*nouns*, (most) *verbs*, *adjectives*, *adverbs*.}

B. Class of function words: {*articles, pronouns, prepositions, etc.*}

And $Part_2$-classes are obtained by further subdividing words into "parts of speech".

2: Classifications by Coordinates. These are given by several *coordinate functions* $c_i(u)$, where determination of $c_{i_0}(u)$ is essentially independent of $c_{i \neq i_0}(u)$ and where the set $I \ni i$ is not necessarily ordered. Different classes are formed by assigning particular values to some coordinates.

For instance, one may have the following functions $\mathbf{c}_1$,$\mathbf{c}_2$, $\mathbf{c}_3$,$\mathbf{c}_4$ on phrase-units u.

$\mathbf{c}_1(u)$ takes values *long*, *medium*, *short* depending on whether u has at most 4, between 5 and 8 or more than 8 word-units in it.

$\mathbf{c}_2(u)$ takes values *yes* or *no* depending if u contains a *content verb* in it.

$\mathbf{c}_3(u)$ assigns the *key word* w in u to u.

$\mathbf{c}_4(u)$ is the expected minimal age (in years) of a child able to understand the phrase u.

We expect our program *PRO* will be automatically generating this kind of functions $\mathbf{c}(u)$ in the course of learning a language.

8 Combinatorics of Libraries and Dictionaries.

We want to represent languages and dictionaries by *multicolored networks U* as follows.

• *Linguistic units u* are represented by *nodes* or *vertices* in such a network.

• *Connections between units* are depicted by *connective edges* between these nodes. We do not exclude a possibility of *several* (or none) connectives joining some nodes.

• Besides connectives, there are *directed edges* between certain nodes; these are depicted by arrows, such, for instance, as *syntactic insertions* between strings and *reductions* $u \mapsto v = class(u)$. A relevant feature of such arrows is that some of them are *composable* as in $u \mapsto v \mapsto w$.

• Nodes and edges carry certain *colours* where such a colour depicts "essential properties" of the node or edge it is assigned to. For instance a colour on a node u may say: "***short string***", a colour on a connective may say among other things, "***similarity***" and a colour on a reduction arrow may refer to a particular classifier algorithm defining/producing this arrow.[193]

From another perspective, colours appear as *descriptors* of *structural/logical units* involved in organisation/construction networks U. The ensembles D of such descriptors are smaller than U:

a dictionary may have 10^5-10^8 nodes (libraries are even larger) while the number of colours is in U, that is the number of units in D, is somewhere between 50 and 500.

But multicoloured network-like structures carried by these D are more sophisticated than the structures in U. (We shall explain this later on.)

COLLECTIONS, ENSEMBLES, SETS.

[193]One should think of a colour as a simple combinatorial entity, e.g. a little tree, rather than a word or phrase.

We *do not* regard *collections/ensembles* of nodes, and even less so of edges, as *mathematical sets* for the following reasons.

1. The presence of a particular node in a network, e.g. of a particular phrase, in the long term memory of a learner is often ambiguous.

2. Basic set theoretic constructions, such as the union $X_1 \cup X_2$ and the Cartesian product $X_1 \times X_2$, can not (and should not) be unrestrictedly performed in our networks.

The set theoretic language may lead you astray;[194] yet, we use fragments of this language whenever necessary.

Colouring Descriptors. A conceptually most difficult problem in building network models of learning[195] is assigning colours to descriptors and connectives between them. We think of these colours as *formalised expressions* of (simple combinations of)

fundamental/universal principles of learning.

Identification and formalisation of these principles is our main task.

8.1 Library Colours.

What we call *a library L* is a a collection of *string-units*, where such a string may be, a priori, anything starting from a fragment of a word to a paragraph with a few dozen words in it.

These strings are "colored" according to their size, say into three basic[196] "colours": ***short, median, long***, where each of these colors may be "subdivided" into three "subcolors":

short$_{short}$, ***short***$_{median}$, ***short***$_{long}$, ***median***$_{short}$, ..., ***long***$_{long}$[197].

(Later on, this structure will be *parsed* by identifying "significant strings" , such as ***word***s, ***phrase***s, etc. and disregarding insignificant ones, with possible subcolors such as ***word***$_{rare}$, for instance.[198])

There are several facets to a *learner's perception* of *geometric structure*(s) that underlines L and that is involved in formation of string-units and which is also needed for defining connectives between strings.

Simple Linear Order. A mathematician may be conditioned to think of a library as a string of symbols indexed by integers $i \in \mathbb{Z}$ with all of geometry of L derived from what he/she knows of $\mathbb{Z}$:

... ••• ...

But the "naive concepts" in the mind of our learner – think of a baby learning to talk – are more general, more powerful and more flexible, something like:

close-one-to-another, ***far-one-from-another***, ***next-to-each-other***, ***in-between***, ***begins-with***, etc., where these "colourful concepts" come in sev-

[194]Bringing forth *random sets* and/or *fuzzy set* may only aggravate the problem.

[195]Possibly, our kind of networks have little to do with "true learning", but one can not rule them out at the present stage.

[196]Digital number bases *two* or *four* are also possible(?) but *three* is preferable.

[197]These can be depicted as nine leaves of a rooted (and ordered) triadic tree with 12 edges and with a single colour ***length*** spread all over it.

[198]An essential (but not the only) intrinsic motivation for doing this by an ergo-learner is economising the memory space.

eral subcolor-flavours similarly to (yet, differently from) how it is with lengths of strings.

Notice that all of these are binary relations except for the ternary ***in-between***.[199]

Besides these relations the large scale geometry in L is reflected in the presence of (relatively large) *contextual units* such as ***page***s and ***book***s, for instance.[200]

Then *closeness* between two strings can be seen as *simultaneous containment* of these *string-units* in the same *context-unit*.

This is essential for enlisting and keeping in memory (pre)syntactic insertions between strings, in particular all pairs of identical words w in L. The number of such pairs is uncomfortably large being *quadratic*[201] But identifying identical words say on a single page goes *linearly* in time.

Among colours carried by *insertions* of string-units into context-units we indicated here only the two: ***frequent*** and/or ***rare***. These are used, both for identification of (statistically homogeneous) context units as well as for classification units with a use of biclusterization.

It is not difficult to complete the above combinatorial description of our library L – one needs a few dozen more colours, about 50 of them all-to-gether, that represent *basic types* of units and of connectors between them in L.

But the principal issue is not so much L per se but a construction of the corresponding descriptor network D on the basis of few, probably 4-8, *"general rules"*, where, recall, the colours-descriptors of L are taken for nodes in D and *"general rules"* are used as colours assigned to nodes and connective edges in D.

9 Appendices

9.1 Ergo Perspective on Chess

> *Arithmetical or algebraical calculations are, from their very nature, fixed and determinate....* [But] *no one move in chess necessarily follows upon any one other.*
>
> Edgar Allan Poe, "Maelzel's Chess-Player", April 1836.

In the early 19th century, when Poe was writing his article on *Maelzel's Chess-Player Automaton*, an ability to play chess was seen by many (all?) as

[199] ***"begins-with"*** *formally* defines simple linear order that implies all other relations. But our baby-like learner is unaware of formal logic, for him/her these "colorful-relations", albeit interconnected, do not "logically imply" one another. If a human baby were born with a logical mode of thinking in his/her head, he/she would learn preciously little of essence in this world.

[200] Originally, these units are classified/coloured by their size, where different classes must roughly fit into the corresponding frames of the *short-term, medium-term* and *long-term memory.* Then the concept of "context" is modified and refined in the course of learning (not quite) similarly to how it happens to strings, where the true ***page***s and ***book***s must be either sufficiently *statistically homogeneous*, or *structurally unified* or to have *pronounced boundaries.*

[201] Squares are unacceptable except of small quantities. We happily live through *million* seconds that make less than 12 days of our lives. But *trillion seconds*, that is million squared, stretch over more than 31 000 years.

a quintessential instance/measure of the intellectual power of the human mind. But the *mere existence* of *chess algorithms* is obvious.[202]

You play white. Let $\mathbf{eva}_0(P^*)$, where $* = \circ$ or $* = \bullet$, depending on whether a move by white($\circ$) or by black($\bullet$) is pending, be a "natural" numerical[203] *evaluation function* of a position P^*, e.g. the sum of judicially assigned weights to the pieces – positive weights to the white pieces and negative to the black ones.

Inspect possible *white* moves wh for all P°, denote by $P^\circ + wh$ the resulting new positions, and similarly consider all $P^\bullet + bl$. Define new evaluation function $\mathbf{eva}_1(P)$ by

$$\mathbf{eva}_1(P^\circ) = \max_{wh} \mathbf{eva}_0(P^\circ + wh)$$

$$\mathbf{eva}_1(P^\bullet) = \min_{bl} \mathbf{eva}_0(P^\bullet + bl).$$

Keep doing this and get

$$\mathbf{eva}_0 \Rightarrow \mathbf{eva}_1 \Rightarrow \mathbf{eva}_2 \Rightarrow \mathbf{eva}_3 \Rightarrow ... \Rightarrow \mathbf{eva}_N \Rightarrow ...$$

Stop, say, at $N = 20$ and let your computer (that plays white) maximize $\mathbf{eva}_{20}(P + wh)$ for all its moves wh. Such a program, probably, will beat anybody, but... no computer can inspect twenty half-moves in realistic time.

As recently as in 1950's, Hubert Dreyfus, a critic of artificial intelligence believed that a child would beat any chess program.

In 1957, Dreyfus was defeated by a $\mathbf{eva}_2$ chess program that was *implemented on a computer* by Alex Bernstein and his collaborators.[204]

Fourty years later in 1997, Deep Blue (non-impressively) defeated the wold champion Kasparov, 3.5-2.5. The computer could evaluate 200 million positions per second; it inspected, depending on the complexity a position, from $N = 6$ to $N \approx 20$ half-moves. The program contained a list of endgames and it adjusted the evaluation function by analyzing thousands of master games.

So, when Poe insists that

... no analogy whatever between the operations of the Chess-Player, and those of the calculating machine of Mr. Babbage,

one might judge him as mathematically naive; yet, Poe's conclusion was fully correct.

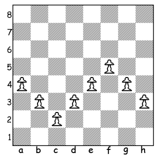
It is quite certain that the operations of the Automaton
are regulated by mind, and by nothing else.
... this matter is susceptible of a mathematical demonstration, a priori.

The idea behind what Poe says is valid: Turing(Babbage) machines and **eva**-algorithms make poor chess players – they can match ergobrain programs *only* if granted superhuman resources of computational power.

[202] This, must(?) have been understood by Wolfgang von Kempelen, the creator of Chess-Player, and by contemporary mathematicians and scientists, e.g. by Benjamin Franklin who played with this "automaton". But I could not find a reference.

[203] This is *unnatural*, there is nothing intrinsically *numerical* in chess. Logically, what we need for an evaluation is (somewhat less than) an *order relation* on positions. But "ergo-evaluation" is more subtle and less logical.

[204] In 1945, the *first*(?) chess program was *written* by Konrad Zuse in his *Plankalkül* – a high-level programming language.



This does not preclude, however, an approach with a quite different, possibly yet unknown to us, (ergo?)mathematics, but some people conjecture that the human (ego?)mind is "fundamentally non-algorithmic".

In his book *Shadows of the Mind*, Roger Penrose, who opposes the idea of *thinking machines,*[205] presents a chess position where

> *black has eight pawns, while white, in addition to eight pawns, has two rooks (and the white squared bishop, if you wish). The black pawns stay on black squares in an unbroken chain (as in the above drawing) that separates the black king from the white pieces. White pawns are positioned in front of the black ones and fully shield them from the rooks.*

Thus, the black king is safe in-so-far as black pawns do not change positions. But if a black pawn captures a white rook, then the chain of the black pawns will be disrupted and the black king *eventually* mated.

Any current computer-chess program would accept a sacrifice of a white rook, since "eventually" shows only in another twenty-thirty half-moves, while no human player will make such a silly mistake.

But Doron Zeilberger, who fights against the *Human Supremacy* idea, insists[206] that

> *symbol-crunching* [program], *valid for an $m \times n$ board rather than only an 8×8 board, will perform as good as a human player.*

Also, Zeilberger is critical of Penrose's use of *Gödel's incompleteness theorem* (see 2.1) as an argument against *thinking machines.*

Chess has been supplying an experimental playground to all kind of people pondering over the enigma of the human mind.

Logicians-philosophers marvel at how *formal rules* but not, say, the shape, color or texture of the pieces, determine what players do with them.

For example, Wittgenstein instructs (mocks?)[207] the reader:

> *The meaning of calling a piece "the king"*
> *is solely defined by its role in the game.*

He continues –

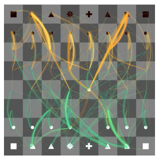
> *imagine alien anthropologists landing on planet earth ...*
> *discover ... a chess king...*
> [It] *remains an enigma to their understanding.*
> *Without ... other artifacts, the chess king is*
> *only a chunk of wood (or plastic, or...).*

(It is hard to resist continuing with *...or chocolate...* . But what the philosopher had in mind was not as trivial as it looks.)

[205] See more on www.calculemus.org/MathUniversalis/NS/10/01penrose.html

[206] www.math.rutgers.edu/~zeilberg/Opinion100.html.

[207] Wittgenstein is often quoted with *A serious and good philosophical work could be written consisting entirely of jokes.*



Unlike logicians, students of the *ego*mind search for the meaning of chess in apparent or hidden urges of players to compete, to win, to grab, where making checkmate for a male chess player substitutes for killing his father in accordance with *Oedipus complex.*[208]

From the *human ergobrain perspective*, the relevance of chess is seen in its *intrinsic attractiveness* to certain people[209] and the central problem of chess in designing a (relatively) simple *learning (ergo)program* $\mathcal{L}$ that would find chess *interesting* and/or will be able (and "willing") to learn playing chess by itself whenever it is given an access to the records of sufficiently many (how many?) chess positions, chess problems and/or (fragments of) chess games.

And since chess, as most (all?) ergo-activities, is *interactive*, a learning will go faster if $\mathcal{L}$ is allowed an access to computer chess programs.

We conjecture the existence of such an $\mathcal{L}$, that is, moreover will be (rather) *universal* – not chess-special in a remotest way. It may come from somebody who has never heard of chess or of any other human game. However, such a program $\mathcal{L}$, being a pure ergo, may behave differently from a human player, e.g. it will not necessarily strive to win.

Such self-taught ergolearner program implemented on a modern computer will play chess better than any human or any existing specialized computer chess program, but this is *not* the main ergo issue. And it is nether the power of logical formality – something trivial from the ergo (and from general mathematical) point of view, what makes chess attractive and interesting.

An ergolearner delights in the beauty of the *structure* of $\mathcal{CHESS}_{ergo}$, some kind of combinatorial arrangement of "all" interesting games and/or positions. An ergolearner tries to understand *(ergo)principles of chess* that transcend the formal rules of the game

These "principles" enable one, for instance, to distinguish positions arising in (interesting) games from *meaningless* positions, as it is seen in how chess masters memorize *meaningful* positions that come from actual games but they are as bad as all of us when it comes to random arrangements[210] of chess pieces on the board. In its own way, chess tells us us something interesting about *meaning.*

9.2 Polynomialas, Equations, Computations.

Formal languages do not walk on the streets occupying themselves with proving Gödel's style theorems one about another. But we humans *are* walking com-

[208]Was it intended by Freud as a macabre joke? Sphinx might have accepted this solution to the riddle of chess, but we feel more at ease with *Flatulus complex* see [SLE] §6.7 [3].

[209]If you remold the piece "king" into "sphinx", the game will not loose its attractiveness to more than half of chess perceptive people.

[210]The number of possible chess positions is estimated around 10^{45}. Probably, 10^{12}-10^{18} among them are "meaningful".

puters that are programmed, among other things, to guess and to imitate each other's mental computations.

"Computation" as it is used in the science of the brain and in science in general is a metaphor for elaborated, yet, structurally organized process. But there is no clarity with this notion.

Does, for instance, a planetary system perform a computation of, say, its total potential energy? You would hardly say so on the microsecond time scale but it may look as a "computation" if the time measured in million years.

In mathematics, there are several specific models of computation but there is no readymade language for describing all conceivable models.

Mind you, there is an accepted class $\mathcal{COMP}_{\mathbb{N}\to\mathbb{N}}$ (that parallels the class of provably defined functions) of what is called *computable* or *recursive functions*, $R(n)$ that send $\mathbb{N}\to\mathbb{N}$ for $\mathbb{N}$ being the set of *natural numbers* i.e. of positive integers $n = 1, 2, 3, 4, 5, ...$ Yet, there is no single distinguished natural description of this class as it is witnessed by the presence of *many* suggestions for its "best" description with the following five being most prominent.

Recursion + Inversion (Skolem, Gödel, Herbrand, Rózsa Péter),
λ-calculus (Church),
Turing machines and programs (Babbage, Ada Lovelace, Turing),
cellular automata (Ulam, von Neumann, Conway),
string rewriting systems (Markov).

These definitions of "computable" reflect their author's ideas on what is *"simple, useful, natural"* with the corresponding schemes of computation being quite different. None of them can be taken for the "normal" or "canonical" form of computation.[211] Besides, all these definitions of $\mathcal{COMP}$ are decades old and they have not undergone the post-Grothendieck *category theoretic* renovation.[212] But the traces of the following ideas, that underly the concept of computation, will be seen in our ergo-models.

• Compositions and Categories. *Composabilty* says that if a computation with the input from some (constructive set? class?) X_1 and output in X_2, denote it $C_1 : X_1 \rightsquigarrow X_2$, is followed by $C_2 : X_2 \rightsquigarrow X_3$, then the tautological composition $C_3 = C_2 \circ C_1 : X_1 \rightsquigarrow X_3$, where C_2 is performed right after C_1, is a computation again.[213] Thus, computations make what one calls *categories* provided *the identity* [non]-computations are in there as we shall always assume.

(Composability is a fundamental but non-specific feature of computations – almost everything you do in mathematics can be "composed" if you think about it.)

Moreover, many computation schemes operate with functions in *several variables* with arguments and/or values in a certain set X, that may be the set $X = \mathbb{N}$ of *natural numbers*, the set $X = \mathbb{Z}$ of *integers* and the set $X = \mathbb{R}$ of *real numbers.*,

[211]It is hard to argue for or against something being "natural". *Feets*, *meters* and *miles* may seem natural physical units of distance for some people. But, probably, there is *neither a truly canonical* normal form *nor a convincing mathematical* concept of equivalence applicable to different models of computation.

[212]Computations are not bound to $\mathbb{N}$ but I am not certain if there is a proper definition of "computable objects", e.g. sets with some structures representing suitable functors, in (yet, hypothetical) "computable categories".

[213]There is no consensus for writing $C_2 \circ C_1$ or $C_1 \circ C_2$. Although the *Zermelo Buridan's ass axiom* allows a choice of one of the two, remembering which one is impossible – how can one tell "←" from "→" in a symmetric Universe?

where "complicated" computable functions are successively built from "simple modules" by composing these "modules".For instance, the following four functions:

two one variable functions: *the constant* $\mathbf{x} \mapsto \mathbf{1}$ and *the identity* $\mathbf{x} \to \mathbf{x}$
and two functions in two variables:
subtraction $(\mathbf{x}_1, \mathbf{x}_2) \mapsto \mathbf{x}_1 - \mathbf{x}_2$ and *multiplication* $(\mathbf{x}_1, \mathbf{x}_2) \mapsto \mathbf{x}_1 \cdot \mathbf{x}_2$

generate, in an obvious sense, all *polynomials* that are sums of products of constants (coefficients) and powers of variables:

$$P(x_1, ..., x_k) = \sum_{d_1,...d_k \le D} a_{d_1,,...d_k} x_1^{d_1} \cdot ... \cdot x_k^{d_k}.$$

with integer coefficients $a_{d_1,...,d_k}$ for all $k = 1, 2, ...$ and all degrees $D = 0, 1, 2, ...$

SUPERPOSITIONS, CLONES, MULTICATEGORIES, OPERADS.

The algebraic skeleta of sets of functions in several x-variables closed under compositions, also called *superpositions* in this context, go under the names of *"abstract clones"* in mathematical logic and universal algebra and and/or "*operads*" in algebraic topology. More generally, if the domains and ranges of maps are *not* assumed to coincide, than one speak of *multicategories.* These, similarly to ordinary categories, are described in terms of of classes of *diagrams of (multi)arrows* that mimic the obvious associativity-like properties of superpositions of functions.

The operad structures underly neural networks models of the brain. They will be also present in our ergosystems, where we shall insist on assigning specific structures to what goes under the heading "several" and/or "multi". (A newly born ergobrain does not know what the set $\{1, 2, ..., k\}$ is and it can not operate with functions presented as $f(x_1, x_2, ..., x_k)$).

• INVERSIONS. Inverting a function $y = P(x)$, that is finding x that satisfy the equation $P(x) = y$ for all y in the range of P, may be frustratingly difficult even for simple function $P : X \to Y$. An instance of this is computing (the integer part of) $\sqrt{y}$ for integer y that is much harder than taking $y = x^2$.

In general, the inverse map P^{-1} sends a point $x \in X$ *not to a single point* but to a (possibly empty) *subset* called $P^{-1}(x) \subset Y$, namely, to the set of those $y \in Y$ where $P(y) = x$. But a composition of such P^{-1} with some map $Q : Y \to Z$ may be a bona fide point map denoted $R = Q \circ P^{-1} : X \to Z$. This happens if $P^{-1}(x)$ is *non-empty* for all $x \in X$, i.e. P is *onto*, and if Q is constant on the subsets $P^{-1}(x)$ for all $x \in X$. In this case

R equals the *unique solution* of the equation $R \circ P = Q$.

Thus, extensions of classes of maps by adding such inverses may be described in category theoretic terms as follows. Let $\mathcal{P}$ be a subcategory of a category $\mathcal{S}$, e.g a class of maps p between sets that is closed under composition.

The invertive extension $\mathcal{R}$ of $\mathcal{P}$ in $\mathcal{S}$ is, by definition, obtained by adding to $\mathcal{P}$ the solutions $R \in \mathcal{S}$ of the equations $R \circ P = Q$ for all P and Q in $\mathcal{P}$ whenever such a solution exists and is unique.

(This $\mathcal{R}$ may be *non-closed* under composition of morphisms and it can be *enlarged further* by generating a subcategory in $\mathcal{S}$ out of it.)

Such an extension may be incomparably greater than $\mathcal{P}$ itself, where the BASIC EXAMPLE of this is as follows.

Let $\mathcal{S}$ be the category (*semigroup* in this case) of functions $\mathbb{N} \to \mathbb{N}$ for $\mathbb{N} = \{1, 2, 3, 4, 5, ...\}$ and let $\mathcal{P}$ consist of *primitively recursive* functions.

Then the invertive extension $\mathcal{R} \subset \mathcal{S}$ of $\mathcal{P}$ equals the set of all recursive (i.e. computable) functions $\mathbb{N} \to \mathbb{N}$.

"Primitively recursive" is a currently accepted formalization of "given by an explicit formula". Such formalizations and the resulting $\mathcal{P}$ may be somewhat different but the corresponding $\mathcal{R}$ are all the same.

A convincing instance of this is

DPRM THEOREM.[214] *The invertive extension $\mathcal{R}$ of the subcategory $\mathcal{P}$ of polynomial maps $\mathbb{N}^k \to \mathbb{N}^l$, $k, l = 1, 2, 3, 4, 5, ...,$ in the category $\mathcal{S}$ of all maps equals the category of recursive (i.e. computable) maps $\mathbb{N}^k \to \mathbb{N}^l$.*

In other words,

every computable function $R : \mathbb{N} \to \mathbb{N}$, can be decomposed as $R = Q \circ P^{-1}$, where

- *$P, Q : \mathbb{N}^k \to \mathbb{N}$ are integer polynomials,*
- *the map $P : \mathbb{N}^k \to \mathbb{N}$ is onto,*
- *Q is constant on the subsets $P^{-1}(n) \subset \mathbb{N}^k$ for all $n \in \mathbb{N}$.*

(Moreover, there is a universal bound on k, e.g. $k = 20$ suffices.)

The way the theorem is proven allows an explicit construction of polynomials P and Q, e.g. in terms of a Turing machine (defined later on) that presents an R. For instance, these P and Q can be actually written down for the *nth prime number function* $n \mapsto p_n$.

However, this theorem does not and can not shed any light on the structure of prime numbers[215]. All it shows is that Diophantine equations, that make a tiny fragment of the world of mathematics, have, however, a capability of "reflecting" all of $\mathcal{MATH}$ within itself: any given (properly formalized) mathematical problem Π can be translated to the solvability problem for such an equation. This, in conjunction with Gödel's theorem, tells you that

solvability of general equations $P(x_1, x_2, ..., x_k) = n$ is an intractable problem.

There is a host of similar theorems for all kinds of (not at all Diophantine) "simple equations" that make mathematics, seen from a certain angle, look like a fractal composed of infinitely many "Gödel's fragments" where each "fragment" multiply reflects $\mathcal{MATH}$ as a curved fractal mirror with every reflected image of $\mathcal{MATH}$ being transfigured by a chosen translation of $\mathcal{MATH}$ to the language of this "fragment".

A translation of a "general difficult problem" Π to a "concrete and simple" equation whenever such a translation is available by a DPRM kind of theorem, does not help solving Π but rather shows that an apparent simplicity of the corresponding class of "equations" is illusory with Gödel's theorem guarding you from entering blind alleys of naive solvability problems.[216]

For instance, the solvability problem for a Diophantine equation $P(x_1, ..., x_k) = 0$ transforms by a particular translation algorithm ALG_{part} built into a given

[214] Conjectured by Martin Davis (Emil Post?) in 1940's and finalized by Matiyasevich in 1970 following Davis, Putnam and Robinson.

[215] Probably, *nothing what-so-ever* about prime numbers can be seen by looking at such P and Q, not even that there are infinitely many of primes.

[216] Even without Gödel, anything as easy to formulate as the solvability problem makes one wary, be these Diophantine or other kinds of equations.

proof of DPRM to the solvability problem for an equation $P_{new}(x_1, ..., x_l) = 0$ with the (integer) polynomial P_{new} being by far more complicated (i.e. with larger coefficients) than the original P, where it is virtually impossible to reconstruct P back from P_{new} even if you know ALG_{part}.

The DPRM theorem itself was a response to David Hilbert who suggested in his 10th problem:

to devise a process according to which it can be determined in a finite number of operations whether the equation is solvable in rational integers.

This idea of a possible effective resolution of all Diophantine problems was in line with Hilbert's (pre-Gödel) optimistic:

Wir müssen wissen – wir werden wissen!

[We must know – we will know!)]

(Hilbert also articulated this position is his 2nd problem: *a direct method is needed for the proof of the compatibility of the arithmetical axioms.*)

But the DPRM theorem showed that Hilbert's suggestion taken literally was unsound and, if followed, must be coupled with a search for *particular classes* of equations where integer solutions are well structurally organized.

The "true Diophantine beauty", as we see it to-day, resides not in integer solutions of $P(x_1, ..., x_k) = 0$, but in *non-Abelian higher dimensional "reciprocity laws"* associated to integer polynomials P. Roughly, such laws can be seen as *analytic relations* between infinitely many numbers $N_p(P)$ for *all prime* $p = 2, 3, 5, 7, 11, 13, 17, ...$, where $N_p(P)$ equals the number of solutions of the congruence $P = 0 \mod p$.

Such relations are expected to generalize *Riemann's functional equation*

$$\frac{\zeta(1-s)}{\zeta(s)} = \frac{\alpha(s)}{\alpha(1-s)},$$

where

$$\zeta(s) = \prod_p \frac{1}{1-p^{-s}} \quad \text{and } \alpha(s) = \frac{1}{2}\pi^{-s/2} \int_0^\infty e^{-t} t^{\frac{s}{2}-1} dt \quad \text{for } s > 1,$$

where both functions, ζ (that harbors the deepest mysteries of prime numbers) and α (an apparently insignificant child of simple minded analysis) admit meromorphic extensions to all complex s-plane and where the thus defined functional equation, applied at different s, encompasses infinitely many relations between the prime numbers $p = N_p(P)$ for $P = P(x_1, x_2) = x_1 - x_2$.

The above is just a hint at what is known as the *Langlands program*[217] that predicts a presence of unexpectedly strong and simple structural constrains (laws) that are satisfied by quite general and complex objects like the above P and that is opposite to the spirit of the DPRM style theorems where one is keen at exhibiting special and apparently simple objects that display an arbitrarily complex behaviour not constrained by any "law".

The lesson we draw from the "Diophantine story", where properties P of *unknown* objects x are expressed by algebrac equations, is that

identifying essential properties P of an x and formulating structurally significant questions about these P is more instructive than straightforward attempts to construct x.

We believe, this as much applies to yet unknown objects x that mathematically represent *thinking and learning processes* as to k-tuples of integers $x = (x_1, ..., x_k)$.

Networks behind Formulas.

Arithmetic operations such as $x_1 + x_2$ and $x_1 \cdot x_2$ becomes progressively more and more elaborate as the numbers x_1 and x_2 grow, but they are decomposable into sequences of a few elementary operations *over the decimals* of these numbers as every schoolgirl knows.

And general "complicated computations" can be realized by networks of "elementary computational steps" with no explicit use of anything mathematically elaborate (e.g. addition and multiplication of integers) as it is done in our computers and, probably, in our brains (where "elementary steps" may go far from computers).

In fact, (almost?) all multivariable functions "in real life" come as sums of simple superpositions of functions in one and two variable such as (generalized) *tensorization* – the most common decomposition of functions into sums of products :

$$F(x_S) = \sum_{i \in I} \prod_{s \in S} f_{s,i}(x_s),$$

where

- S is a finite set that enumerates the x-variables:
- each x-variable x_s, $s \in S$, runs over a given *domain set* denoted $X \ni x_s$;
- x_S is denotes the totality of the variables x_s, that is an element of the Cartesian S-power of X:

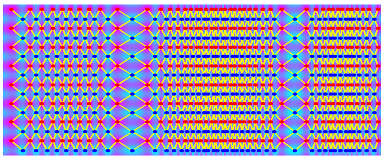
$$x_S \in X^S = \underbrace{X \times X \times ... \times X}_{S};$$

- $I \ni i$ is a finite set that enumerates functions $f_{s,i}$ on X;
- the functions $f_{s,i}$ on X and F on X^S take values in a set called Y.

In order to be able to make "sums of products" $\sum_i \prod_s$, we endow the set Y with two *binary operations* that are functions in two variables $Y \times Y \to Y$ denoted $y_1 \boxplus y_2$ and $y_1 \boxtimes y_2$ and we suppose that

Y contains two elements, called "zero", denoted $0 \in Y$ and "one" $1 \in Y$ with their usual arithmetic properties:

[217]This program goes along the lines of Hilbert's 9th and 12th problems on *the most general law of reciprocity in any number field* and on *Abelian extensions* of such fields.



$y \boxplus 0 = 0 \boxplus y$ for all $y \in Y$,
$y \boxtimes 0 = 0 \boxtimes y = 0$ for all $y \in Y$,
$y \boxtimes 1 = 1 \boxtimes y = y$ for all $y \in Y$, $y \neq 0$.

If the functions $f(x) = f_{s,i}(x_s)$ are *atomic* i.e. different from "zero" (at most) at a *single* element in X (depending on s and i) and if, for each $i \in I$, there is at most *single* $s \in S$ such that the non-zero value of $f_{s,i}$ is different from "one" (this is a kind of "multiplicative atomicity"), then the above $\sum_i \prod_s$ makes sense. Moreover, (this is obvious)

every function $F(x_S)$ that equals "zero" away from a finite subset in X^S admits a "tensorial" decomposition into atomic functions in the variables x_s.

In fact, one does not truly need $\boxplus$ as it can be obtained by composing $\boxtimes$ with the one variable function $y \mapsto y^\perp$ that interchanges $0 \leftrightarrow 1$ and keeps all other $y \in Y$ unchanged. Indeed, the operation

$$y_1 "\boxplus" y_2 =_{def} (y_1^\perp \boxtimes y_2^\perp)^\perp$$

does satisfy the above properties required of $\boxplus$.

Polynomials in real variables $x_s \in \mathbb{R}$ are most common examples of such "tensors", but computationally/logically the simplest case is where $X = Y = \mathbb{F}_2 = \{0,1\}$, that is the *field of integers* mod 2 (where $1+1 =_{def} 0$), rather than $\mathbb{R}$. Here, observe, (this equally applies to all finite fields $\mathbb{F}_{p^n}$)

every atomic $\mathbb{F}_2$-valued function on $\mathbb{F}_2$; hence, every $\mathbb{F}_2$-valued function on $\mathbb{F}_2^S$, is representable by a polynomial.[218]

The architecture of tensorial representations of functions is clearly visible. For instance, there are $2^{2^{card(S)}}$ different polynomials over $\mathbb{Z}_2$ in $x_s, s \in S$, that correspond to subsets in the power set 2^S.

But general representations of functions by a superposition of two binary operations, say of "+" and "×" are combinatorially more elaborate: these are *formulas*, such as

$$((*o(*o*))o*)o(((*o*)o(*o*)o*)),$$

where

$*$ substitutes for x_s-variables;

and

o stands either for + or for ×.

These formulas express computations as *strings* in four types of symbols:

$$* \quad o \quad (\quad)$$

[218]One can *discard multiplication* in writing down polynomials over the fields $\mathbb{F}_{p^n}$ for the primes $p \neq 2$, because $x_1x_2 = \frac{1}{4}\left((x_1+x_2)^2 - (x_1-x_2)^2\right)$. This does not work for $p = 2$, e.g. in $\mathbb{F}_2$ where, instead, one can discard addition by expressing it as a superposition of the functions $x \mapsto x+1$ and $(x_1,x_2) \mapsto x_1 \cdot x_2$.

but depicting computation by *one dimensional strings* is an artifact of the way we speak and write[219] that does not faithfully reflect the geo-combinatorial structure behind such formulas.

The essential reason for his is that the *rule of brackets* that restricts admissible configurations of right and left brackets symbols ")" and "(" is *non-local in the string geometry*[220] unlike, for instance, the prohibition of *oo* or $)*$ for consecutive pairs of symbols in the strings.[221]

"Spaces" between pairs of brackets, (regardless of symbols written in there), such as ((...)... (...(..).. ...)), make a *nested* family of subsets (intervals). The natural *partial order* (by inclusion) between these subsets can be depicted by a tree – a binary tree[222] in the present case, where some vertices are labeled by two "colors" that are + and ×.

In general, let $(V, \vec{E})$ be a directed graph where there are three kinds of vertices:

• *input vertices* $v = v_{in}$: these have no incoming edge-arrows at them;
• *output vertices* $v = v_{out}$: these have no outgoing edge-arrows;
• *operational vertices* $v = v_o$ also called *o-vertices*: these have two incoming and several (a single one for the trees depicting the above string-formulas) outgoing edges.

An additional structure in this graph is an *o-labeling*:

o-vertices are labeled either by + or by × and, accordingly, denoted by v_+ and $v_\times$.

Now, given a set X with symmetric binary operations + and ×, this graph defines

A STATIONARY DESCRIPTION OF COMPUTATION
over X-values of functions on the input vertices in V.

This computation, that applies to arrays of values $x_{in} = x(v) \in X$, $v \in V_{in}$, is represented by

an X-valued function on V, say $x(v)$ that *extends* x_{in} from V_{in} to all of V, such that

$$x(v_+) = x(v_1) + x(v_2) \text{ and } x(v_\times) = x(v_1) \times x(v_2)$$

for all +-vertices and ×-vertices in V and the of vertices pairs of vertices v_1 and v_2 adjacent to them by $v_1 \longrightarrow v_o \longleftarrow v_2$, where o substitutes for the corresponding + or × label at this v_o.

If such an extension exists, as it is the case for trees, the computation is truly defined and one may speak of

the output of the computation that is the restriction of $x(v)$ to the subset $V_{out} \subset V$ of the output vertices in the graph.

Notice that this output consists of a *single* $x \in X$ in the case of string-formulas where these graphs are *rooted trees* with V_{out} being a single vertex – the root of this tree.

---

[219]Like it or not, there is a non-trivial reason why languages are condemned to this one-dimensionality. Nature herself could not write her messages on 2D DNA.

[220]An essential (not the only) aspect of this rule is the equality:

the number of the left brackets "(" = *the number the right ones* ")" .

[221]Only eight (out of 16) consecutive pairs are allowed: $*o$ $o*$)) (((* *))o o(.

[222]The (artificial) linear order structure in strings disappears when we pass to this tree.

In order to arrive at the *polynomial* (rather than a computation of its particular value) defined by such a graph one needs three extra ingredients.

(1) Division of the input subset of vertices, $V_{in} \subset V$ into two disjoint subsets corresponding to constants and to variables $V = V_{in,const} \cup V_{in,var}$;

(2) Giving specific values, say c_v to all $x(v)$, $v \in V_{in,const}$;

(3) Joining some vertices in $V_{in,var}$ by edges, representing equalities between the values of x at the corresponding vertices, where the set of these edges is denoted $E^{=}_{in}$.

This latter serves to remove the set S indexing the variables x_s from the definition of a polynomial by (implicitly) replacing it with the set of the *connected components* of the graph $(V_{in,var}, E^{=}_{in})$:

The graph (V, E) with extra (1), (2), (3) ingredients, defines a polynomial in variables indexed by the connected component of this $(V_{in,var}, E^{=}_{in})$, at least in the case where this (V, E) itself is a rooted tree.

(Supression of "amorphous" sets like S from the description of mathematical objects is very much in the spirit of ergo logic, where an alternative is giving "interesting structures" to such sets S.)

The following problem raised about half a century ago remans wide open.

Let $\mathcal{P}_1$ and $\mathcal{P}_2$ be two sets of polynomials over $\mathbb{F}_2 = \{0,1\}$ that are defined by two "simply combinatorially describable" classes $\mathcal{C}_1$ and $\mathcal{C}_2$ of *o*-labeled graphs $(V, E \cup E^{=}_{in})$, where, to simplify, we assume that $X_{in} = X_{in,var}$ and so the corresponding polynomials contain no constant terms.

When can one tell that $\mathcal{P}_1$ equals $\mathcal{P}_2$ or that $\mathcal{P}_1$ is included into $\mathcal{P}_2$?

In particular,

can one bound from below the "complexity" of graphs from $\mathcal{C}_1$ (e.g. the minimal possible number of vertices in such a graph[223]*) that represent the same polynomials as given graphs in $\mathcal{C}_2$?*

9.3 Games of Life.

The idea of decomposing logical reasoning into elementary computational steps was suggested by Leibniz three centuries ago, who introduced what is nowadays called *the Boolean algebra* that is generated by three operations over *binary* variables x, i.e. with two possible values $\circ$ and $\bullet$. The one of the three operations, called NO, is *unary*, i.e. it maps $\{\circ,\bullet\} \to \{\circ,\bullet\}$ and the other two are binary, $\{\circ,\bullet\}^2 \to \{\circ,\bullet\}$, called AND and OR:

NO: $\bullet \overset{\perp}{\leftrightarrow} \circ$,

AND: $(x_1, x_2) \mapsto x_1 \wedge x_2$*; this, by definition, equals* $\bullet$ *if* $x_1 = x_2 = \bullet$ *and* $x_1 \wedge x_2 = \circ$ *if either* x_1*, or* x_2*, or both equal* $\circ$,

OR: $(x_1, x_2) \mapsto x_1 \vee x_2 =_{def} (x_1^{\perp} \wedge x_2^{\perp})^{\perp}$*; this equals* $\circ$ *if and only if* $x_1 = x_2 = \circ$.

Since $x_1 \wedge x_2$ equals the product $x_1 \cdot x_2$ in the field $\mathbb{F}_2 = \{0,1\}$ for $\{\circ \leftrightarrow 0, \bullet \leftrightarrow 1\}$ and since addition as a function $\mathbb{F}_2 \times \mathbb{F}_2 \to \mathbb{F}_2$ can be expressed via multiplication,[224] *all* binary functions $\{\circ,\bullet\}^S \to \{\circ,\bullet\}$ are superpositions of the above

[223] The famous $P \neq_? NP$ dilemma is an instance of a precise formulation of this complexity problem.

[224] OR is similar to the sum in in $\mathbb{F}_2$, except that $1 \vee 1 = 1$ while $1 + 1 = 0$.

three. (This, apparently, was one of Leibniz' motivations for introducing these $\wedge$ and $\vee$.)

General purpose programmable computers based on similar decompositions were designed by Charles Babbage in the mid-1800's and the modern perspective on machine/network computations was opened by a 1936 paper by Alan Turing.

The idea of Turing was to show that *an arbitrary* computational process X can be effectuated by a *"simple minded bug"* who crawls back and forth along a given string of letters (or digits) and performs a computation over such a string by modifying the letters along its path according to an instruction given to it. (A definition is given later on in this section.)

However, contrary to what Leibniz dreamed, decomposing "complicated processes, call them $\Psi^\star$ into "elementary (bug's) steps" *does not* render $\Psi^\star$ amenable to any kind of "simple algebraic analysis", but rather shows that the apparent simplicity of decomposed $\Psi^\star$ is illusory. In fact, even a simplest operation, say squaring of integers $n \mapsto n^2$, may become nearly incomprehensible when translated to the language of instructions executed by a "Turing bug".

The network "core" of the bug computational model is so simple that it is almost invisible: it is the ideal *memory tape* $[0,\infty)$ decomposed into unit segments $s = [i, i+1]$, $i = 0, 1, 2, ...,$

0 –|–|–|–|–|–|–|–|–|–|–|–|–|–|–|–|–|–|–|–|– |– |–|–|–|–|–|–|–|–|–|–|–|–|–|–|–|–|–|–|–..., (i, $i+1$)

that is represented by the graph where the edges correspond to the pairs of adjacent segments.

A geometer as well as a "true bug" would feel rather constrained on the 1D tape and be more comfortable on the 2-plane divided into squares.

Langton Ant (1986).

This "ant" crawls from square to square on an infinite sheet S of squared paper, with a letter **L** or **R** written in each square s.

The ant is depicted by an arrow that may be oriented in four different ways:

$$\{\uparrow, \downarrow, \rightarrow, \leftarrow\}.$$

At a given moment, the ant makes one step ahead from its current position at s to the *adjacent* square s'.

Upon stepping to s', the ant turns either left (counterclockwise) or right (clockwise) by 90° depending on **L** or **R** written at s'.

Simultaneously, the letter written at s' switches to the other one:

$$\mathbf{L} \mapsto \mathbf{R} \text{ or } \mathbf{R} \mapsto \mathbf{L}.$$

The path taken of by the ant in S may be amazingly complicated even for an *initially constant* $\{$**L**,**R**$\}$ valued function on S, where the bug starts by creating *rather symmetric*, $\{$**L**,**R**$\}$ patterns that become progressively more and more *irregular.*

Eventually the ant runs to infinity along a "highway" made of 104 steps that repeat indefinitely. (This is seen in the picture below with white □ for **L** and ■ for **R** .)

Conjecturally, the bug eventually takes such a "highway" to infinity for *all* initial **L-R** distributions (functions) on S that are *constant at infinity,* i.e. either with **L** written at all but finitely many squares or, similarly, with **R** written everywhere at infinity.

(It does not seem to be even known if the bug eventually goes to infinity, but it is clear that the route of the bug must be *unbounded*. In fact, if a square s is consecutively visited *twice*, necessarily with *different* letters **l** and **r** written in s at the two visits, then it must, obviously, have *three* (rather than only two) adjacent squares s' also visited twice or more. This shows that the set M of multiply visited squares can not haver corners, in particularly, must be unbounded. But, a priori, M may be equal to all of S or to something like S minus a square.)

Langton ant, when positioned somewhere on S, starts moving and performs sequence of transformations, say $\Psi_1, \Psi_2, \Psi_3, ..., \Psi_i, ...$ of $\{\mathbf{L},\mathbf{R}\}$-valued functions $\sigma = \sigma(s)$ on S. If the ant eventually goes to infinity, these transformation stabilize and define a transformation that can be denoted Ψ_∞, such that $\Psi_\infty(\sigma)(s) = \Psi_i(\sigma)(s)$ for all sufficiently large i depending on s. Notice that there are many such transformations depending on where the starting positions of the ant on S are.

Then every correspondence C between the set of function σ on S constant at infinity with the set $\mathbb{N}$ of integers $1, 2, 3, ...$ will make Ψ_∞ act on $\mathbb{N}$, where such an action can be regarded as *numerical computation*.

There are lots of simple correspondences C between our functions σ on S and numbers, but none of them is natural/canonical in any way. This makes it rather awkward to nicely formulate the question of which numerical computations can be performed by this ant, and this awkwardness persists with other models of computation.[225]

Langton ant is an instance of what is called a *cellular automaton*, but it is a very atypical one, since there is "true (unsolved) mathematical problem" associated to it. Usually (always?) what one proves or even conjectures about such automata is that they are "universal", i.e. may behave in an arbitrarily complicated manner.

Conway's Game.

Unlimited complexity arising from apparent simplicity does not make mathematicians happy: what we try to do is quite opposite – finding simple regularities in the sea of an apparently unlimited complexity.

But in 1970, a mathematician John Conway found an unexpected beauty among monsters of computational complexity, called

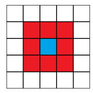
Conway's Game of Life

with an amazing balance between "chaotic" and "regular" behaviour resembling the real Game of Life on Earth.

Conway's Game is played on the same field where Lannton's ant roams, that is on the plane divided into unit squares s, but the network/graph structure on the set S of these squares is different: all eight squares s' that touch an s, either at a side or at a corner of s are regarded as *adjacent* to (joint by an edge with) s.

The states of this Game are similar to what we saw before: these are functions $\sigma(s)$ on S with values in a two element set, but the essential difference of

[225] Apparently, there is no mathematical theory of "computable symmetries" of $\mathbb{N}$ responsible for this ambiguity that would be comparable in beauty and power to Galois theory in algebra or to the theory of fiber bundles in topology.



this game from a wandering ant is that action may take place at many locations s simultaneously: Conway's S is inhabited not by a single live entity but by a dynamic ecology of interacting cells.

Formally, the game is defined as a transformation Ψ acting on functions $\sigma = \sigma(s)$, where the value of $\Psi(\sigma)(s)$ depends *only* on the values σ at *s itself* and eight cells σ' *adjacent to s.* The two possible values of σ are seen as

$$\square = \textit{dead} \quad \text{and} \quad \blacksquare = \textit{live},$$

where Ψ says what happens to the "life" $\sigma(s)$ at the moment $t+1$ depending on the immediate surrounding of s at the time t.

This dependence is assumed being the same for all s; thus, Ψ is defined by

a single $\{\square, \blacksquare\}$-valued function ψ on the set of
binary (namely, $\{\square, \blacksquare\}$-valued) functions in nine variables.

These variables correspond to the nine "modes" of adjacency between squares in S and they may be depicted as $\{\bullet, \uparrow, \downarrow, \rightarrow, \leftarrow, \nearrow, \searrow, \swarrow, \nwarrow\}$, where $\bullet$ signifies the adjacency of a square to itself.

If you feel it is easy to find an "interesting cellular game" of this kind by a brute force computer search and/or by trial and error, just imagine how long it would take to single out any "interesting" binary function in nine variables,

$$\psi : \{\square, \blacksquare\}^9 \to \{\square, \blacksquare\},$$

out of $2^{2^9} = 2^{512} > 10^{150}$ (!) possibilities.

Apparently, human (ergo)brain is able to make such a choice by blinding its eyes to the enormity of the problem. Thus, closing his eyes, Conway takes his ψ that

does not depend on the actual $\square$ or $\blacksquare$ values of the variables corresponding to $\{\uparrow, \downarrow, \rightarrow, \leftarrow, \nearrow, \searrow, \swarrow, \nwarrow\}$, but only *on the number of variables where these values equal $\blacksquare$.*

Since there are 9 possible values of this number: $0, 1, ..., 8$, the total number of possibilities for ψ is reduced to $2 \cdot 2^9 \approx 1000$, where the extra "2" comes from two possible values for the $\bullet$ variable, that is the value $\sigma(s)$ itself.

From this moment on, conceivably, there remains a single "interesting" possibility – the one suggested by Conway:

$\psi(\square, n) = \blacksquare$ for $n = 3$, and
$\psi(\square, n) = \square$ for $n = 0, 1, 2, 4, 5, 6, 7, 8$;
(Dead come to life only in the presence of exactly three live neighbours.)

$\psi(\blacksquare, n) = \blacksquare$ for $n = 2, 3$, and
$\psi(\blacksquare, n) = \square$ for $n < 2$ as well as for $n > 3$.
(Both underpopulation and overpopulation is deadly for live cells.)

The dynamics of the *iterates* of resulting transformation Ψ, denoted

$$\Psi^{\circ N} = \underbrace{\Psi \circ \Psi \circ \ldots \circ \Psi}_{N}$$

is amazingly rich even if the action of Ψ; hence of $\Psi^{\circ N}$, is restricted to the space of *finite configurations/patterns* σ that are functions $\sigma(s)$ that are "dead", i.e. equal □ at *all but finitely many* $s \in S$.

The following (taken from Wikipedia) gives a rough idea of possibilities of this dynamics.

• Some finite configurations, e.g. the following one that is localized on a single line,

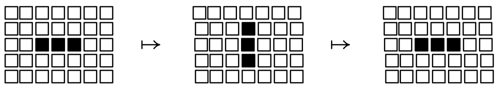
... □□□□□□□■■■■■■■■□■■■■■□□□■■■□□□□□□■■■■■■■□■■■■■□□□□□□...
grow indefinitely in the Conway plane, (contrary to the original conjecture by Conway).

• There are many N-*periodic points* of Ψ, that are finite configurations σ, such that $\Psi^{\circ M}(\sigma) = \sigma$ exactly for M being multiples of a given N.

The simplest among these – *fixed points* of Ψ, that are N-periodic points with N=1, are represented by *stationary configurations* where

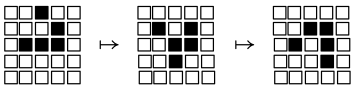
no ■ *has exactly three* □ adjacent to it (thus it remains "dead") and *each* □ *has either two or three* adjacent □.

The smallest 2-periodic pattern σ is the triplet:



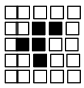
and examples of N-periodic σ are known for *all but finitely many* N.

• Some configurations, called *gliders*, move regaining their shapes:



and they eventually drift away to infinity, necessarily along a straight line (that is a diagonal for the above "pentom").

• The following "pentom" discovered by Conway stabilizes in 1103 generations at population of 116.



• The 3+4 pattern sketched in blue on the above figure, call it σ_{3+4}, with seven ■ in it "dies" (that is a special kind of stabilization) after 130 generations: $\Psi^{130}(\sigma_{3+4})(s) = □$ for all $s \in S$.

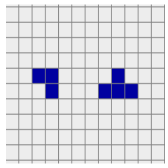
The above examples suggest that "almost everything" can be "seen" in Conway's game.... but it is hard to find a proper mathematical *language* for comprehensive expression/development of the following question:[226]

What can and what can not happen in Conway's game?

Should this question and expected answers be formulated in terms of the geometry/topology of the *space* of functions $\sigma(s)$ and the *dynamics* of Ψ acting on this space?

What are interesting (sub)spaces of functions on S, that are invariant under the action of Ψ?

[226] This is the fundamental problem we face everywhere in science.



(Instances of these, besides "dead at infinity" are periodic functions $\sigma(s)$ and those depending in a "simple manner" only on the combinatorial distance from s to some $s_0 \in S$.)

Evolution of what kind of observables (properties of configurations σ) do we want to understand?

What data, besides properties of the initial configurations, must be used for the answers?

Could formulating/answering this kind of questions be helpful for understanding the real world systems?

One knows (this took a while to prove) that any computational process, e.g. any other cellular automaton, such as Langton's ant, for instance, can be "simulated" by the Game of Life, but such a "simulation" the way this is usually understood, would not look at all as the "real" ant.

Yet, the REAL GAME OF LIFE does have this ability as it is witnessed by the *images* of computer simulations designed by human players of this GAME.

Is there something mathematically discernible in the REAL GAME OF LIFE *that is absent from Conway's game?*

This deficiency of Conway's game is shared by most models of universal computation.[227] The apparent (but not the only) reason for this is that "general modeling" suppresses the time (and space?) factor processes being modeled that is a most essential feature of a machine computation.

What would you make of an omniscient computer with 2^N minute time delay in answering your Nth question?

CELLULAR AUTOMATA OF ULAM AND VON NEUMANN.

A couple of decades prior to Conway's game, von Neumann described "automata" that can build other "automata" and later on modeled his "construction" by Conway's like game on the squared paper S. (The latter idea is attributed to Stanislaw Ulam.)

The "game" suggested by von Neumann, that mimicked his imaginary engineering of such automata, required a 29 letter alphabet X representing possible states x of the cells s (instead of Conway's two) and the von-Neumann's $\Psi = \Psi_\psi$ acting on function $x(s)$ with values in this X, where this action depends, via ψ, only on $4 + 1$ (rather than $8 + 1$ as in Conway's game) relevant neighborhood/adjacency relations between cells: $\uparrow, \downarrow, \rightarrow, \leftarrow$ and $\bullet$ for self-adjacency of s to s.

There are $29^{29^{4+1}} > 10^{28\ 000\ 000}$ different $\psi : X^{4+1} \rightarrow X$ in this case – no surprise that all kind of beasts roam the super-duper universe of these Ψ-games.[228] Apparently, *everything* can be *faithfully* modeled by this kind of a

[227] "Representation of reality" by the brain is no better in this respect.

[228] What is more difficult, if possible at all, is finding *non-trivial* "laws" that would constrain

"game" if there is no restriction on the number of states x of cells, but... it is hard (impossible?) to formulate mathematically what these "everything" and "faithfully" signify. Even the original "self-replication construction" by von Neumann has not been *formulated* as a *true mathematical* theorem with such a formulation *not being tied up* to a specific class of models *beforehand.*[229]

Von Neumann-Conway "games", often called *cellular automata*,[230] are associated with the values of the following (Ulam-Neumann) *UN bifunctor* the two entries of which are:

($\mathbf{1}_{UN}$) *A Y-valued function $y = \psi(x_d)$ in x-variables that are indexed by $d \in D$,*

$$\psi : X^D \to Y,$$

where X and Y and D are given sets.[231]

For instance, $X = Y = \{\square, \blacksquare\}$ in the Conway game and, in the von Neumann case, $X = Y$ is an alphabet with 29 letters.[232]

($\mathbf{2}_{UN}$) *A D-labeled bipartite (S,T) graph, denoted $S \leftleftarrows_D T$.*

The vertex set of this graph is the disjoint union of the sets S and T where all edges go from $t \in T$ to $s \in S$ and where the edges issuing from each vertex t are *D-labeled* by $d \in D$ for a given set D.

In other words, such a graph is defined by a map

$$G : T \times D \to S,$$

that we interpret as a D-family of maps

$$G_d : T \to S, \text{ for } G_d(t) = G(t,d), d \in D,$$

where the pairs $(t, G_d(t))$, $d \in D$, are seen as edges of the graph $S \leftleftarrows_D T$ that issue from $t \in T$. Thus, all vertices $t \in T$ have the same number, call it k, of edges issuing from them for $k = card(D)$ that is the number of elements in D.

Conversely, every graph with $k = card(D)$ edges at all $t \in T$ admits a D-labeling, with the Cartesian T-power $perm_k^T$ of the (permutation) group $aut(D) = perm_k$ of automorphisms of D acting on the set of these labeling.

(One could equally interpret maps $G : T \times D \to S$, as *T-labeled (D,S) graphs* that are represented by T-families $G_t : D \to S$, but in the "real life" the set S may be large but D is small. Moreover, in most examples, S and T are not just larger than D, but, as a objects with structures, they lie in a category that is *different* from the one to which D belongs to.)

For instance, $T = S$ of Conway's game equals the (infinite) set of the unit squares in the plane, and D can seen as the set of *distinguished directions* or *adjacency rules* that are depicted by the arrows $\uparrow, \downarrow, \rightarrow, \leftarrow, \nearrow, \searrow, \swarrow, \nwarrow$ and by $\bullet$ signifying the adjacency of a cell to itself.

---

the behaviors of all these beasts.

[229]Yet, on the technical side, it would be interesting to have a von Neumann-Conway style mathematical model of an ecology of *interacting replicators* with the resulting cut-off of the exponential growth of population.

[230]Compare http://mathworld.wolfram.com/CellularAutomaton.html.

[231]Our definitions apply to X and Y from (a class of) categories $\mathcal{X}$ different from the category $\mathcal{S}$ of sets, and, this is less straightforward, the exponents D, as sets with certain structures, also may be taken from categories $\mathcal{D}$ different from $\mathcal{S}$.

[232]Anything as large as this alphabet – 29 is a lot in ergo-terms – does not come as "just a set" but always endowed with some structure(s).

Definition of the Ulam-Neumann Bifunctor.

The value of the UN bifunctor on the pair (G,ψ) for the D-family of maps $G = G_d : T \to S$, $d \in D$, that represent the graph $S \leftleftarrows_D T$ and the function $y = \psi(x_d)$ is

the map Ψ *from the set* X^S *of* X*-valued functions* $x(s)$ *on* S
to the set Y^T *of functions* $y(t)$

that is defined by composing $x(s)$ with $G_d(t)$ and $\psi(x_d)$ as

$$t \underset{G_d}{\mapsto} s \underset{x(s)}{\mapsto} x \underset{\psi}{\mapsto} y = y(t);$$

namely

$$y(t) = \Psi(x)(t) = \psi(x_d) \text{ for } x_d = x(s_d) \text{ and } s_d = G_d(t).$$

Cellular automata are defined as such maps $\Psi = \Psi_{G,\psi}$ under the assumptions $S = T$ and $X = Y$.

Usually one also assumes that

[1] *The set* X *is finite.* In this case set X^S with its product topology is compact, moreover, it is homeomorphic to the *Cantor set*.

[2] *The set* D *is finite.* Then then the map $\Psi : X^S \to X^S$ is continuous in the Cantor set topology.

[3] *The automorphism group* Γ *of the graph* $S \leftleftarrows_D S$ *is transitive on* S. Then the map Ψ is Γ*-equivariant*, i.e. it *commutes with the natural action of* Γ *on* X^S.

The games of von-Neumann and Conway, clearly, satisfy [1], [2], [3] with Γ being the group $\mathbb{Z}^2$ of pairs of integers.[233]

Being a bug

Tutng's bugs are described in the language of cellular automata as follows[234].

$*$ Turing's $S = T$ equals the set of natural numbers $\mathbb{N} = \{1, 2, 3, ...\}$, and the set X comes with a $\{\circ, *\}$ valued function $\sigma(x) \in \{\circ, *\}$ such that $\sigma(x(s)) = *$ signifies that the cell s is occupied by the bug.

$*$ The set D of direction in the Turing case is taken to be $\{\bullet, \leftarrow, \rightarrow\}$ with the edge $\leftarrow 1 \in \mathbb{N}$ interpreted as the loop $1 \longleftarrow 1$ rather than as $0 \longleftarrow 1$ for 0 being not in $\mathbb{N}$.

$\star$ (*Localizaton.*) Functions $\psi : X^3 \to X$ admitted in this model must be such that the bug can move only one step at a time either to the left or to the right; it may stand still only at $s = 1$; the only way a cell s may become occupied at the moment $t+1$, it is by the bug moving to s from one of the adjacent locations $s+1$ or $s-1$ (with convention $1 - 1 = 1$) at this moment.

[233]Since Conway's $\psi : \{\square, \blacksquare\}^D \to \{\square, \blacksquare\}$ is invariant under the group *perm*$_8$ acting on the eight arrows in $D = \{\uparrow, \downarrow, \rightarrow, \leftarrow, \nearrow, \searrow, \swarrow, \nwarrow, \bullet\}$, the corresponding UN bifunctor is defined for *all* graphs G with *no labeling* of the edges where one does not even have to assume that there are exactly eight edges at all $s \in S$. This points toward a functorially finer version of our general definitions but a satisfactory concept of "computational network" still remains beyond our grasp.

[234]This description may be unaccceptably abstract for a practically minded computer scientist, but being mathematicians we want to find a proper place for the Turing bug in a *maximally general*, hence *simplest possible* environment. An engineer may laugh at a mathematician who start designing a square box with making up an abstract theory of symmetries but, with a luck, we may be the ones who laugh in the end.

(This localization is exactly what turns a general cellular automaton into a "bug".)

EVENTUALIZATION.

Turing imposes these mathematically artificial conditions (in slightly different terms) in order to achieve *maximal specificity* of his computation model that he, shows, is powerful enough to perform all conceivable computations where "unlimited complexity" achieved by an repetition of a a simple operation *unspecified number* of times.

In general terms, let Ψ be a map of a space $\mathcal{X}$ into itself, where Turing's main example is the space of X-valued functions x on $S = \mathbb{N}$ that are *sequences* $x(s)$, $s = 1, 2, 3, ...$, and the map

$$\Psi = \Psi_{\psi} : \mathcal{X} \to \mathcal{X}$$

is defined by some function $\psi : X^3 \to X$, via the above (UN bifunctor) construction, where the exponent "3" is a shorthand for the three-element set $\{\bullet, \leftarrow, \rightarrow\}$.

Turing's transformations $\Psi = \Psi_{\psi}$ themselves are no more complicated than the underlying functions ψ and if X is a finite set one sees a finite level of complexity in them. But the iterates of these maps, that are

$$\Psi^{\circ N} = \underbrace{\Psi \circ \Psi \circ \ldots \circ \Psi}_{N} : \mathcal{X} \to \mathcal{X}$$

may become unexpectedly complicated, similarly to what we see in the familiar image of *the Mandelbrot set*.

Say that Ψ *stabilizes on the orbit of* $x \in \mathcal{X}$ if

$$\Psi^{\circ N}(x) = \Psi^{\circ N_\star}(x) \text{ for all } N \geq N_\star \text{ and some } N_\star \text{ that depends on } x.$$

(A weaker concept of stabilization is possible for x that are *functions* $x = x(s)$, where the equality $\Psi^{\circ N}(x)(s) = \Psi^{\circ N_\star}(x)(s)$ holds with $N_\star$ that may *depend on* s as well as on x.

More generally, if $\mathcal{X}$ is a *topological space*, one can define stabilization as *convergence* of $\Psi^{\circ N}(x)$ to a *fixed pont* of Ψ for $N \to \infty$.)

If Ψ stabilizes on the orbits of all x in (necessarily Ψ-invariant) subset $\mathcal{U}$ in $\mathcal{X}$, we say that Ψ *stabilizes on* $\mathcal{U}$ and define the *stabilized map*

$$\Psi^\star : \mathcal{U} \to \mathcal{U}$$

by

$$\Psi^\star(u) = \Psi^{\circ N_\star}(u) \text{ for the above } N_\star = N_\star(u),$$

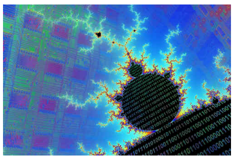
where an essential property of such $u_\star = \Psi^\star(u) \in X$ is being *fixed points* of Ψ.

(The stabilized map $\Psi^\star$, whenever it exists, can be described in purely algebraic – semigroup theoretic, terms as being Ψ-*bivariant*, i.e. such that

$$\Psi \circ \Psi^\star = \Psi^\star \circ \Psi = \Psi^\star$$

and such that all other Ψ-bivariant Φ are, necessarily, $\Psi^\star$-bivariant as well.)

Passing from ψ to $\Psi^\star = \Psi^\star_\psi$ via Ψ_ψ, say in the Turing case, may lead to computations of mind-boggling complexity. The reason for this is *non-specificity*



of what can be called *eventualization* $\Psi \rightsquigarrow \Psi^\star$ that makes it *impossible* to decide a priori

[A] *for which x the sequence $\Psi^{\circ N}(x)$ stabilizes*;

and/or to give

[B] an "*effective*" upper bound on *the first* $N_\star$ *where the stabilization begins.*

This impossibility, is "equivalent" to Gödel's Incompleteness Theorem and when "impossibility" is understood as *non-computability* it comes as

TURING'S HALTING THEOREM.[235]

Another manifestation of the unlimited complexity inherent in $\Psi^\star$ is TURING MODELING THEOREM that says, in effect, that

every computable (i.e. recursive) function can be "written" as $\Psi^\star_\psi$ *for some finite set* X *and a function* $\psi : X^3 \to X$.

In fact, there is a computable transformation $\mathcal{T}$ from the set of functions Φ that define recursive functions x as solutions of equations "$\Phi(x) = 0$" to the space of Turing's ψ, such that the *functional equation* "$\Phi(x) = 0$" transforms to the *fixed point equation* $\Psi(x) = x$ where $\Psi = \Psi_\psi$ for $\psi = \mathcal{T}(\Phi)$, where, moreover, this fixed point, call it $x_\star \in \mathcal{X}$, is a *unique attractive* one.

The construction of Turing's $\mathcal{T}$ is straightforward except that one needs to specify how Turing's sequences $x(s)$, $s = 1, 2, 3, ...$ are related with numbers – arguments of recursive functions. There are many simple ways of doing this and the conclusion of the theorem is valid (and pretty obvious) for any one of them.

Yet, there is no *natural*, *distinguished* or *canonical* $\mathcal{T}$ nor there is a natural correspondence between numbers and finite sequences;[236] also there is no *natural* extension of finite sequences that represent numbers to infinite ones where Turing's $\Psi^\star$ resides.

SERIAL AND PARALLEL COMPUTATIONS.

In general, a cellular automaton performs many (similar) computations in parallel, but the above *localization* property $(\star)$ enforces such a computation to be sequential:

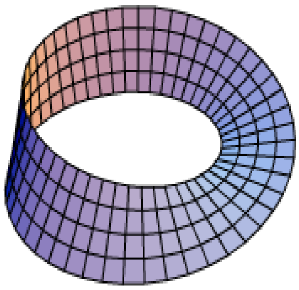
Turing's "bug" executes a single procedure at a time,

where this procedure can be one of the three Boolean operations AND, OR, NO over binary variables: $x_1 \wedge x_2$, $x_1 \vee x_2$ and $x \mapsto x^\perp$.

On the other hand, the brain's "computations" are manifestly *parallel* with millions of neurons firing simultaneously, that, apparently allows your ergo-

[235]The equivalence Gödel ~ Turing is rather obvious. On the other hand, there is no simple general framework where this "equivalence" would admit a mathematically acceptable formulation. The same applies to the even more obvious equivalence [A]~[B].

[236]Digital representation of integers is practically convenient but there is nothing intrinsically nice and natural about it. This, possibly, is why it was not accepted by ergo-oriented Greek mathematicians, even though Archimedes came close to it in his *Sand Reckoner*.



brain to absorb and understand flows of signals such as visual images and words/phrases with their internal structures being by far more elaborate than what is nodded for distinguishing binary digits, say for telling $x = \circ$ from $x = \bullet$.

This undeniable supremacy of the brain over machines may, however, be illusory, since, according to Turing's modeling theorem, *every* conceivable computation *can be* reprogrammed into a sequential form. Besides, allowing "*parallel*" *does not* ameliorate the computation time in many cases.

However,

there is *no natural* reprogramming *parallel* $\rightsquigarrow$ *sequential*.

Such reprogrammings destroy relevant structures in interesting non digital flows of signals while introducing some irrelevant artificial structures. Consequently, the *stability* under small random perturbations – an essential feature of algorithms that direct natural ergo-processes, will be lost when you go sequential.

Besides,

there is no *automatic* process for introducing *an ordering* of branches[237] of a parallel computation always preceding such a reprogramming.

Who, on Earth, can *order* millions of active neurons in the brain, where even their number is unknown to us?

A mathematician would try to bypass this issue by introducing the set PO of *all* possible orders, but the enormous size of this PO makes this idea computationally unusable. In fact, $card(PO) = N! = 1\times 2\times 3\times \ldots \times N$ for a computation running in N parallel branches.[238]

Maybe it is the *equivalence parallel* ~ *sequential* that is illusory.

Discouraging Conclusion. Most (all?) classical statements concerning general classes of computations are easy to prove, yet, they are rarely (ever?) set into a *mathematically* (as opposed to "logically") satisfactory general framework. On the other hand there are many open problems, such as $P \neq_? NP$, with no progress being achieved toward their solutions.

[237]Rephrasing Hermann Weyl one may say that indiscriminatory ordering mathematical objects is *an act of violence whose only practical vindication is the special calculatory manageability.*

[238]If $N = 2$, the ordering problem, often being attributed to *Buridan's ass* (1340), goes sixteen centuries back to Aristotle. Today, we all know for sure that this problem admits *no algorithmic solution* as it follows by contradiction from the existence of *Möbvous' strip*. And it is amazon to see how "a geometric unfolding" of this "asinine idea" has turned into magnificent theory of *fibered spaces* accompanied by *gauge theories* while in algebra it has developed into the *Galois theory*.

This difficulty, probably, is *inseparable* from our non-understanding of the logic of Life and of the Mind: we do not know what are the right questions to ask.

As far as computations are concerned we do not know, for instance, what are the true objects to which computation should be applied: are they numbers, strings of digits, or we must allow general finite combinatorial structures such as finite graphs, or some kind of (controlled?) infinite recursive (non-recursive?) objects, or what?

Even though all such computation theories are mutually equivalent in some philosophical sense, such "equivalence" is useless when it comes to modeling ergo-systems by such objects.

10 Bibliography.

References

[1] William H. Calvin, "A Stone's Throw and its Launch Window: Timing Precision and its Implications for Language and Hominid Brains," Journal of Theoretical Biology 104, 121-135 (1983). See also http://WilliamCalvin.com/1980s/1983JTheoretBiol.htm.

[2] J-P Cheng, The Origin of Probability and The Problem of Points, www.math.rutgers.edu/~cherlin/History/Papers2000/cheng.html

[3] [SLE] M.Gromov, Structures, Learning and Ergosystems: Chapters. 1-4, 6. http://www.ihes.fr/~gromov/

[4] M. Gromov Quotations and Ideas. http://www.ihes.fr/~gromov/

[5] M. Gromov, In a Search for a Structure, Part 1: On Entropy. http://www.ihes.fr/~gromov/

[6] G. Shafer and V.Vovk. The Sources of Kolmogorov's Grundbegriffe, Statist. Sci. Volume 21, Number 1 (2006), 70-98.

[7] A Hald, A History of Probability and Statistics and Their Applications before 1750, Wiley Series in Probability and Statistics, 1990.

[8] F. Lawvere and S. Schanuel, Conceptual Mathematics, Cambridge University Press, 1997.

[9] Oudeyer, P., Kaplan, F., Hafner, V.V.: Intrinsic Motivation Systems for Autonomous Mental Development. IEEE Transactions on Evolutionary Computation 11:1, (2007) and [www.pyoudeyer.com].

[10] Daniel McFadyen, Review of Liber De Ludo Aleae (Book on Games of Chance) by Gerolamo Cardano, www.link.cs.cmu.edu/15859-s11/notes/Mcfadyen_review.pdf.

[11] Wolpert, Daniel M., Diedrichsen, Jörn; Flanagan, J. Randall "Principles of sensorimotor learning". Nature Reviews Neuroscience Nature Reviews Neuroscience 12, 739-751 (December 2011) | doi:10.1038/nrn3112 http://www.nature.com/nrn/journal/v12/n12/abs/nrn3112.html

[12] Qin Zhua, Geoffrey P. Binghamb, Human readiness to throw, Evolution and Human Behavior 32 (2011) 288-293.

[13] Teaching and Learning: International Best Practice, edited by Dennis M. McInerney and Arief Darmanegara Liem.

[14] First probabilists: Galileo, Cardano, Fermat, Pascal, Huygens,

www.bricol.net/downloads/misc/HistProb/02.First_Probabilists.pdf

THE ECOLOGY OF THE DHOLE OR ASIATIC WILD DOG (Cuon alpinus) IN PENCH TIGER RESERVE, MADHYA PRADESH

[15] Sources in the History of Probability and Statistics

www.cs.xu.edu/math/Sources/

Tony Cicoria,

Alonzo Clemons

Derek Amato http://www.popsci.com/science/article/2013-02/when-brain-damage-unlocks-genius-within

Doctors diagnosed Fagerberg with a traumatic brain injury. He suffered memory loss and had problems with processing language.