Memorandum Ergo.*

Misha Gromov

October 29, 2015

Contents

*Two thirds of this "memorandum" is compiled from slightly edited extracts from the two our earlier ergo-articles.

1 Ergo Project.

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

The ultimate aim of the *ergo project* is designing a *universal learning program* that upon encountering an *interesting flow of signals*, e.g. representing a natural language, starts *spontaneously interact* 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 must 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.

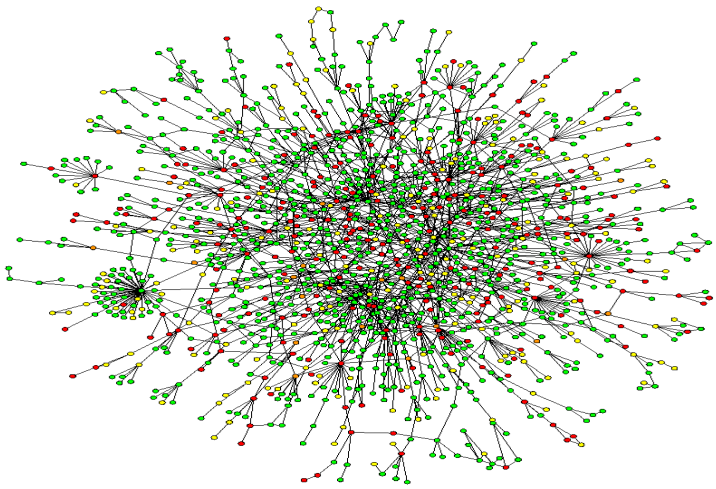
When approaching these issues, one should follow the principles of what we call the *ergo logic*, thus, *distancing ourselves* from the *common sense* ideas about the Human Mind that are dominant in our *ego-mind* and that are pervasive in our *(Popular) Culture.*

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 is composed of two subgraphs (very roughly) corresponding to the ergo-brain and the ego-mind,

$$\text{MIND} = \text{M}_{ergo} + \text{M}_{ego}$$

where M_{ergo} is a kind of a *core* of the MIND, that is a *union of cycles* and M_{ego}, a *periphery*, is a disjoint union of *trees*[1] such that each of these trees meets M_{ergo} at a single vertex – the root of this tree from where it grows.[2]



Ego is *rational*: common sense – the logic of "ego" carries accumulated evolutionary wisdom needed for our personal survival and that of our genes. Common (popular) culture is a kind of the collective ego-mind.

Human ergo is *irrational*.[3] It is after beautifully interesting structures in the world, not practically useful ones, it is enthralled by play, art, science. Science and mathematics are at the core of our *collective ergo.*

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

[1] *Trees* are connected graphs without cycles.

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

[3] Man is least rational of all animals. No matter what cockroach does, even it gets killed in the process, its behaviour is 100% rational. You can not say this about people.

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

This is great for your survival. But nothing of this kind – nothing suggested by your common sense is good as an idea in science.

We do not claim we know which model of the human mind is the nearest to the truth, but it must be as dissimilar from what the intuition and common sense whisper in your ear as one is capable to imagine.

2 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*?

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 recognise 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 human capacity for learning *languages*[4] as well as the ability to learn *mathematics*.

At the moment, one may only speculate in favour of universality by appealing to "evolutionary thrift of Nature" and to "brain plasticity". It may strike you as paradoxical that capacity to understand 1000 pages of math needed for the

[4]In order to be learned, native languages do not have to be embedded *in auditory* flows of signals: deaf and deaf blind people to learn languages relying on *visual* and/or *tactile* cues.

proof of the Last Fermat Theorem points toward simplicity and universality of the programs run by the ergo-brain.

But a specialised and/or complicated learning program, besides being evolutionary unfeasible, could hardly do mathematics that is far removed from the mundane activities the brain was "designed" for.)

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;[5] and only "clever mathematics" may furnish universality in learning.[6]

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 flavours – well organised 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.

Of course, *nontrivial* structures can be found by a learning system, (be it universal or specialised) only in *interesting* flows of signals. For instance, nothing can be extracted from fully random or from constant flows.

But signals that are modulated by *something meaningful* from "the real world" carry within certain *mathematical structures* that the brain of a human infant can detect and reconstruct .

UNIVERSAL PATTERNS IN ANIMAL BEHAVIOUR.

When, as by a miracle, the lovely butterfly
bursts from the chrysalis full-winged and perfect, ...
it has, for the most part, nothing to learn,
because its little life flows from its organisation
like melody from a music box. DOUGLAS SPALDING.[7]

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.*

[5]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.

[6]Our objectives are different from those taken by *mathematical psychologists* (e.g Robert Duncan Luce, and James Tarlton Townsend) as we are not so much concerned with modelling Human Mind but rather the "invisible" processes that shape the Mind.

[7]Douglas Spalding (1840(41?)–1877), the founder of *ethology*, was arguably, along with Gregor Mendel, the most original biology (and psychology) thinker of the 19th century. He discovered *imprinting* in baby animals (popularised by Konrad Lorentz in 1930s) and he began the study of anti-predator reactions. Unlike Darwin and Freud, he had not exposed his ideas to laymen and his name remained unknown to the general public.

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 most general law of Life from a biologist point of view is *the genetic code* – a *specific* correspondence between the two set accidentally frozen in time, one is comprised of 61 out of 64 triplets of four bases: Adenine, Thymine, Cytosine and Guanine in DNA and the second one is the set of 20 basic amino acids in proteins.[8]

$\{AAA, AAT, ..., TTT, AAC, ..., GGG\} \rightsquigarrow$ {❀, ✲, ❈, ✾, ❁, ❋, ✳, ✹,✸, ✤,✺,✵, ❀,✧,✶, ❃,❆,❁, ✽,✿,✖}.

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

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

Evolution, Universal Grammar and Chomskyan Theory.

> *...ideas which consist of "symbolic images". The first step to thinking is a painted vision of these inner pictures ... which are produced by an "instinct to imagining" and ... re-produced by different individuals independently...* Wolfgang Pauli

According to Chomsky, Lenneberg and their followers, *poverty of the stimulus* (i.e. limited data) would prevent children, who have amazing innate ability for language acquisition, from learning mother tongues as quickly as they do unless they have *universal grammar* "unscripted" in their LAD – *Language Acquisition Device* – a "language organ" in the brain or rather a module of the human mind that emerged in the course of human *evolution* as a result of some peculiar *mutation*.

Linguists usually do not bother furnishing any specific genetic or neurophysiological data on this "mutation" but rather operate with concepts of "evolution" and "mutation" *metaphorically*.

It is more likely, contrary to what Chomsky insists upon, that

> *language* ... [is] *not a wholesale innovation, but ... a ... reconfiguration of ancestral systems.*[9]

And from our ergo perspective, "universal deep learning mechanisms" are not limited to language, but also carry a variety of other functions, such as

- learning to read and to write by ~95% of (non-dyslexic) people on Earth;

[8] There nucleotide triplets: TAA ,TAG and TGA are *stop codons* that do not correspond to amino acids. Yet TGA, may represent the 21st amino acid : *selenocysteine*.

[9] *The eloquent ape: genes, brains and the evolution of language*, Fisher&Marcus [14].

• learning mathematics by mathematically inclined students that make 10% - 30% of all people;[10]

• learning playing chess by particularly gifted children (Morphy, Capablanca, Tal, Waitzkindo) by observing adults play.

The presence of these abilities in human populations forcefully denies naive arguments of the Darwinian[11] adaptive evolution.

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 movies.[12] 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.

3 Freedom, Curiosity, Interesting Signals 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, *formulating* 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,[13]

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.

[10] It may be close to 30%(100%?) at the age 3 - 5; eventually it declines, probably, below 1% partly under the pressure (unconsciously) exerted by "mathematically dyslectic" parents and teachers.

[11] More often than not, "Darwinian" is used synonymously to "truly scientific". But our "Darwinian" refers to how evolution was understood prior to the genetic revolution of the last decades.

[12] 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.

[13] 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 [32] [38].

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 [28] and by Oudeyer, Kaplan and Hafner [24] 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 programmed to) optimize (in a specific formally defined way) the quality of this prediction by varying B.

Thus, "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 bug's "freedom".

Similarly,[14] 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.

And a most dramatic instance of being prevented from learning is described by Helen Keller:[15]

Once I knew only darkness and stillness... my life was without past or future... but a little word from the fingers of another fell into my hand that clutched at emptiness, and my heart leaped to the rapture of living.

The idea of "*interesting*", that is the feature of a structure that excites "curiosity" of a learner can be best grasped by looking at the extreme instances of *uninteresting* flows of signals – the constant ones:

●●

[14] The distribution density of bugs positions crawling on a leaf is similar to that of your eye scanning this very leaf.

[15] Helne Keller who lost her sight and hearing went she was 18 months, was not exposed to *tactile sign language* until nearly age seven.

○○○

There is (almost) nothing to predict here, nothing to learn, there is no substance in 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.

○●○●●○○●●○●○○○●●○○●●●○●●●○○●●○○○○●●●○●○○●●○●○●○●●●●○●○

This appears "non-interesting" because one loses control over incoming signals, but there is much to learn from the following string that makes you curious ergo much happier.

Our 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*[16] would enjoy propelling your body through a rugged terrain with occasional random jumps from one rocky stone to another making the trip enjoyably dangerous.[17])

4 Information, Prediction and a Bug on the Leaf.

The Optimal Prediction idea of Schmidthuber-Oudeyer-Kaplan-Hafne is central in our thinking on ergosystems but we emphasise "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 learner system LEARNER *and two fragment, say* $\overleftarrow{past}$ *and* $\overrightarrow{future}$ *in the flow of signals, where*

LEANER *predicts "something" from* $\overrightarrow{future}$ *on the basis of its knowledge of* $\overleftarrow{past}$.

This "something" refers to the result of some reduction – a kind of simplification, procedure applied to $\overrightarrow{future}$ *where such a reduction may be suggested by* LEARNER itself or by another ergosystem, e.g. by a human ergobrain.

An instance of that would be predicting a *class of a* $[word]$ in a $[text]$ 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, animal, inanimate object or something else.

[16] 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.

[17] *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.

And "degree of predictability" of a class of $[word]$ derived from correlations of this class with words that *follow as well as precede* $[word]$ is also structurally informative.

(The *proper direction*, that is "follow" versus "precede" relation, is *not intrinsic* for (a record of) a flow of speech: it is non-obvious if strings have to be read left to right or right to left in an unknown language.

But, possibly, the direction can be reconstructed via some *universal* feature of the "predictability (information) profile" of such a flow *common to all* languages[18], similarly (but not quite) to how the arrow of time is derived from evolution of macroscopic observables of large physical ensembles.)

Let us apply the prediction idea to a bug crawling on a leaf or of your eye inspecting a *green* spot on a *brown* background.

We assume (being unjust to bugs) that all the bug can perceive in its environment are two "letters" G and B – 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 green G-locations from the brown ones, B.

The four "words" our bug (eye) creates/observes on contemplating the meaning of its two consecutive positions are GG, GB, BG, BB

But can the bug tell GB from BG?

Is GG "more similar" to GB than to BB?

An algebraically minded bug will translate these questions to the language of *transformations* of the "words in colours", that are:

1. *Switching the colours*: GG ↔ BB, GB ↔ BG.
2. *Interchanging the orders of the letters*: GB ↔ BG with no action on GG and BB.[19]

These 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. (Is this "equality" recognised by animals?)

Alphabet of Bug's Moves. Besides perceiving/distinguishing colours, our bug (as well as our eye) has a certain repertoire of moves but it knows as little about them as it knows about colors. Metaphorically, bug moves by pressing certain "buttons" and then records the colors of the locations these moves bring it to.

[18]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 instance, the correlations of *short* words with their neighbours are stronger for the neighbouring words that *follow short-words* rather than precede them; but this may be not so in other languages.

[19]There is yet another transformation: changing the color of the first letter :
GG ↔ BG, BG ↔ GG. This 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.

The bug does not know its own position on the leaf and the *equality of two moves* – pressing the *same button* – translates in bug's mind to the constancy of bugs direction and speed of motion.

(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 stochastic.)

Amazingly, albeit obviously, this is exactly what is needed for reconstruction of the *affine geometry* of the Euclidean plane that tells you which triples of points • • • in the plane are positioned on a straight line and, moreover, when one of the points, say •, is positioned exactly *halfway* from • to •.

And if the bug can "count" the numbers of repetitions of identical moves, it can evaluate distances and, thus, reconstruct the full *Euclidean (metric) structure* of the ambient space, that is the 2-plane in the present case.

Which buttons has the bug to press in order to efficiently explore the leaf and learn something about its meaning – the shape of the leaf?

The bug feels good at the beginning being able to predict that the color usually does not change as the bug makes small moves. (The eye, unlike the bug, can make fast large moves.) But it becomes bored at this repetitiveness of signals, until it hits upon the edge of the leaf. The bug 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 to be close to straight on bug's scale.)

Thus, the bug 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"↔"right" adds yet another involution to the bug's world symmetry group.)

Amazingly (accidentally?), 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 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.

3. Besides "repeat", there is another *distinguished* move available to the eye, namely *reverse*.[20] Apparently, approximate back and forth movements of the eye appear unproportionally often, especially when comparing similar images.

But the problems faced by out bug are harder than evaluating a metric in a *given* space.

Imagine, you are such a bug at a keyboard of buttons about which you know nothing at all. When you press a button, either nothing happens – the color does 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 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 the 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 significant (overwhelming?) probability, you know you got it right.

But what the bug has to do is even more difficult, since, apparently, there is *no a priori idea of spacial geometry* in bug's brain.[21] Bug's geometry is the grammar of the "button language".[22]

[20]Geometric incarnations of "reverse" are more amusing than the affine spaces associated with "repeat". These are *Riemannian symmetric spaces*

[21]Some animals, e.g. mice, have *map making programs* in their brains.

[22]This is similar in spirit but dissimilar in every single detail from axiomatic representations of geometries by mathematicians.

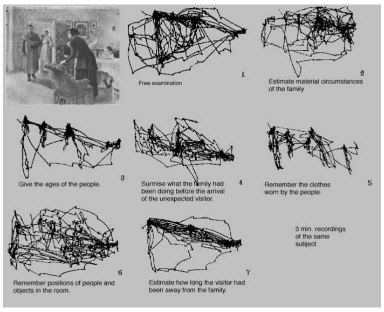
Because of this, bug's brain (and an ergobrain in general) can not use a strategy tailored for a particular case, but must follow *universal* rules, as the real bugs, we believe, do. The success depends on the relative simplicity/universality of the plane geometry, more specifically on the group(s) of 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.)

And the bug is able to make an adequate picture of the world, because, incredulously, *universality is universal*:

> the mathematical universality of bug's strategies matches
> the universal mathematical properties of the world.

This universality shapes the mechanisms of your mental processes as much as those of a bug. Your eye spends more time focused at the *edges* of images as much as bug who crawls along the edges of leaves and you pay most attention to the ends of words and it usually doesn't mcuh mttaer in waht oredr the ltteers in a wrod are.

"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.)



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. The "architecture of understanding" in the human mind is built from "bricks of predictability" that come in all shapes and colors; this is hard to reconcile with Rene Thom's "*Prédire n'est pas Expliquer*".)

5 Stones and Goals.

Reach high, for stars lie hidden in you.
RABINDRANATH TAGORE

Some maintain that mastering accurate throwing, a uniquely human[23] capacity, could have been, conceivably, a key factor in the early hominid brain evo-

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

lution.[24] 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. [5], [39].

We cannot judge the neurophysiological plausibility of this conjecture[25] but there is a chasm of differences between learning to throw and learning to speak as far as mathematics of the two learnings is concerned.

Achieving precision throwing is a *single layer problem.* Any conceivable algorithm for it, no matter how naive – that my not even need the knowledge of the laws of mechanics – is going to work. And now-a-days, it is no big deal to build a mechanical contraption that will beat any human in the throwing contest many times over, the best of Palaeolithic hunters included.

It is more difficult, but probably feasible, to design a similar program that would imitate whatever controls your tongue and larynx for proper production of sounds. But Language is a multilayer structure, wheel within wheels. No one is close to designing "a speaking algorithm" that would come close to a silliest of human conversation.[26]

The unitary hypothesis, regardless whether it is right or wrong, is hardly(?) relevant to our "ergo" but looking at throwing from a position of a goal free learner is instructive.

For a thrower, 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 and the destination point is even less so.

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

We – physicists and mathematicians with all our science would not stand a chance against *Homo heidelbergensis*[27] in a spear throwing contest; however, for instance, we, at least some of us, shall do better in mathematically designing *gravity-assist trajectories* from Earth to other Solar system bodies.

But from the position of Homo heidelbergensis[28] it would be unreasonable, not to say plain stupid, to aim at an inedible target.

Well, let us make it clear, *goal free learning* is far from being "plain and rea-

[24] 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

[25] There is a parallelism between spontaneous drives to learn to speak and to learn to walk/run/jump by children, but precision throwing is not in the same basket.

[26] An exception is what a patient expects to hear during a seance of psychoanalysis: ELIZA – a program written by Joseph Weizenbaum in the mid 1960s successfully imitates a psychotherapist.

[27] *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 years ago.

[28] This position is articulated by Lev Tolstoy in an essay on science where speaks for *a plain and reasonable man.* But of course not all (if any) of Heidelberg men were plain and reasonable. Those who were have returned to the trees.

sonable" 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* no matter how hungry he/she is, *nor in the initial condition*, although much skill is needed to achieve it. But the *transformation*

$$\textit{initial position} \mapsto \textit{aim}$$

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 and initial conditions but not so many fundamental laws$L$ and of transformations *initial position* $\mapsto$ *aim* associated to them.[29] 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* to the *final aim* 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" does as it brings you from the initial input/instruction to your aim:

What is the "space" where all this happens?

What is the structure of "the trajectory" for *initial input* $\underset{?}{\rightsquigarrow}$ *aim of learning*?

And, unlike a teaching instructor, we are not concerned with *observable* inputs and aims but with mathematical models of *invisible* intrinsic structures of transformations *inputs* → *aims* 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 "transformation formula" that is responsible for the arrow *inputs* → *aims*. 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.

6 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 neurophysiological 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.

[29]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".

"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 reasonable and 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 behaviour 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 interactions; also it suffices for expressing ideas in *the theory of mind* of *folk psychology*.

Ego-mind is responsible for WHYs about your thoughts; 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 behaviour 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.[30] This makes most of ergobrain's activity invisible.

[30] Dramatic effects of accidental breaking this wall are described in [?], [2], [25], [33], [34].

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.[31]
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 *without structurally organizing them* ("chunking" in the parlance of psychologists) equals three or four.[32] This is seen on the conscious level but such 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.[33]

Animal (including human) emotional responses to external stimuli are rather straightforward with no structurally elaborate ergo mediating between neuronal and endocrine systems.
Think of emotions as colours 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!

Ergo-moods also come in different colours:

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

serve as indicators as well as dynamic components, of the activity of the ergobrain. These indicators tell us how far our ergo-brain is from the 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.

---

[31]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.*
[32]Some people claim their $N_\circ$ is as large as (Miller's) "magical seven" but this seems unlikely from our mathematical perspective; also some psychologists also find the number four more realistic.
[33]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 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.)

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.

7 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 also the common-sense way.*[34] 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 a 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-centred 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[35] our personal survival and for guiding us socially and sexually toward successful reproduction and not at all for a structural modelling of the world including the mind itself.

The first principle is that you must not fool yourself –
and you are the easiest person to fool. Richard Feynman.

The self-gratifying ego-vocabulary of

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

[34] According to Lev Tolstoy, this is the way of thinking by *a plain, reasonable working man*.
[35] This embarrassing "for" is a fossilized imprint of the teleological bent in our language.

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 intelligence* – 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 the 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,
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 will direct us toward developing ergo-programs that would model leaning processes in a child's mind. After all, this mind 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,[36] 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.[37]

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

[37] Abstract "more" and "less" are not ingenious of ergo-brain as we shall argue later on.

‘

8 Ergo in the 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 olden 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.

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.[38]

Judging phenomena of "Ramanujans" and "Mozarts" *statistically insignificant*[39] is like signing off *explosions of supernovae* to *mere accidents*, 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)

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 interpretation of this experiment besides this and the obvious one.

[38]George Andrews and Bruce Berndt collaborated on publishing four 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.

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

can not come by an accident.[40]

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 favour 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. *Spontaneous 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.*

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

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

5. *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.*

6. *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.

7. *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.*[41] by reading a thousand-page *written proof* of it.

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

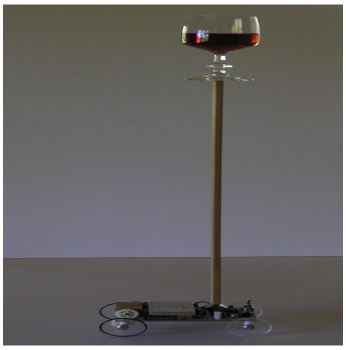
A 4-5 years old child who sees somebody balancing a stick on the tip of the finger, would try to imitate this; eventually, without any help or approval by adults, he/she is likely to master the trick.[42]

What is the mathematics behind this?

[40]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.*

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

[42]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.



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 Ramanujan's $\sqrt{1+2\sqrt{1+3\sqrt{1+\ldots}}}$.

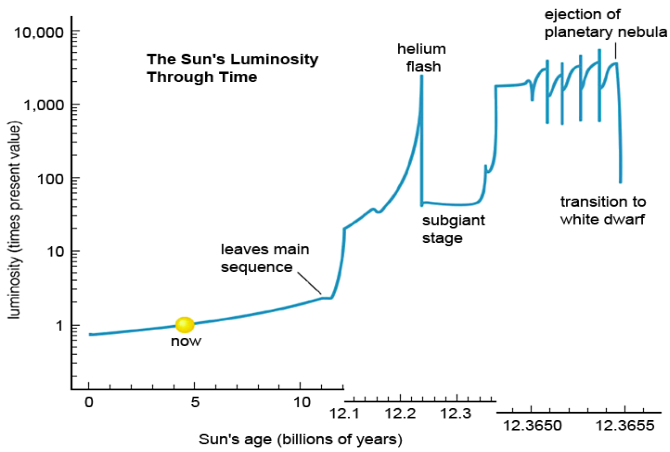
A better (ergo-style) solution of the balance problem with a single degree of freedom – the inclination angle α, may be obtained by following $grad_v(T)$ for $T = T(\alpha, \alpha', v)$ being the empirical "falling time" where α' denotes the angular velocity and v is the control parameter – the (horizontal) velocity of the support. 3

But even this distracts you from the key issue:

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

What drives Ramanujans to invent impossible formulas?

(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 and being unable to write Ramanujan style formulas, you would have to prove your "non-animal level of mentality" by putting a stick vertically in the *centre* of you cage.)



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 2 - 4 supernovae explosions per century in our Galaxy.

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 ≈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.

9 Language and Languages .

...may well have arisen as a concomitant of structural properties of the brain.
Noam Chomsky.

Most of what we know about the structure of the human ergo-brain is what we see through the window of human Language.

But what is Language? Is it *conversing, writing, reading?*

What are essential *mathematical* structures characteristic for languages?

What would make signals coming from Space classified as *"language"* rather than *"music"* or *a record of an elaborate chess-like game*?

We forfeit any idea of definition[43] but rather sketch in a few words a picture of languages of the world.

One counts ≈7000 currently spoken[44] or having recently become extinct[45] languages where *distinct languages* (rather than distinct dialects) are defined by linguists as *clusters* of vertices in the graph $\mathcal{L}_{Earth}$ where the nodes of $\mathcal{L}_{Earth}$ represent "something" (parlances) spoken by small communities of people and where the (weighted) edges of $\mathcal{L}_{Earth}$ express (degrees of) *mutual intelligibility* of these "somethings".[46]

Languages with traceable common origins are organised into ≈150 *language families* and besides the ordinary spoken languages there are several types of less common ones.

Whistled languages. There are several dozen regions in the world where the speakers of native languages can also communicate by whistling, thus transmitting messages over several km distances.

Silbo based on Spanish is practiced by 20 000 inhabitants on the island of La Gomera in the Canary Islands.

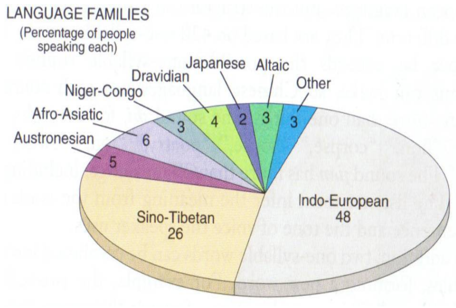
[43] A definition of *language* with a reference to human activity is as helpful for us as defining *plants* by their cooking recipes for a molecular biologist.

And defining *language* as a set of strings of symbols is no better than defining *plant* as a set of atoms.

[44] There are data on writing systems for ≈3500 languages (not all of them are widely used) and ≈700 languages are known to be unwritten.

[45] There are ≈1000 languages each spoken by $\lessapprox$1000 people with about 20 - 30 languages disappearing every year.

[46] Such an $\mathcal{L}_{Earth}$, of course, is only a dream of a linguist; besides, identifying languages as clusters depends on a *clusterization algorithm* in use.



Silbo replaces Spanish phonemes with whistling sounds distinguished by pitch and continuity. (Two whistles replace the five Spanish vowels, and four whistles used for consonants.)

MRI monitoring shows that whistling sounds are processed in the same localities of the brain as Spanish sentences.

Mazatecan languages, spoken by 200 000 people in southern Mexico, are *tonal*; thus, well adapted to whistling. Whistled communication is used predominately by men but understood also by women in Mazatecan speaking communities.

Pirahã – the language of 200-300 *Pirahã* people, living in the Amazon rainforest in Brazil, can be whistled, hummed, or encoded in music.

The Pirahã language, unrelated to any other living language, has been studied by Keren and Dan Everett who lived with Pirahã people for nearly 10 years.

According to Everetts there is no specific names for colors, no plural and no concept of *number* in Pirahã, possibly not even for *"two"*. Many other features of Pirahã, such as sentences not being produced according to Chomskyan style *transformational grammar,* also remain controversial.

Pidgin and Creole Languages. Pidgins that serve for communication between people, e.g. traders, having no common language, are built from words and other units, of several other languages.

Children who learn a pidgin as their first language create in the process learning a creol language with as elaborate a grammatical structure as those of natural languages and that are missing from pidgins.

Sign languages. There are more than 100 different sign languages in deaf communities in the world. They are mainly independent of spoken languages; their grammars have little (if any) resemblance to that of spoken languages in the same areas. For instance, *British Sign Language* (more than 100 000 users) and *American Sign Language* (more than 300 000 users) are mutually unintelligible.

Sign languages are structurally in the same league as the spoken ones, but they have a high *non-sequential* component: many "phonemes" are produced simultaneously by combining shapes, orientations and movements of the hands, arms and body as well as facial expressions.This makes developing writing forms of sign languages quite difficult; and most sign languages have no written counterparts.[47]

[47]Probably, there is an (ergo)algorithm encoding flows of signs by sequences of "phonemes"; thus, delivering *phonetic* representations of sign languages, such that children would be able

Learning and developing sign languages follow ergo-routes similar to those of spoken languages and a well minded ergo-blind teacher intervention only serves to block the learning – even more so the language creation – process.

ISN. A unique(?) instance of the emergence of a new language is provided by *Nicaraguan Sign Language* (ISN) that was developed by about 400 deaf children in Nicaragua in the1980s after the attempts to teach children "sign Spanish" had failed and children became linguistically disconnected from their teachers.

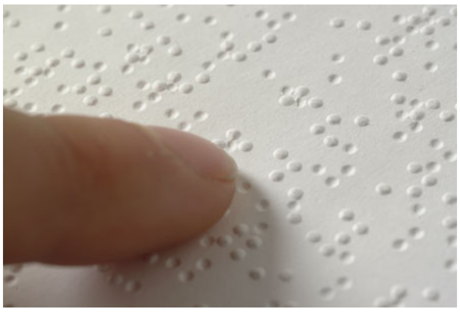
This is considered by some as a strongest evidence for *innate human language capacity.*

LANGUAGE ACQUISITION BY DEAF-BLIND CHILDREN.

> *Deafness is a much worse misfortune* [than blidness]. *For it means the loss of the most vital stimulus – the sound of the voice that brings language, sets thoughts astir and keeps us in the intellectual company of man.*
>
> HELEN KELLER.

The bulk of information we receive enters the brain through the eyes and the ears[48] (with more than 50% of the human cortex dedicated to vision), while the external world reconstructed by the brain of a deaf-blind, the world that is defined by what and who he/she touches, must be dissimilar to the world of those who see and hear.

However, given opportunity, deaf-blind children muster languages up to the point of writing poetry. This is amazing.



Life is either a great adventure or nothing. HELE KELLER

The great adventure of learning by Helen Keller (1880 – 1968), who was left blind and deaf at the age of 18 months, commenced when she was six under the guidance of 21 year old Anne Sullivan.[49]

After an initial painful failure to understand the meaning of tactile signs, a dramatic breakthrough came when Keller connected the feeling of cool water

to learn languages by listening and reproducing such "flows of phonemes"; but hardly an experiment is possible. Compare [17].

[48]Babies and young children are hungry for *tactile information* received from the lips, the tongue and the hands.

[49]Anne Sullivan (1866 – 1886) – a brilliant educatior, dubbed "miracle worker" by Mark Twain – was herself visually impaired. Born in a poor illiterate family of Irish immigrants she contracted trachoma at the age 7 and nearly fully lost her sight. Anne's education began in 1880, first by learning to read and write and to use the manual Braille alphabet. At that time, she had undergone several eye operations, which improved her sight. Soon after her graduation in 1886 she became the tutor of blind-deaf Helen.

running on her hand with the Braille sign for water.

Supported by Anne's creative teaching adapted to her needs, Helen had learned in a few months the Braille, 600 words and the multiplication table. Eventually, Keller had mastered touch-lip reading, typing and finger-spelling. Later on she learned to speak.

In the course of her life Helen wrote a dozen books as well as a multitude of essays: on faith, on blindness prevention, birth control, the rise of fascism in Europe, atomic energy. Also she gave many public speeches campaigning for women's suffrage, labor rights, social equality.

Poetry by Deaf-Blind People.

The following gives an idea of how deaf-blind people perceive the world.

....

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,...

From *A Chant of Darkness* by Helen Keller[50]

....

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

From *My Hands* by Amanda Stine.[51]

The free use of the ordinary language by deaf-blind people whose internal model of the external world is quite different from the rest of us points toward a a significant independence of language from non-linguistic stimuli in accordance with ideas of Chomsky.

Language may occupy a tiny part of the brain compared to that committed to vision but, for a human being as Ludwig Wittgenstein says,

The limits of my language mean the limits of my world.

And the inexplicable ability of children to learn and to use language is most astounding in deaf blind people. But only two lines out of 800 pages in the Handbook of Linguistic[52] are dedicated to this miraculous phenomenon:

Even a child like Helen Keller, who has lost both hearing and sight, can still acquire language through symbols expressed in touch and motion.

[50] http://www.deafblind.com/hkchant.html
[51] http://www.deafblind.com/myhands.html
[52] Blackwell Publishing (2007), p. 466.

In the words of Helen Keller,

The only thing worse than being blind is having sight but no vision.

10 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[53]

Meanings of words are determined to a large extent by their distributional patterns.

ZELIG HARRIS.

The idea of "*meaning*" advocated by 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 if you suggest 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.[54] 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 UTT is derived similarly to that of the meaning of a position POS in chess: the latter is determined by the combinatorial arrangement of POS 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

[53]In the original, one has "real" instead of "meaningful" and "reality" instead of "meaning".

[54]Semiotically minded *vervet monkeys* would not hesitated to say that the meaning of the *word-signs* of their language resides in (dangerous) object-events as these come to their fields of vision: a leopard, an eagles, a python, a baboon.

*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.[55]

(There are many different answers to the questions "What is meaning?", "What is understanding?" offered by linguists, psychologists and philosophers.[56] 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,

[55]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.

[56]References can be found on the corresponding pages of Wikipedia.

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 it should be a necessary ingredient of universal learning programs.)

None of these have counterparts in chess[57] 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. when it is communicated from mind to mind and in its logical foundations such as *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, 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.[58]

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 $\mathcal{PROPRIOCEPTION}_{ergo}$, that may include $\mathcal{TACTILE}_{ergo}$, coupled with the motor control system, since a significant disfunction of these systems at early age makes one unable to communicate.)

[57]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.

[58]There are most intriguing differences in the first language acquisition by sighted and blind children, see [3]. Probably, a comparable peculiarities would be present in the language acquisition by a sighted child born in a society of blind people who were unaware of their blindness.

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.[59]

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.[60]

But if you have no ergo counterparts to such concepts as "*some piece on a certain line*"[61] 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.

---

[59]Possibly, some bi-clustering algoritm may help.

[60]The geometry of the board can be reconstructed from a moderate list of sample chess games with *Poincaré's-Sturtivant space learning algorithms* but these algorithms are slow.

[61]Such "abstractions" are probably acquired by the visual ergo-system of a child well before anything as "concrete" as *white knight in a particular position on a chessboard*, for example.

11 Play, Humour and 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 play, e.g. *dholes*.[62]

A BEAR AND A DOG.[63]

[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.[64]

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*.[65]

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

[62]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.

[63]Taken from:
http://www.onbeing.org/program/play-spirit-and-character/feature/excerpt-animals-play/1070.

[64]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.

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

"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*.

> *The ... phrase that heralds new discoveries is not "Eureka!" but "That's funny".* Isaac Asimov.[66]

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 – theatre, 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,[67] 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.

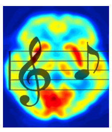
Also, there are active fields of *Neuroscience of Art, Neuroesthetics, Cognitive neuroscience of music*, with many publications openly accessible on the web.

For instance,

"positron emission tomography scanning, combined with psychophysiological measures of autonomic nervous system activity"

shows that

endogenous dopamine release in the striatum at peak emotional arousal during music listening".[68]



And an avalanche of superlatives that music lovers pour on you when they speak of music[69] tells you something about the levels of endorphins release into

[66] Fifty-fifty maybe?

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

[68] See http://www.ncbi.nlm.nih.gov/pubmed/21217764.

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

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.[70]

12 Ergo in Science.

...a scientist... is a curious man looking through the keyhole of nature.
JACQUES YVES COUSTEAU.

It is the harmony of the diverse parts, their symmetry,
their happy balance; in a word it is all that introduces order,
all that gives unity, that permits us to see clearly and
to comprehend at once both the ensemble and the details.
HENRI POINCARE

Seduction by an unadulterated beauty of the world overrides the pragmatic dictum of evolution and lures us to chess, to arts and to the ultimate human game: the mad pursuit of harmonious structures in science and mathematics. This is "ergo" in the human character that shapes the mental set-up of a scientist. "Ergo" makes the very existence of science possible.

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.[71] 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?

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

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

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

[71] 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.

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*

We are in no position to brush aside what these people say about science. Penicillin had saved about 100 million human lives.[72] Without either Fleming or Florey half of us would not be alive today and younger ones would not be even born.

Science produces ignorance, and ignorance fuels science.
STUART FIRESTEIN[73]

Science never solves a problem without creating ten more.
BERNARD SHAW

Show meant this as a mockery, but science *is* an art of *not* understanding. We strive to understand but we are not satisfied with kindergarten explanations like: *teeth are for chewing* and *wings are for flying*. We search for *new, unknown, invisible* to understand[74] what we see in front of our eyes.[75] We are happy to discover ten new problem where originally we could discern only a single one.

A four year old asks:

Why the grass is green? Why do we breath? Why the water is wet and the stones are hard? Why we do not see in the dark? Why do we not fall upward?

A *plain and reasonable man* of Lev Tolstoy would smile at nativity of these questions but a 21st century scientist would readily admit that the he/she understands none of these either, but he/she may continue with further and better focused questions:

How does the chlorophyll assisted photosynthesis work?
When and exactly how had Great Oxygenation Event occurred?
How does the cell transfer chemical energy of oxidation into mechanical energy?
What is a workable microscopic model of liquid water?
What is the nature of divergences in the quantum electrodynamics model of light and matter?
Is there a self consistent theory of quantum gravity? ....

Of course, *a plain and reasonable man* would have none of this. If anything, he would like to understand Nature in "a few simple words".

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

[73]Firestein is a biologist. In his lab, they study *the vertebrate olfactory receptor neurone as a model for investigating general principles and mechanisms of signal transduction.*

[74]Here, "to understand" refers to what is called "understanding" by students of natural science. This is different from "to understand" of mathematicians and has little in common with "to understand" of humanities scholars.

[75]A rare for humanities instance of an explanation of known by unknown, be it only a *conjectural* one, is that by Julian Jaynes who suggested that the major Mediterranean Religions had resulted from *the Breakdown of the Bicameral Mind* about three thousand years ago. Jaynes' bicameral mind conjecture is, in principal, falsifiable – it my be either true or false.(One can not assign truth value to "teeth are for chewing" theories of religion or of anything else for this matter.) But conducting an actual experiment verifying Jaynes' conjecture is ethically prohibitive.

Well, the way the world is run may be beautifully simple, but our mind was not designed by Nature for contemplating her beauty. Perceiving this beauty requires an utmost intellectual effort on our part.

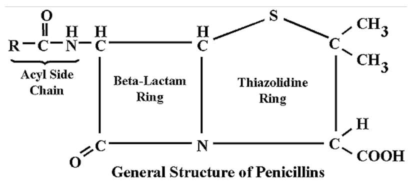
Even most familiar and apparently simple things in science are intuitively hard to accept, such as the second law of Newton that presents a manifestly mathematical (ergo)way of thinking about motion:

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.[76] 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*.

But the main reason why our brain resists absorbing scientific knowledge is the complexity of the combinatorial organisation of this "knowledge" that the brain needs to represent within itself.[77] Probably, the number of synaptic connections needed for *understanding* "N units of knowledge", of, say, *string theory* in mathematical physics grows (at least) as N^2 rather than $const \cdot N$ which is required for *absorbing*[78] the same number of "units of knowledge", say, in *cultural anthropology* that conveniently fit into pre-prepared niches in your ego-mind.

13 Unreasonable Men and Alternative Histories.



It [science] *triumphantly tells him: how many million miles it is from*

[76]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.

[77]Partly, the difficulty in understanding "abstract" mathematical ideas is due to the protective wall separating ego from ergo.

[78]Memory in the brain is not straightforward unlike that on the magnetic tape. For instance, remembering long loosely structured (quasirandom) sequences, such as pages of telephone directories and arrays of the dates of "great historical events", is difficult – *excruciatingly difficult* for mathematically inclined among us.

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.

LEV TOLSTOY.

Nothing about penicillin – the miracle drug that had cured millions over millions of people was ever reasonable. And those who had discovered and developed this drug were anything but *plain and reasonable people*.

The modern chapter of penicillin history starts with a *Stafilococci plate* in Fleming's lab that, between 27 July and 6 August of 1928, was unreasonably and unaccountably contaminated by an unusual in its antibacterial activity strain of *Penicillium notatum*, probably by the spores that escaped from the lab of La Touche – the mycologist at St Mary's Hospital working in the room below Fleming.

In 1928 –1929, Fleming determined that the secretion from the mold *Penicillium notatum*, that he called *penicillin*, was effective against many bacteria.

(Penicillin suppresses growth of the so-called *gram-positive* bacteria e.g. *streptococci* and *staphylococci* that have no outer cell protective membrane, by blocking the cell wall growth when bacteria replicate.)

Fleming suggested that penicillin could serve as a low toxicity[79] disinfectant; also he indicated the laboratory use of penicillin for the isolation of *Bacillus influenzae*. But despite a few therapeutic successes,[80] he became disappointed with instability of penicillin and turned to his other projects.[81]

In 1938, Ernst Chain (1906 – 1979), upon reading the 1929 paper by Alexander Fleming proposed the study of penicillin to Florey.

In 1939, Howard Florey (1898 – 1968) 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), with the key roles played by three biochemists: Chain, Heatley[82]and Abraham.[83]

[79]Penicillin kills bacteria but it is harmless for humans. However it is toxic to a few mammalians, e.g. to Guineia pigs whose intestines are inhabited by gram-positive bacteria. Luckily, Fleming and people from the Florey's team, Chain and Heatley, tried penicillin on mice, apparently because they have limited amounts of the drug.

[80]In 1930, Cecil George Paine, treated a gonococcal infection in infants and achieved the first recorded cure with penicillin. He then cured four additional patients of eye infections, and failed to cure a fifth.

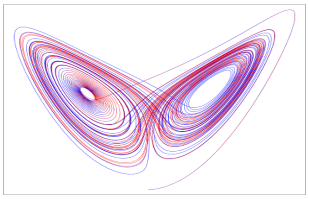
[81]Alexander Fleming (1881–1955) was a member of the research department at St Mary's Hospital in London that was organised and directed by Almroth Edward Wright (1861 – 1947).

In the first year of Wold War 1, Wright and Fleming worked out a treatment of infected wounds; ever since Fleming had been searching for antibacterial agents. In 1921, prior to penicillin, he discovered *Lysozyme* – an enzyme, present in tears, saliva, human milk, and mucus, that protects from gram-positive pathogens.

[82]Norman Heatly (1911– 2004) has devised main steps for producing therapeutic quantities of penicillin. This, combined with the know-how in fermentation technology of organic acids, has led to fast development of production of penicillin on the industrial scale.

[83]In 1943, Edward Abraham (1913 –1999) determined the structure of penicillin which in-

Does the Flap of a Butterfly's Wings in Brazil set off a Tornado in Texas?
EDWARD LORENZ



There had been several decisive moments in history when ideas, insights and decisions by such people as Fleming, Florey, Chain, Heatley, Abraham were turning the path taken by humanity to its present course.[84]

Below are two other, even more vivid instances of (potential) instability of the course of the human history.

1. In 1896-1897, Ernest Duchesne (1874 – 1912) conducted the first(?) 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* that he sent to *Institute Pasteur.*

If Duchesne's work had been taken seriously by people (a single person?) at Institute Pasteur the development of antibiotics (and of the world pharmacological industry, in general) could have started a few decades earlier.[85]

It is hard to imagine what kind of world we would be then living today.

2. A boy fails to recall the name of the river the city Berlin is located on and as a result he is denied entrance to a gymnasium in Odessa. Seeking education, he is compelled to move to the United States where a few decades later he starts a research on the soil bacteria. Around 1940 this "boy" develops a comprehensive program for screening and testing *actinomycetes* for antibacterial activity. This leads to discovery of a dozen antibiotics including, in 1943, *streptomycin* – a drug harmful to *gram-negative* bacteria, the first one effective against *mycobacteria* that cause tuberculosis.[86]

The name of the "boy" was Selman Waksman (1888 – 1973).[87] Streptomycin was discovered by Albert Schatz (1920 – 2005) who worked in Waksman's group.

---

volved *beta-lactam ring*; this was confirmed in 1945 by Dorothy Hodgkin by X-crystallography. In the 1950s, Abraham essentially contributed to isolation and development of *cephalosporin*, that kills penicillin-resistant bacteria.

[84]The perturbative effects of these people on human history were more subtle than of those depicted by Stefan Zweig in *Sternstunden der Menschheit*.

[85]This may be compared to what could have happened if Nägeli had understood Mendel. Possibly, nothing would have changed in both cases: the whole of the scientific community, stabilised by inertia, was not ready for these ideas.

[86]Hundred years ago, 10-15% deaths in Europe were inflicted by tuberculosis. In 2013, about nine million people in the World fell ill and 1.5 million died from TB.

[87]A significant factor turning Waksman's interest toward antibiotics, besides the achivements by Florey's group in Oxford, was the work of René Dubos (1901 – 1982) who, in 1939, under influence of Oswald Avery (1877 –1955) and with the help of Rollin Hotchkiss (1911 – 2004) at the Rockefeller University in New York, isolated a bactericidal substance from the spore-forming bacillus of the soil, that he called *gramicidin*.

Were Waksman's examiner in Odessa less pedantic, the discovery of streptomycin and other antibiotics could have been delayed by several years and hundreds of thousand people would have died of tuberculosis during this time.

A FEW REFERENCES ON ANTIBIOTICS AND THEIR DISCOVERERS.

Gwyn Macfarlane, a haematologist who worked with Florey, wrote down two masterful accounts on the lives and works of discovers of penicillin, [20], [21].

More can be found in [11], [23], [6], [8], [35], [36], [16], [4], [37], [9], [29], [12], [30], [27].

14 Mathematics and is Limits.

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

We share our inborn[88] ability to count with *pigeons* and *vervet monkeys*, but it is long way from *quantity and shape perception* by animals (humans included) to something like *Ramanujan mysterious formula*

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

This formula, similarly to equally incredulous but more familiar *Leibniz formula* $1-\frac{1}{3}+\frac{1}{5}-\frac{1}{7}+\frac{1}{9}-\ldots = \frac{\pi}{4}$, relates the *geometrically* defined number $\pi = 3.14159265\ldots$ to an *arithmetically generated* infinite sum, that in the Ramanujan case is comprised of the impossibly complicated terms.[89]

What allows such miracles in mathematics?
What is mathematics from the ergo perspective?
What is mathematics that underlies ergo?

The relations between mathematics and ergo, that are by necessity circular, may be summarised as follows.

Mathematics at its core, is "just" an instance of an ergo-structure.
Mathematics is "just" a fragment of the collective[90] human ergo.

[88] Inborn? – not quite, at least not quite inborn in *humans*. Our number sense is intertwined with the language learned in the cradle.

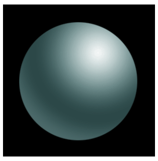
[89] The series $1103 + \frac{4!(1103+26390)}{396^4} + \frac{8!(1103+26390\cdot 2)}{(2!)^4 396^8} + \frac{12!(1103+26390\cdot 3)}{(3!)^4 396^{12}} + \ldots$, unlike Leibniz' $1-\frac{1}{3}+\frac{1}{5}-\ldots$, converges exponentially fast; this allows a practical computation of the decimals of π with Ramanujan's formula.

[90] We, mathematicians are a tiny community, something of order 0.001% of the total population, and, probably only a couple of hundreds among us (myself is not in the club) understand the Ramanujan formula.

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.

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.

The main transitions in the evolution of mathematics were not achieved by reduction of unknown to known but, contrary to the common sense, by inventions of "irreal objects" such as *negative* and later *complex imaginary* numbers, *infinitesimals*, *ideal numbers*, *n-dimensional spaces*, etc., similarly to how the progress in physics was driven by "irreal ideas" of *atoms, wave functions, quantum fields*.

It is amazing how mathematics manages to contain these symmetries and to represent, for example, "*roundness*" immanent in something like $\pi = 3.14159265...$ by a Ramanujan-like, let it be infinite; yet, *arithmetic* formula or in the combinatorial, nearly digitalised, form of axioms, lemmas, theorems and proofs

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 as well.[91]

This "combinatorial" nature of mathematics may be also compared to that used by Life for encoding shapes of organisms by DNA sequences, 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]

[91] 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.

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 "ergos", we 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.

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.

3. In building 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. And remember that

you can not apply mathematics as long as words still becloud reality.

as Hermann Weyl said.

Our choice of the components of a logical reconstruction of human ergo-brain, call it $\mathcal{HEB}$, follows the criteria used everywhere in mathematics:

NATURALITY, UNIVERSALITY, LOGICAL PURITY, CHILDISH SIMPLICITY.

Mathematical universality of $\mathcal{HEB}$, in particular of its learning strategies (programs) may be seen in how we *enjoy and learn many different* logically complicated games. Thus, for instance a chess learning program in somebody's $\mathcal{HEB}$ 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.[92] What is not simple is discovering and formulating 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 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 our (even if conjectural) ergo-brains.

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

But beware, mathematics that directs your thoughts toward " logically perfect structures", may mislead you when it comes to the "real life".

(The arithmetic of numbers is seductively beautiful, but the surface temperatures of the sixteen named stars in the constellation of the Great Bear, even if taken in Kelvin, are not meant for addition and multiplication.)

The structures of "ergo" grow on the soil of mathematics but they diverge unrecognisably far from their pure mathematical prototypes.[93]

And it may appear – this has been the dominant philosophy in the mainstream AI – that it is not so much mathematic proper, but rather such concepts as *axiomatic systems, automata, Turing's machines, Gödel's theorems, etc.*, are essential for understanding mental processes.

However, these ideas – this is witnessed by their poor record in implementing Turing's original program – are as insufficient for understanding the nature of the human thinking processes as for illuminating the nature of the human mathematics.

15 Numbers, Symmetries and Categories.

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

[93]This is not so in theoretical physics but similar to the position of mathematics in shaping (mostly unknown) fundamental principles of biology.

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.[94] 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 neurones 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.[95]

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 endow an ergo-learner with 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.

---

[94]Certain power/energy quantities, e.g. those measured in *decibels*, comes on the multiplicative (logarithmic) scale, and their doubling is perfectly meaningful.

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

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

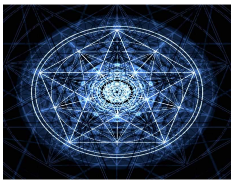
Numbers are not in your ergobrain but the idea of symmetry is in there. Much of it concerns the symmetries of our (Euclidean) 3-space, the essential ingredient of which – the group of the *(3-dimensional Lie) group* of all possible motions, call them *rotations* of the Euclidean *round 2-sphere* within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

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*[96] *of signals* that *break* spatial and temporal symmetries.

For instance, the input of the visual system may be represented by the set of samples of a distribution of (not quite) a probability measure on the set of subsets of the light receptors in your retina. *In principle*, this is sufficient for reconstruction of the Euclidean geometry similarly to how Alfred Sturtevant obtained the partial genetic map of drosophila X chromosome on the basis of distributions of phenotype linkages.

However, your brain would not(?) be able to "map" the 3D-space without receiving along with visual signals also signals coming from the firing of the motor neurones controlling your motions, especially those of your eyes,

The eye, by the necessity of having rotational freedom, is spherical.[97] The movements of the eye, have, apparently, prepared the brain to the idea of spatial symmetries and have helped the brain to learn how to identify images that move in the visual field.[98]



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 what mathematician called *the five element Galois field* that can be visualised as the set the five vertices of ⬠ with with 20 transformations acting on it, where only 10 of them are geometrically apparent – [5 rotations]×[2 reflections]. But there is an extra one, that is there because 5 is *a prime* number and that can be depicted as ⬠ ↦ ⛤.

A fantastic vision, unimaginable to ancient mystics and to mediaeval occultists, emerges in the *Langlands correspondence* between arithmetic symme-

[96] The initial leaning programs we envisage contain no counting mechanism, but an ergo-learner must be able to distinguish "significant/persistent" signals and ignore "accidental" ones.

[97] The eye shares the spherical symmetry with the shoulder and hip joints that also have three degrees of freedom, while the cylindrical knee joint allows only circular motions. And the elbow hinge-joints is "designed" with *exactly two* degrees of freedom.

[98] This is explained in §IV of Poincaré, *La science et l'hypothèse.*

tries 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.[99]

Categories, Functors and Meaning. Mental ergo-objects, e.g. sentences of a language, rarely (ever?) possess perfect internal symmetries, that move objets within themselves, such as the rotations of spheres, pentagons or of icosahedra[100] preserving their geometric structures. (Icosahedron admits 120 transformations where – this is not fully accidental – $120 = 5! = 1 \times 2 \times 3 \times 4 \times 5$.)

However, certain transformations of such objects, e.g. of sentences, can be depicted, albeit only approximately, in the *mathematical category theory.*

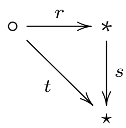
The simplest, geometric rather than syntactic, transformations are *insertions* of strings of letters into longer strings positioned somewhere else in a text,

$$...ABC... \longrightarrow ...DEABCMN....$$

Mathematical Categories extends the concept of symmetries by allowing this kind of transformations that "move" one object into another one.

Such transformations are depicted by arrows between objects, e.g. $string_1 \to string_2$, or symbolically $\circ \xrightarrow{r} *$, where the structure of a particular category is given by a *composition rule* between incoming and outgoing arrows for all objects in this category.

Namely, one distinguished certain triples of arrows, say (r, s, t) between triples of objects, call them $(\circ, *, \star)$, such that $\circ \xrightarrow{r} *$, $* \xrightarrow{s} \star$ $\circ \xrightarrow{t} \star$, for which one declares that t equals the composition of r and s and one depicts this by the following diagram.



(In geographic terms, objects are "locations" and arrows between pairs of "locations" – there may be many of these – are possible routes from one location

[99]Seen by an outsider, the symmetry in mathematics is diluted to the point of invisibility by *useful formulas, difficult computations, efficient algorithms, logical axioms, reliable (or unreliable) statistics...* .

[100]Viruses – they are not bound by ergo – love the icosahedral symmetry, since this minimises the area of their proteins shells that must contain all of DNA encoding these very proteins.

This is how Life works : information and geometry (of physical matter) go hand in hand. But mathematicians lags far behind viruses in the solution of *inforimetric problems* to another one, where composition of routes is the route that is obtained by consecutive following these routes, say, first r from $\circ$ to $*$ and then s from $*$ to $\star$.)

The combinatorics of (large) arrangements made from triangles of arrows carries an unexpectedly rich amount of information about the *internal* structures of the objects in our category. For instance, a seemingly vague idea of "*naturality*" of a certain mathematical construction can be non-ambiguously expressed in terms of *functors* between categories.

When applied to a language, this leads to a workable (sometimes called "holistic", see p. 242 in [19]) definition of "meaning" of a text without any reference to "real meaning" along the ideas maintained by Zelig Harris: *meanings of words are determined to a large extent by their distributional patterns*.

16 Logic and Illusion of Rigor.

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.

Evolution of mathematical concepts in their convergence to clear shapes they acquired in the 21st century suggests how one may design ergosystems. Yet, not all roads we explored had lead us to the promised land; understanding what and why did not work may be more instructive than celebrating our successes.

The problem with our mathematical ideas is not that they are too abstract, too difficult or too farfetched, but that we lack imagination for pulling abstract difficult and farfetched ideas out of thin air. Nor do we have a foresight for predicting how an idea will develop.

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 rigour 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 tunes of formal logic.[101] We are suspicious of "intuitive mathematical truth" and we do not trust *meta*mathematical rigor[102] 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

[101] Not everything in logic is collecting, cleaning and classifying morsels of common sense – it is hard to believe but logical thinking can be creative. But this "creative logic" is what we call *mathematics*. We happily embrace *model theory, set theory, theory of algorithms* and other logical theories that became parts of mathematics.

[102] The concept of *logical rigour* unlike that of *mathematical rigour* can not be defined even with minimal requirements for being precise and rigourous.

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. The miracle of the Leibniz formula $1 - \frac{1}{3} + \frac{1}{5} - ... = \frac{\pi}{4}$ (of 1682) achieved by an appeal to infinitesimals makes the criticism of insufficient rigor in mathematics by George Berkeley (1734) as well the idea of "redemption" of Leibniz' calculus by Abraham Robinson (1966) look puny.[103]

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.

(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[104] and Edmond Halley famous for computing the orbit of *Halley's Comet*.[105])

We can not take seriously anything like $(a, b) := \{\{a\}, \{a, b\}\}$[106] To get "convinced" that this definition is worth making, you must accept logicians' appeal to metamathematical intuition: *nothing* of meaning can be communicated *without a use of natural language* the metaphoric essence of which dissolves logicians' idea of "perfect rigour".

We communicate mathematics not by coping scores of logical symbols from one mind to another but by making another's ergo-brain resonate to the tunes we hear within ourselves.

[103] 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.

[104] 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*.

[105] 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.

[106] This is the 1921 definition of an *ordered pair* by Kuratowski.

And we do not bow down before "power of intuition" so extolled by logicians and mathematicians but rather try to understand the source of this intuition in the human ergo-brain/mind.

Who argues, establishing "truths" of certain kind needs perfect precision, all 200% of it; below is an example of such a "truth".

$$4579 + 8763 - 3459 + 4686 - 6537 + 7763 - 4579 + 1099 - 8765 + 1238 - 3677 = 1111.$$

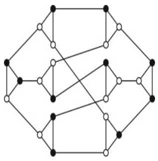
But – this is a miracle from a logician's point of view – a formally imprecise outline of the "core idea" of a *meaningful* statement, even of a relatively structurally plain one such as *Gödel's first incompleteness theorem*, ascertains its validity in the eyes of a mathematician.

(It is how "*something alive*!" in an incomplete and distorted image of an animal, that may be even unknown to you, catches your eye.)

(The century old foundational dust finds its way to our textbooks under pretext of rigour,[107] 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...*

Brr..., there is more sense in simply drawing a graph than giving such definition.)



Problem with Order.

Concept of *logical rigour* doesn't fare well in the ergo environment. For instance, the abstract concept of *order* can not be defined logically rigorously but only with a reference to a preexistent order in the physical or phycological medium where the idea of order is expressed.

Ordering a pair involves breaking $-1 \leftrightarrow +1$ symmetry similar to that encountered by *Buridan's Ass*. (To see it picturesquely switch from the habitual $A \prec B$ to ▲ ❖ ▼ that is graphically but not contextually invariant under rotation by 180°.[108]

Besides mathematical impossibility of algorithmically resolving the "order problem" (the reader is invited to count the number of logical flaws in the above "solution" $(a, b) := \{\{a\}, \{a, b\}\}$) there is a good evidence that there is no innate "idea of order" in our ergo-brains. For instance, children often use

[107]The persistent urge for "rigour" in certain minds begs for an explanation by a Freudian style psychologist.

[108]One may only wonder what face of mathematics one would see in the world where "formulas" were invariably represented by *symmetric* combinatorial arrangements of symbols on the plane.

mirror images of letter-signs when they beginn writing and mathematicians tend to revert (mentally as well as graphically) directionality of their inequalities.

Accordingly, we should not(?) postulate *a primary idea* of order in our design of ergo-systems.

In general, we need to be choosy in our terminology and in assigning basic concepts/operations to our learning systems:

no elementary structure, no matter how simple looking and "obvious", can be taken for granted.

For instance, a child who has learned to count on fingers and can figure out that that $2+3=5$ and $2\times 3=6$ resists $3=3$ and $1\times 5=5$. If you think the child should be instructed to accept $[3{=}3]{=}[3{+}2{=}5]$, this is yourself, not the child, who lacks proper education: $[3=3]\neq[3+2=5]$ – mathematics has arrived at the point of accepting child's attitude and developing means (around *category theories*) for differentiating between various kinds of "equalities". And an *hierarchy of equalities* is essential for our models of ergobrain.

Logic in Science.

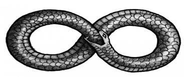
Mathematical rigor and logical certainty are absent not only from logical foundations of mathematics but also from all natural sciences even from theoretical physiscs. Einstein puts it in words:

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 rigour" is higher on certainty than the logical one, since reproducible experiments are more reliable than anybody's, be it Einstein's or Gödel's, intuition.[109]

17 Infinite inside, Finite outside.

If any philosopher had been asked for a definition of infinity, he might have produced some unintelligible rigmarole, but he would certainly not have been able to give a definition that had any meaning at all. Bertrand Russell



Mathematics is the abode of infinity. Almost all our theorems are about infinite objects, e.g. associated with *real numbers* such as *Lie groups*[110], or about infinities of finite objects, e.g. *prime numbers*.[111]

And basic properties of finite objects such as prime numbers can be understood only in the ambiance of transcendental infinities of real as well as *p-adic*

[109]We do not bow down before "power of intuition" so extolled by logicians and mathematicians but rather search for the source of this intuition in the human ergo-brain/mind.

[110]The basic example of such a group is that of *rotations of an imaginable rigid body* in the 3D-space.

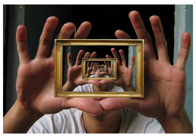
[111]Even individual (especially transcendental) real numbers, such as
$\pi = 3.14159265358979323846264338327950288419716939937510582097494459230781640628 62\ldots$
would barely exist without an ocean of infinity that surrounds and support them.

numbers[112] and of uncountably infinite *adelic groups*[113] represented by rotations of *infinite dimensional geometric (Hilbert) spaces.*

Also this kind of infinity is indispensable for how we describe the physical universe and its fragments. Most (all?) physical systems are modelled by (uncountably) infinite sets of *real numbers.* (Physical constants, even dimensionless ones such as 0.00729735257[114] .., are, of course, not quite numbers, but it is hard to pinpoint what exactly this quite is.)

But looking from outside, the whole body of mathematics, call it $\mathcal{M}$, is a humble mathematical object describable in *finitely many* words.[115] These words however, generate a language that is represented by something *infinite*, call it $\mathcal{M}'$ – a "fragment" of $\mathcal{M}$ taken from it and positioned outside of it.

Thus, $\mathcal{M}$ starts looking as a giant kaleidoscope full of tiny mirrors $\mathcal{M}'$ each with the ability to fully reflect all of $\mathcal{M}$ with the mirrors, including $\mathcal{M}'$ itself, in it.



The coexistence of "equivalence" $\mathcal{M}' \sim \mathcal{M}$ with *strict* "inclusion" $\mathcal{M}' \subsetneqq \mathcal{M}$ allowed Gödel to untangle selfreferentiality within *Liar Paradox: I am unprovable is unprovable* and, thus, prove his theorem on the existence of *formally unprovable mathematical propositions*.

(Gödel's $\mathcal{M}'$ is an infinite set of the strings of finitely many, say 10 symbols/letters representing proofs in $\mathcal{M}$ described by sentences say in ten "letters" a_i. This $\mathcal{M}'$ "embeds" into $\mathcal{M}$ by means of *Gödel's enumeration*, where the "letters" are depicted as 0, 1, 2, 3, 4, 5, 6, 7, 8, 9; thus, "translating" formulas in $\mathcal{M}'$ to *decimally represented* numbers from $\mathcal{M}$ and properties of these formulas, notably their provability/non-provability, to arithmetic properties of these numbers.[116]

Gödel's proof depends on "rigid" physical space-time of our Universe that serves as supporting background for *positional* representation of numbers. No such proof would be possible in a "Liquid World" with *nothing* "rigid" like our space-time in it.)

All of mathematics with all these *uncountable sets* that are incomparably greater than $\mathbb{N} = \{1,2,3,4...\}$ can be reflected in the countably large (small?) mirror of arithmetics according to a childishly simple theorem called *Skolem's paradox*:

[112]These are limits of rational numbers where the usual rule $\varepsilon^i \to 0$ for $|\varepsilon| < 1$ and $i \to \infty$ is formally replaced by $p^i \to 0$ for a given prime number p and $i \to \infty$. Amazingly, this leads to a meaningful concept of a "number".

[113]These groups are infinite products of certain Lie groups with all their p-adic counterparts, $p = 2,3,5,7,11,....$

[114]This is the *fine-structure constant* $\alpha = \pi \cdot [\textit{elementary charge}]^2/hc$, introduced by Sommerfeld in 1916 to account for the spectrum of the hydrogen atom.

[115]From a non-mathematical scientist point of view such an $\mathcal{M}$ is *a mathematical model* (not of a kind defined by mathematical logicians) of "the real mathematics" that is practiced by human mathematicians on the planet Earth.

[116]Logicians, apparently to show their respect for Gödel, compose this with the map $\mathbb{N} \to \mathbb{N}$ defined by $\{a_i\} \to \prod_i p_i^{a_i}$ for the prime numbers $p_1 = 2, p_2 = 3, ..., p_i, ...$

"all" mathematical objects comprising uncountable sets can be "adequately represented" by countably many their "verbal descriptions".

But this is still about "infinities"; in "reality", all these infinities are fictional: the number of mathematical statements that can be ever produced by intelligent beings in our or any other conceivable universe is pathetically small[117] compared to immeasurable infinity of "objects" allowed by the grammar of the mathematical language.

Also, since all physical, biological and/or mental structures may function only on specific space/time/complexity scales, meaningful philosophical interpretations of Gödel theorem seem difficult, if not impossible, to attain.

Most (all?) speculative (rather than purely mathematical) arguments about these structures (e.g. in the philosophy of artificial intelligence) with a whiff of a hint at Gödel's or similar "infinity theorems" such as *Turing halting theorem* or *Kolmogorov-Chaitin complexity* inevitably harbour gross misinterpretations of the concepts underlying these theorems.

Ergo-Logical Model of Mathematics.

Out of an infinity of designs a mathematician chooses
one pattern for beauty's sake and pulls it down to earth.

Marston Morse

A logician's model $\mathcal{M}$ of mathematic, e.g. *Peano arithmetic*, has an unattractively amorphous structure that may be compared to that of the morass of *all positions* of chess pieces on the board that may be obtained from the initial position by the rules of chess.

(Peano arithmetic is an axiomatic description of the structure of integers $n = 1, 2, 3, ...$ that is defined via the *successor operation* $n \mapsto n'$ (where, secretly, $n' = n + 1$),

$\overset{1}{\bullet}\mapsto\bullet\mapsto ...,$

where the basic axiom is that of *induction*:

if a "proposition (or a definition) expressed by a formula" F is valid for $n = 1$ and F-for-n implies F-for-n', then F is true for all n.

For instance the addition $m + n$ is *defined* by the formula

$$m + 1 = m' \ \& \ m+n' = (m + n)'.$$

There is no means in the traditional logic to see what is "interesting" and to describe the structure of "the set" $\mathcal{M}^{ergo}$ of all "interesting" theorems.[118]

One of the goals of ergo-logic is to achieve this end but we can not say yet with a convincing degree of precision what *interesting* theorems (and/interesting chess positions) are but there are illuminating examples.[119]

For instance, Harrington and Paris found out several bona fide mathematical propositions, such as *Goodstein's base change theorem* that are unprovable in

[117] The number of strings of symbols that can be generated during the life cycle of any conceivable universe can be safely bounded by $10^{10^{10}}$.

[118] "Interesting" theorems do not naturally comprise what we call "a set".

[119] It may be also worthwhile to borrow ideas from evolutionary biology (and immunology?) and pursue the ideas of *competition, selection, adaptation* (partly) responsible for emergence of "interesting structures" in living systems and, on a smaller scale, in combinatorial games such as chess.

Peano arithmetic.[120]

18 Small, Large, Inaccessible.

Mathematicians treat all numbers on equal footing, be these

$$2,\ 3,\ 4,$$

or

$$10,\ 20,\ 30,\ 40,$$

or

$$1\ 000,\ 10\ 000,\ 100\ 000,\ 1\ 000\ 000,$$

or

$$10^{10},\ 10^{20},\ 10^{30},\ 10^{40},$$

or

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

or

$$2^{3^{4^{\cdot^{\cdot^{\cdot^{399^{400}}}}}}}$$

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

Partly this is because properties of numbers (and formulas in general) depend on how their representations by symbols spreads over the background either in the physical space/time or in the human ergomind.

Even quite small numbers, in fact, everything above four, unless they are represented by "structured entities" in some way, are not accessible to the human (ergo)mind.

Non-accidentally, 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, syntactically speaking, go to the same basket as infinity.[121]

Yet, some psychologists maintain the idea that a few larger numbers

•••••, •••••• and •••••••

are also humanly perceptible but nobody(?) would claim he/she can immediately grasp

••• or even ••••••••••.

Habituation to the decimal notation deceives us. We have no idea what such numbers as 65 536, even less so 7 625 597 484 987, are[122] unless we write them as $2^{2^{2^{2}}}$ and $3^{3^{3}}$.

Even recognition of "*seven trillions*" in 7 625 597 484 987, unlike that of "*million*" in 1 048 756 (= $2^{2^{2^{2}}+2^{2}}$), needs a few rounds of the mental counting algorithm that is not innate to the human mind.

On the other hand, the human (ergo)mind that feels uncomfortable with, say 177148, readily accepts the same(?) number structurally organised as $3^{3^{3}+2}+1$.

[120] See *Brief introduction to unprovability* by A Bovykin [7].

[121] 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.

[122] Browsing through

$$\underbrace{\bullet\ldots\bullet}_{7625597484987}$$

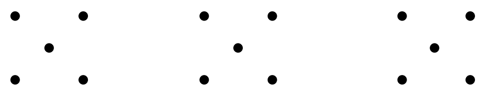
point by point would occupy you for a couple of *hundred thousand* years.

Probably, this is because the *"combinatorial geometry of the (ergo)mind"* is nearer to that of the arrangements of exp-towers in the 2-plane than to the uneventful linearity of •••••••••••••••••••••... as well as of the ordinary positional depiction of numbers and formulas.

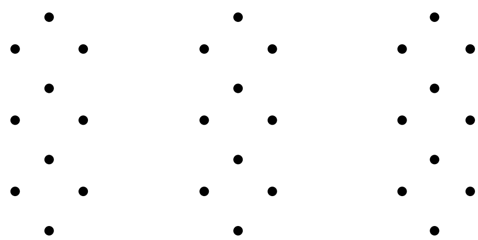
Look closer on how it works. One instantaneously evaluates the cardinality of • • • •, one needs a fraction[123] of a second to identify "(almost) unstructured five" • • • • •, it takes a couple of seconds for • • • • • • • (it is much faster if the symmetry is broken, for instance, as in • • •   • • • •) and it is impossible with

• • • • • • • • • • • • • • •.

But a little structure helps:



And slightly larger numbers, such as



if perceived, then only 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?[124]

What is greater the number of bacteria living in your guts or the number of atoms in a bacterium?[125]

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[126] 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.[127]

[123] This may be a significant fraction, well above 200 milliseconds.

[124] Both numbers are about 100 000.

[125] 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.

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

[127] 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.

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.*[128]

• BRAIN. The number of neurones in the human brain is estimated between ten and hundred billion neurones with hundreds synaptic connections per neurone, 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 neurones – something about 100 times per second. Thus, say hundred million active neurones can perform 10^{10} "elementary operations" per second[129] that is what an average computer does.[130]

•LANGUAGE. There are 10^{22}-10^{25} grammatical sentences in five words in English.

• 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.[131]

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.[132]

IS IT MATHEMATICS?

(1) It seems not hard to show that there exist lots of *provable* theorems T that can be formulated on a page or two in the standard mathematical language but such that their shortest proofs must be enormously long; hence, *humanly*

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

[129] 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.

[130] 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.*

[131] Archimedes, recall, 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.

[132] These turn around the following question. What makes it so difficult to locate/construct *individual* objects O with a property P from a given class C of similar objects, when we know that this P is satisfied by *majority* of members of C?

unprovable. (To be unprovable one does not need "enormous" – "modest" 100^{100} will do.)

(2) On the other hand, pinpointing a *specific* a priori provable T that it is *realistically unprovable.* may be humanly impossible.

(3) A somewhat more problematic (and more interesting) possibility is that there exists a theorem T' that admits a hundred page proof but a search for the proof of T' would need at least 10^{50} human+computer hours; hence, being unfindable.

Explanation. Unlike (1) and possibly (2) that (almost) entirely belong with mathematics proper, this (3) hardly can be approached without making some conjectures on the innate resources of the human (ergo)mind/brain.

In fact, all kind of arguments and ideas – computations, proofs, etc, – that we can imagine and/or design, let this be done with a use of computers, are limited to *compositions* of what is available in the "pool" of *relatively small number* of innate atomic ideas evolutionary installed by Nature into our (ergo)brain along with a collection of composition/selection rules of these ideas.[133]

Therefore, there exist "ideas" expressible in strings of 10^5 decimal symbols, that will be NEVER articulated in any form by humans on Earth.[134]

Shadows of Examples. By their very nature, *humanly inaccessible ideas* can not be exemplified. But the following is suggestive.

($\star$) *"Colorless green ideas sleep furiously"* was, probably meant by Chomsky as a *random grammatical*; hence, absurd sentence. But this is, in truth, a bona fide (sarcastic) English sentence due to pronounced, let them be negative, correlations between the words in it.[135]

(There is no *100% random* option within you ergobrain, but, given a list of words in front of you, you can(?) put your finger blindly on one of them.

NONSENSICAL SEVEN THUS CHOSEN ON RANDOM.[136]

Illegal silly inflations mislead recklessly.
Extended meager materials emphasize fortunately.
Fine joyous departments choose physically.
Wet coordinated articles complain coherently.
Wooden illustrious faults cost halfheartedly.
Wooden coordinated departments emphasize recklessly.
Illegal meager departments complain halfheartedly.[137])

($\star\star$) To scientifically evaluate the intelligence of a monkey, a psychology professor suspended a banana high up under the ceiling, such that it could be reached only from a chair that had to be put on the table.

[133] Some of these "compositions" have been already implemented in the course of human cultural history with incorporation of bits and pieces from the "information flows" from the "external world", but this does not change the essence of what we say.

[134] If – which is unlikely – the number of equivalence classes formed by "meanings" carried by these strings is (very much!) significantly less than 10^{10^5}, then this is objectionable.

[135] Possibly, this sentence came up looking meaningful because it is too short.

[136] These are taken from the pool of, what I guess, 5 000-10 000 most common words out of "full" English vocabulary of more than 150 000 words.

[137] Notice, that the last two sentences are the diagonals of the first five and observe that the longer you look at this sentences the better sense they make – the human brain is apt to detect non-existent meaning in all kind of nonsense.

But the monkey, upon entering the room, discovered a better use of professor's head than the professor himself: the animal jumped on the shoulder of the man, leaped up for the fruit using man's head as a trampoline and safely landed on the table.

Apparently, this solution to the problem was absent from the toolbox of ideas under professor's scull.

19 Probability: Particles, Symmetries, Languages.

Can one reconcile Maxwell's theses that

The true logic of this world is the calculus of probabilities.

with Naum Chomsky's assertion that

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

Human languages carry imprints of the mathematical structure(s) of the ergobrain and, at the same time, learning a natural (and also a mathematical[138]) 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 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.

A good definition must tell you *the truth, all the truth and nothing but the truth*, but there is no common agreement, not even among mathematicians, what constitutes a TRUE DEFINITION.

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,*
e.g. of the equation $2x_1^2 + 4x_2^4 + x_1x_2^3 = 0$,

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 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

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

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.[139] And Titus Lucretius (99 –55 BCE), a follower of Democritus, describes in his poem *De Rerum Natura* what is now called *Einstein-Smoluchowski stochastic model* of Brownian motion[140].

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' ergo-brain (and ours as well) to solve Bouridan's ass problem is to go random.[141] Emanation of the "miraculous decision power of random" intoxicates a gambler's ergo.[142]

The first(?) documented instance of the *calculus* of probabilities – "*measuring chance*" by a European[143] 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*[144] in Luca Paccioli's *Summa de Arithmetica, Geometria, Proportioni et Propor-*

[139]Our sketchy outline of the history of probability relies on [22] [10], [13], [15], [31], [26] with additional *References for Chronology of Probabilists and Statisticians* on Ming-Ying Leung's page, http://www.math.utep.edu/Faculty/mleung/mylprisem.htm

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

[141]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) needs a "seed of randomness" in its software.

[142]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.

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

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

tionalita.[145]

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.

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 encyclopaedias 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 them is known to be a girl?[146])

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*[147] stands among the ten (±2) greatest mathematical theorems of all time.

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

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

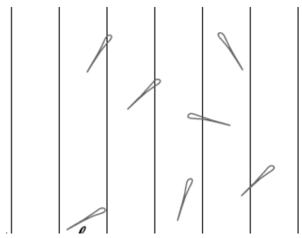
[145] Paccioli became famous for the system of *double entry bookkeeping* described in this book.

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

[147] 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.

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



GEORGES-LOUIS LECLERC BUFFON.

Besides opening the fields of *geometric probability* and *integral geometry*, Buffon, who understood physics of his time, contributed to theoretical and practical optics. As an experiment, he constructed a large (about 2m) concave mirror built of 360 small ones that, by focusing sunlight, could melt iron at 10m distance and ignite wood at 50m.

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 physicists after Buffon.[148]

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.

ON SYMMETRY IN RANDOMNESS.

[148] 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.

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

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.*

Atoms and Bacteria. 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

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

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.[149]

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 Languages.

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.

Probability applicable to languages must be not an assignment of *numbers* to events, but rather a "*functor*" from a "linguistic category" to a "small and simple" category, yet, in general, different from the "category of numbers".

149 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.

And forfeiting 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.

Adaptation of probability to needs of (ergo)linguistic will also require a revision of the concept of "event" the probability of which one measures.

The now-a-days canonised definition of "event", 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 $\blacksquare$ in the plane. You drop a points to $\blacksquare$, 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 $\blacksquare$ 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. But a mathematical interpretation of the idea of probability in *languages* and in *learning* needs a more radical deviation from (modification? generalization of?) this $\blacksquare$.

CARDANO, GALILEO, BUFFON. The very existence of these people challenges our vision on the range and extent 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?

20 Signal Flows from the World to the Brain.

The signals entering the ergobrain via vision, hearing and olfaction are "written" on certain physical/chemical backgrounds the structures and symmetries of which are mathematically quite transparent.

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 three dimensional $time+frequency$ $+amplitude$ 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. Structurally, olfaction, has much in common with gustation as well as with perception of warm, cold and pain, that are, however, more primitive than olfaction in most animals. But olfaction is fundamentally different from vision and hearing.

Scents, unlike images and sounds can not be written and rewritten on a variety of backgrounds they are supported by *specific to each of them* physical/chemical substrates.

The "letters of smells" are aromatic molecules entering our noses. The flows of these, *unless they are bound to a rigid substrate*,[150] have more internal symmetries than the flows of visual and auditory signals: N different kinds of aromatic molecules that freely float in a gas or in a liquid can be permuted in $N! = 1 \times 2 \times 3 \times ... \times N$ different ways.[151]

No, rewriting, no natural digitalisation, no universal structuralization of smells comparable to that of images or sounds is possible.

Because of this smells can not be positionally codified and the olfaction inputs have significantly lower information contents than better organised flows of visual and auditory signals. (This is an apparent reason of why the olfactory perception depends on so many *different*[152] receptors.)

For instance, combinations of two different kinds of molecules ⬡ and □, can encode only a few dozen distinct signals since there is no perceptible order in how groups of these molecules, say from the following string, that may enter your nose.

⬡□⬡⬡□⬡⬡□□□⬡□⬡□□□⬡□□⬡⬡□⬡□□□□⬡⬡□

Even though there are 2^{30} – more than a billion – of such binary strings, your nose can only perceive the relative amount of ⬡ and □. (This is a humble 13/15 for the ⬡□-string.)[153]

Possibly, bloodhounds, who have 200-300 million olfactory receptor neurones in their nasal cavities, may differentiate smells from a pool of the size 10^6(?), 10^9(?), 10^{12}(?)[154], but they cannot beat human visual system that can easily distinguishes randomly taken images from a pool of 10^{20} and with a little effort (training?)[155] up to 10^{1000}.

The internal library of smells is simpler than how we remember visual images, sounds, words and ideas – no ergo is needed for organisation of this "library"

Besides, 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 think we can recall them.

[150] Scents attached to a rigid surface, allow, for instance, tracking by dogs.

[151] This is, of course, an exaggeration. Nothing on Earth has been ever permuted in 50! $>10^{64}$ ways.

[152] There are over 1000 different receptor proteins in these neurones, as it was established by Richard Axel and Linda Buck.

[153] Variation of smell intensity in time does not contribute much, although this may work as a clock for a dog.

[154] Apparently, there is neither a univocal definition of "olfaction sensitivity" no reliable data data for a convincing estimation of the olfaction capacity of people and other animals.

[155] Strings of common words are easy to discriminate but see how long it takes to distinguish the following two.

अ ल ो ए व अ न अ गअ रई हअ तए ह अइ and अ ल ो ए व अ न अ अग रई हअ तए ह अइ.

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 odours, no possibility to encode and "freeze" the flows of olfactory perception.[156] (This is what stopped bloodhounds from developing a language based on olfaction.)

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, 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 in it feel similar to those found in the mathematical theory of *n-categories.*

The combinatorial structure is intertwined with a geometric one; both an individual network and the totality of these, 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 modelled by some transformation(s) in the *space of networks*; the resulting "educated learner" appears as an *attractive fixed point* (or a class of points) in this space under this transformation(s).

The basic property of this (ergo)learning process (transformation) is a compression (or rather, folding) of the *information* carried by the incoming signals, where much of this is achieved by some *coclustering algorithms.*

The result of such learning, call it *(mental) syntactic structure*, probably, possesses significant symmetry, e.g. (non-Lorentzian) symmetry of *space+time* in the case go vision. But this symmetry in mentally structuralized languages seems more sophisticated than what is expressible in terms of the mathematical *theory of groups.*

Apparently, "syntactic structures" in our ergobrains are more "elementary" and more "abstract" than those studied by mathematicians and by linguists, where "elementary", "abstract", "fundamental", "rudimentary" are synonymous concepts from our (ergo) point of view.

Seven Flows.

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

[156] Perfumes do not count.

Incoming flows of signals are naturally classified 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.*

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.[157]

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[158], and it is not *directly* bound to the flow of time[159]. Persistence of written literature is hard to reconcile with a naive selectionist 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 ergobrains. 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 music".[160]

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 sew a button on your shirt.

---

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

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

[159] The time arrow is implemented by *directionality* of what is written.

[160] 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.

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 sensory inputs.

Three flows among these: *Language, Mathematics, Music* have an essential feature in common: the receiver of such a flow $\overrightarrow{F}$ develops an ability, with no external reinforcement, to creatively generate a new flow $\overleftarrow{F}$ in the class of $\overrightarrow{F}$. (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.)

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

The most studied among these is the learning of native languages by children the (still unknown) mechanism of which is must be not far from how mathematics is learned by mathematicians.

The structure of a most sophisticated mathematics we build in our minds is likely to be simpler than that of natural languages (and smaller than that of 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[161] and a presence of criteria for assessing its understanding.

Transformations and Reconstruction of 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_{rec} you receive is an essential aspect of understanding the message carried by F_{orig}. For example, understanding a flow 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.

This reconstruction can be expressed, albeit incompletely, as *annotation* to F_{rec}.

For instance, upon receiving a flat image 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 this image.

[161] 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.

And the background tree structure in a (record of a) flow of speech can be indicated with *parentheses* properly inserted into sentences. (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 neurones involved in eye movements and "understanding" of these images depends on structural matching this activity with similar actions of these neurones 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.

21 Characteristic 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* 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.

For instance, if a sentence starts with "*There are more ... in ...* " one may rightly expect "*than in* " coming next with abnormally high probability.[162]

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.[163]

[162]Try: *there are more * in* on Google.

[163]A universal learning program faces here the difficulty of "understanding" *his* and *again*.

(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.[164] 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.[165]

(5) *Imitation, Repetition and Generation of Linguistic Signals.* Humans, especially children, have the ability to reproduce linguistic signals [*sign*] they receive, including those emitted by themselves, where, to be exact, *not signals* [*sign*] *themselves* are generated but members of the *same class/cluster* as [*sign*] and where the choice of a particular *classification rule* is not a trifle matter.

One can hardly analyse languages without being able to generate them[166], 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 typical 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) that correspond to similar mirrors in human ergobrains.

(Such "mirrors" in vision are, probably, implemented by projections from "deeper" regions of the brain to the primary visual cortex.)

22 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*

[164]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.

[165]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.

[166]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.

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 Descartes 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 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.

To go further, we conjecture that

∘ most (all?) structures we encounter in life, such as natural languages, mathematical theories, etc. are UNDERSTANDABLE;[167],
∘ this UNDERSTANDING universally applies to a large(?) class of *structural*, entities;
∘ UNDERSTANDING is an elaborate structural entity in its own right, thus, being a subject to mathematical scrutiny.

Granted all this we start searching for mathematical models[168] of UNDERSTANDING.[169]

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?

[167]Overoptimistic? Yet, in line with the remark *"... mystery of the world is its comprehensibility"* by Einstein.

[168]When we say *mathematical structure* or *mathematical model* we do not have in mind any particular branch of the continuously growing and mostly hidden from us enormous tree that is called MATHEMATICS.

[169]Impossibility of resolving – not even formulating – the problems of UNDERSTANDING and of *thinking machine* in simple words does not abate one's urge to make the world know what one's *gut feeling* about these issues is – an incessant flow of publicised opinions on this subject matter is a witness to this.

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 H. Florey with his coworkers, with the results published in 1929 under the title *The Vascular Reactions of the Colonic Mucosa of the Dog to Fright* (This the same Florey who was responsible for bringing penicillin to the therapeutic use in 1940's)

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?[170]

QUESTION 8. Can Turing-like tests be performed with *algorithmically* designed formal questions that would trick a computer program to give senseless answers?[171]

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?

Apparently, UNDERSTANDING is composed of three ingredients:

$[\bullet]_U$ *a certain combinatorial(logical?)* ***structure*** $\mathcal{U}$ *in the understander's: mind/brain/program;*

$[\bullet]_{[IU]}$ *a* ***process*** $\mathcal{IMPU}$ *of implementation of* $\mathcal{U}$ *by an ergosystem representing "an understander";*

$[\bullet]_{[RU]}$ *the* ***result*** $\mathcal{RESU}$ *of such implementation,* $\mathcal{RESU} = \mathcal{IMPU}(\mathcal{FLOW})$, *where* $\mathcal{IMPU}$ *is seen as a transformation applied to flows of signals.*

We do not know for sure if *understanding* is a formalizable concept and, if "yes", there is no clear idea of what kind of structure this $\mathcal{U}$ could be.

The only convincing argument in favor the existence of $\mathcal{U}$ 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.

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 $\mathcal{U}$ are organized as a structural community that can be partly described in category theoretic terms.

However, we have a realistic(?) expectation of what space/time characteristics of this structure(s) could be : $\mathcal{U}$ is much smaller in the volume content than the totality of the flow $\mathcal{FLOW}$ this $\mathcal{U}$ "understands" and implementation of $\mathcal{U}$, that is application of $\mathcal{IMPU}$ to a flow $\mathcal{FOW}$, is much faster than achieving understanding $\mathcal{U}$.

It takes, probably, $\approx l \log l$ elementary steps for learning $\mathcal{FLOW}$ of length l that translates to months or years when it comes to learning a language or a mathematical theory.[172] 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 $\mathcal{FLOW}$ is completely novel and even less time to see that a string is meaningless.

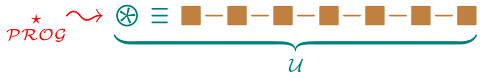
On the other hand, the space/volume occupied by an understanding program $\mathcal{U}$ is a few orders of magnitude greater than a learner's starting program $\mathcal{PROG}$,

[170]Designer's own ability to pass a test is a poor criterion for designing such a test.

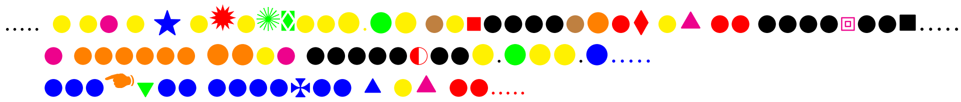
[171]Such questions must refer to an earlier part of a conversation. For instance, the examiner says *"...example..."* at some point and then responding to a question by the program, the examiner says; *"I have already given an example of this in my previous five word sentence."* But to trick a sufficiently sophisticated program one must design a "logical *Russian doll*" of such questions.

[172]The true measure of time, call it *ergo-time*, should be multi-(two?)-dimensional, since it must reflect parallelism in programs modelling learning and other mental processes.

where such a program is *universally* (independently of the total number of signals from $\mathcal{FLOW}$ received/inspected by a learner) bounded by something like 10^6 bits. Picturesquely,



where ⊛ represents the "core understanding" – a few thousand page "dictionary+grammar" of $\mathcal{FLOW}$ – the flow of signals, where this ⊛ is augmented by several (tens, hundreds or thousands depending on $\mathcal{FLOW}$) "volumes"■ of loosely (imagine RAM on your commuter) organized "knowledge", while the available $\mathcal{F}$ itself may number in tens or even hundreds of millions of nearly unrelated units – volumes, internet pages, images – a mess like this:



The branch of mathematics that would support, say human UNDERSTANDING, of the language, must be quite elaborated – it is not grown yet on the "grand tree of math."

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 AlCaLa, between two 4-element sets *understands* the meaning of the alarm calls even if this AlCaLa 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 Descartes and Diderot, not to speak of Turing himself, attach such significance to the Turing test?

Is there an *essential* difference between the correspondence *"questions"* ⟶ *"correct answers"* in a human language and the AlCaLa correspondence?

The answer is:

Yes, there is an essential difference, an ENORMOUS DIFFERENCE.

Operating with tiny sets, e.g. composed of four utterings – alarm calls of vervet monkeys and with correspondences between such sets needs no structure in these sets. But one CAN NOT manipulate human utterings and even less so longish strings of utterings and/or written texts in a structureless way.

It is tacitly assumed by scientifically minded people — Descartes, Diderot, Turing... , that the above correspondence "⟶" 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*

$$D^{30} = \underbrace{D \times D \times D \times \ldots \times D}_{30}.$$

Such structurality is indispensable for an implementation of a "thinking automaton" and/or the program running it in a *realistic* space time model [173] 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.[174]

("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.[175])

23 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.[176]

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.

4. Universality also implies that the actions of the learner – building internal structures and generating signals, both within itself and/or released outside,[177] *are not governed by goals* expressible in terms of the external world.

---

[173]The property of being physically realistic is often missing in philosophical discourses on artificial intelligence.

[174]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?

[175]There is no accepted definition of *"function"* that would separate the *wheat*: $\sin x$, $\arctan x$, $\sqrt{x}$, $\log x$, Riemann's $\zeta(x)$, Dirac's $\delta(x)$,..., from the *chaff* , such as the characteristic function of the subset of rational numbers.

[176]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.

[177]These are the "actions" the human brain is engaged in.

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.[178]

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.[179]

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.*[180]

Our main conjecture is that

[178] It is the rare words in texts that are significant, not the most frequent ones. For instance, *gifku mfink otnid* on three different pages of a book will impress you more than ooooo ooooo ooooo on twenty different pages.

[179] Partition of stars in the sky into constellations is an instance of this.

[180] 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.*

universal learning algorithms that converge to UNDERSTANDING *exist*

and that, 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.[181]

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.

24 Learning to Understand Languages: from Libraries to Dictionaries.

Mathematically, the process of *learning a language, call it* $\mathcal{TONG}$ is represented by an *orbit* of the *universal learning program* $PROG$ that acts on the *linguistic space of* $\mathcal{TONG}$ and where this orbit must eventually (approximately) converge to "*I understand* $\mathcal{TONG}$" state/program.

The principle existence of such a $PROG$ 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".[182]

A closer to our experience scenario is that of a visitor from another Universe[183] who attempts to "understand" what is written in some human "library" LIBR 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* EDI – a kind of "concentrated extract" of the combinatorial structure(s) that are present (but not immediately visible) in flows/arrays of linguistic signals.

Making a dictionary involves several *interlinked* tasks where a starting point is

[181]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.

[182]Bridging linguistic signals to non non-linguistic ones is an essential but, probably, not indispensable ingredient of "understanding Language" as it is suggested by the linguistic proficiency of deafblind people.

[183]No imaginable Universe appears as dissimilar to ours as what the brain "sees" in the electrochemical world where the brain lives.

Annotation &Parsing, that is identification and classification of ***textual units*** that are persistent and/or significant fragments in short strings s (say, 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.[184]

One may think of such annotations as being written in strings positioned on several *levels*[185] 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.

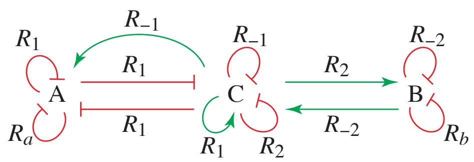
An ergo-dictionary is obtained by several consecutive *reductions* – kind of *compressions*, applied to a library of annotated texts where the resulting combinatorial structure of the dictionary is more elaborate that of the library of annotated texts. The overall reduction in volume is thousandfold, roughly, from 10^9-10^{10} to 10^6-10^7 linguistic units.

The grammar of a language makes a part of EDI where the structural position of this grammar in EDI is supposed to imitate how it is (conjecturally) organised in the human mind.

A particular ergo-dictionary

$$EDI = EDI(\text{LIBR}) = EDI_{PROG}(\text{LIBR})$$

is obtained from a collection of texts, this what we call a *library* LIBR, according to some universal (functorial?) process/program $PROG$ that drastically reduces the size of LIBR 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 edge and vertex colored graph*.



This EDI can be thought of as (a record of) *understanding* of the underlying language by the learner behind $PROG$. This understanding, call it U_t, is time dependent, with EDI being an *approximate fixed point* of the learning process

$$U_{t_1} \underset{PROG}{\rightsquigarrow} U_{t_2},\ t_2 > t_1,$$

where, a priori, $PROG$ can be applied to "understandings" U that were not necessarily built by this $PROG$.

[184]An annotation may include references to *non-linguistic* signals but this would contribute to one's *knowledge* rather than to one's understanding.

[185]These levels l can be regarded as numbers $0, 1, 2, 3, \ldots$ 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.

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 $PROG$ 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.[186]

The apparent ingredients of ergodictionaries EDI that encodes "understanding" and processes for assembling these dictionaries are the following.

Short range correlations,[187] *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.[188]

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.[189]

Recognition of self-referentiality.[190]

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.[191]

More seriously, a validity of a particular program and the resulting linguistic competence of a learner in $\mathcal{TONG}$ can be certified only by comparing the outcomes of several programs, say comprised and executed by visitors from 100

[186]We know that such programs are fully operational in the brains of 2-4 year old children.

[187]Relative frequencies of "events" are essential for learning a language but such concepts as "probability", "correlation", "entropy", can not be applied to languages, without reservation.

[188]The essence of understanding is not so much extracting "useful information" but rather understanding the structure of redundancies in texts. Non redundant texts, such as tables of random numbers and telephone directories do not offer much of what is worth UNDERSTANDING.

[189]Playfulness is the first visible manifestation of "ergo" in humans (and some animal) infants.

[190]Omnipresent self-referentiality, along with "playfulness", distinguishes languages from other flows of signals. The simplest instance of this is seen in *noun*↔*pronoun* linkages.

[191]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".

different Universes. Conceivably, 70 among them will not be able to communicate in $\mathcal{TONG}$ with anybody, but 30 will be able to talk to each other in $\mathcal{TONG}$ and find it interesting. These 30, by definition, UNDERSTAND $\mathcal{TONG}$.

(Competent in $\mathcal{TONG}$ and linguistically naive native speakers will judge differently.)

25 Libraries, Strings, Annotations and Colors.

Libraries LIBR we have in mind may contain 10^7 – 10^{11} *input units* that are possibly overlapping stings of letters starting from a fragment of a word to a paragraph with a few dozen words in it.[192]

Annotation may be seen as an assignment "*colors*" to these strings as well as connecting some units with colored edges, where these colours depicts "essential properties" of the node or edge it is assigned to and serve as *descriptors* of *input units* in (annotated) LIBR and later on in EDI.

A colour on a node u may describe a *presyntactic* property of a string, such as

short string, median string, long string and *frequent string*,

while syntactic/semantic features of strings will be assigned different classes of colors from different classes.

Similarly, a colour on a connective may signify a type of a geometric/temporal relation between strings, such as

overlap, ***contained one in another***, ***close-one-to-another***, ***far-one-from-another***, ***next-to-each-other***, ***in-between***, ***begins-with***,

etc., where these "colourful concepts" come in several subcolor-flavours similarly to (yet, differently from) how it is with lengths of strings.[193]

These connectives are tied together themselves by relations between them. For instance, *closeness* between two strings often comes as *simultaneous containment* of these string into a longer string.

This is essential for enlisting and keeping in memory (pre)syntactic insertions between strings, in particular all pairs of identical words w in L.[194]

And another class of colors such as for similarity $\leftsquigarrow$, and for reduction arrows may refer to particular classifier algorithm defining/producing these arrows.

.

Colors themselves are structural entities but much simpler ones than incoming units such as sentences. One may think of them as simple graphs, e.g. little trees with few branches.

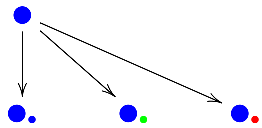
[192]There are also larger units, call them "pages", "volumes", "shelves", but these play different role.

[193]The ergobrain, that processes incoming information in parallel, divides long strings into pieces and keeps track of their mutual positions by means of binary (and ternary such as ***in-between***) relations.

Although the linear order •••••••••••••• can be *formally* reconstructed from ***begins-with*** a baby-like learner is unaware of formal logic and can not iterate indefinitely a single relation.

[194]The number of pairs of (short) identical strings is uncomfortably large being *quadratic*, but identifying identical words say on a single page goes *linearly* in time.

This is gratifying. 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.



The library colors – about 100 in number – pass to EDI, our ergo-dictionary where they are organised into a more elaborate coloured network with additional descriptor-colors, maybe up-to 1000 of them.

A more difficult/interesting aspect of annotation is identifying "significant strings" (this is essentially what is called *parsing*) such as ***word***s, ***phrase***s, etc. and disregarding insignificant ones[195] These also come in coloros and subcolors such as $\textbf{\textit{word}}_{rare}$, for instance.

COLLECTIONS, ENSEMBLES, SETS.

Collections/ensembles of linguistic units and of connections between them can not be indiscriminately operated with as we do it with sets in mathematics 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;[196] yet, we use fragments of this language whenever necessary.

Besides "localized" relations between strings and connectors between them, there is a large scale geometry in LIBR that is seen in the presence of (relatively large) *contextual units* such as ***page***s and ***book***s.

At the first stage of annotation process 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.*

Coloring Colors. It seems not difficult to make a complete combinatorial description[197] of LIBR with a few dozen (about 100?) "colors" that are *descriptors* of *basic types* of units and of connectors between them in LIBR.

But the principal issue is not so much LIBR per se but a construction of an adequate *coloured network of descriptors* on the basis of few, probably 4-8, "general rules". Eventually this network will be laid in the foundation of EDI – that is a *stationary model* of understanding $\mathcal{TONG}$ as it represented by the library LIBR.

[195] An essential (but not the only) intrinsic motivation for doing this by an ergo-learner is economising the memory space.

[196] Bringing forth *random sets* and/or *fuzzy set* may only aggravate the problem.

[197] Implementing this, i.e. achieving a proper annotation and/or parsing is my no means easy.

26 Teaching and Grading.

A universal language learning problem *PRO* is supposed to model a mind of child and it needs only a minimal help from a "teacher", such as ordering texts according to their complexity[198] 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 of 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.[199]

Similarly, understanding languages depends on compression of structural redundancies[200] 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 *PROG* 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 *PROG* minimally proficient in English, would "know", upon browsing through Encyclopaedia Britannica, that cows eat grass and cats

[198]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.)

[199]Ancients astronomers came to *understand* periodicity of planetary motions and were able to make rather accurate predictions.

[200]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.

eat mice.[201]

Also, the following can be seen as a hallmark of understanding.

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 *PROG* may be graded according to their "internal characteristics", such as the *volume of the memory* a *PROG* has to use, the *number of elementary operations* and *the time* needed for it to make, for instance, a particular prediction.

Eventually, full-fledged ergo-systems must contain *self-control programs*, such as an evaluation of *increase* of quality of predictions with extra information getting available to the learner. (Such control programs are much easier to design than the core ergo-programs.)

27 Atoms of Structures: Units, Similarities, Cofunctionalities, Reductions.

Much of learning and understanding consists in *structuralizing* incoming flows of signals you perceive that is achieved 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 a few their ingredients are visible.

Let us make a short (and incomplete) list of four "logically (quasi)atomic constituents" of (ergo)operations applied to flows of signals with no attempt at this point to give precise definitions of these "atoms", to justify their reality, and/or to explain how one finds them in flows of signals.

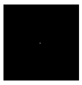
1. DISCRETIZATION AND FORMATION OF UNITS BY SEGMENTATION AND PARSING.

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". (The *internal* correlations/connections in these "parts" must be significantly stronger than *mutual* correlations/connections between different "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.* Also 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.*

[201] 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 Encyclopaedia Britannica by *PRO*.

And even if, say, paintings and ■'s, both are regarded as "units", 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".

On the other hand, *proprioception sensory system*[202] and motor control of skeletal muscles may also depend on *continuity*, since the incoming signals may be not(?) naturally decomposable into "discrete units".)

But our ergo structure is built from invisible *internal* units that may be grossly dissimilar to the units of incoming flows; one needs truly universal *discretizers* – "meaningful segmentation" algorithms to discern these.

Naively, *unit* is anything that can be 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.

We follow the lead of natural languages that make units of everything: *qualities, states, actions, processes,...* by what is called *nominalization*:

everything deserving a name becomes a unit.

Example: Non-Textual Syntactic Units.[203] In order to implement this "definition", one needs to to design a program that would "understand" what "deserves" is. This is most essential for understanding Languages, that unlike understanding non-linguistic arrays of signals, decisively depends on formation of units that are *not geometric fragments of texts.* For instance, the groups of superficially dissimilar words, such as

yes, no, maybe; we, us, our; big, large, huge; smelly, tasty, crunchy.

are kind of "outlines" of such units. Identification of these is an essential aspect of of passage from a library to an ergodictionary.

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

[202]This is the perception of motion, of stresses and of position of parts of the body.

[203]The word "syntactic" is understood in the present article as "characteristic of languages".

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 logically indoctrinated mathematician – these two *"equal"* are not mutually equal and the common language has no means to express this inequality. For example,

5 is the same as 5

does not make it 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*.[204]

On Composability of Similarities. Customary, one defines an *equivalence* as a *symmetric binary relations* on a set[205] 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* $\sim$" 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

the composition $f \square g$ 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, Reduction, Clustering, Compression.

Equivalence relations E on a set S go hand in hand with partitions of this S into the corresponding *equivalence classes* that can be conveniently described via the *reduction map* $R = R_E$ from S *onto* a C, such that

$$s_1 \underset{E}{\sim} s_2 \text{ if and only if } R(s_1) = R(s_2).$$

However, implementations of a *binary* relation $s_1 \underset{E}{\sim} s_2$ and of a *unary* operation $R(s)$ are quite different from a working ergosystem point of view.[206]

It is much harder to record $\approx N^2/2$ bits encoding an equivalence relation on a set S with N units, than $\approx N \log N$ bits needed for defining $R(s)$; *similarities* and *reductions* must be treated separately.

[204] The spirit of this is close to how different levels of "equivalence" are treated in *the n-category theory*.

[205] This definition does not cover equivalencies between theories and/or between categories since these are *are not* relations on *sets*.

[206] 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*.

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 with a 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 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 without fully degrading its structure.

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 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 it is dangerous to force categories into languages prematurely.

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.

To see what we mean, imagine you have two books, each approximately 200 pages long. Choose anyway you want the numbers of pages, say 150 numbers altogether, and count how many times *"the"* appears on each chosen page i of the books 1 and 2.

Similarity of these two *"the" contents* N_1 and N_2 such as

$$N_1 - 2 < N_2 < N_1 + 2,$$

for a single pair of pages is not informative, but if this relation holds for all 150 pairs of your chosen pages, you bet that 1 and 2 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 it is helpful to assume so when defining *co-clustering*); 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 $s_1 \underset{f}{\longleftrightarrow} s_2$.

CONNECTIONS BETWEEN UNITS:
THEIR IDENTIFICATION, NOMINALISATION AND CLASSIFICATION.

Deciding which units are essentially independent and which have non-trivial *connections/relations* between them, is one of the first priorities of an ergo system that must be accomplished by several algorithms, that can be called *connectors*.

Such connectors, being themselves particular kinds of units, need to be classified by universal algorithms as it befits all decent units, where the coarsest classification would separate the following three classes of relations.

- *Similarity.*
- *Cofunctionality.*

(The latter is common not only for pairs but also for triples and possibly, quadruples of units that *perform together* certain functions. This "togetherness" is manifested by systematic co-appearance of the corresponding units.)

- *Reductions.*

WARNING. Treating relations on an equal footing with initial units opens pandora box of self-referentiality in our ergo system(s). Apparently, this is necessary for the kind of ergo-behaviour we want to achieve but this is also inoculates our systems with logical paradoxes. Somehow one has to strike balance between dumbness and madness of ergo-programs.

After we have developed algorithms for the structural analysis of " incoming flows of signals" along the above guidelines, 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 while thinking mathematically, we also need to keep in mind possibilities and limitations of the brain, ergo-algorithms must be "broad and shallow": they 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 hundred (thousand?) operations running in parallel.[207]

28 Fragmentation, Segmentation and Formation of Units.

Certain fragments of incoming signals[208] e.g. particular strings of letters such as *"words"*,[209] or some distinguished regions in visual fields, such as "*perceived objects* or "things"[210] 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.[211]

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

[207]This parallelism is the "technical reason" why our basic mental (ergo)processes are inaccessible to our sequentially structured conscious minds.

[208]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"*.)

[209]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.

[210]The rigid concept of *object-unit* modifies by classification/reduction and applies 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.

[211]The number of different books in the world is estimated at about 100 million.

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.[212]

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 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 incorporates into the full "understanding space" on the later stages of learning.

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 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.

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 $10 \leq \underline{n} \leq 30$. 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[213] 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 di-

[212]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.

[213]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.

vided 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.

29 Presyntactic Morphisms, Syntactic Categories and Branched Entropy.

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 $100^{10} << n^{10}$ for n being the number of symbol-words in a language.[214] As for L one might think 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.

JUSTIFICATION.

[+] *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

[214] Never mind the saying: "*there are infinitely many possible sentences in a natural language*".

strings would enhance the *robustness* of categorical descriptions of *written* languages as well.

FEATURES

[$*$] *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.[215]

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 "non-linguistic" 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 is achieved by operations of *reduction*[216] applied to (the sets of) strings and arrows.

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 $\mathcal{D}^{\hookrightarrow\downarrow}$

[215]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 $\mathcal{L}$.

[216]This is also called *clusterization, classification,categorisation, factorization.*

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 $\mathcal{D}_{\hookrightarrow}$.

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.

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.

DIMENSION OF VISION. Visual signals[217] 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.

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.[218]

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.

[217]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.

[218]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.

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 rectangles 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

$$p_{\vec{s}}(t) = p_{L,\vec{s}}(t) = N_L(st)/N_L(s),$$

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.

Such entropies apparently are quite different for the strings *"birds-fly"* and *"pigs-fly*[219] while *"bats-fly"* will be close to *"birds-fly"* in this respect.(This is more pronounced for extensions not of the strings themselves but of their "syntactic variations".)

30 Similarities and Classifications, Trees and Coordinatizations.

Many (most?) linguistic "units", are classes of other units. For instance, *the words* are, in reality, *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-..., inoculation-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...*

[219]The two strings have comparable frequencies on Goggle

represent two different "bat" class-words.[220]

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. For instance, the similarity $\bullet \sim_1 \circ$ is different from $\bullet \sim_2 \blacksquare$ as well as from $\triangle \sim_3 \square$, where $\sim_2$ and $\sim_3$ are similar themselves and dissimilar from $\sim_1$

Not all similarities lead to what may be called " true classification", partly because the "equivalence axiom" $A \sim A$ is not satisfied in ergo-logic. Indeed you would become mad if you fill you brain with $A \sim A$ for all A in you head.

Also some similarity concepts are applicable only to small groups of objects, such as what brings together

{*sweet, bitter, salty, sower, tangy*},

and that do not extend to majority of words.

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

{*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.

The two common classification structures are as follows.

1: Classification as a Tree. This is may be seen as a sequence of partitions of the, say of words units into finer (smaller) classes, where the rule defining each following partition depends on the previous (coarser) one.

A linguistically rather artificial instance of that, is classification/positioning of words in alphabetically organised dictionaries.

More significant example is where the first partition divides words into the two classes **A** and **B**:

A. Class of content words: {*nouns*, (most) *verbs, adjectives, adverbs.*}

B. Class of function words: {*articles, pronouns, prepositions, etc.*}

The classes of the second partition are obtained by subdividing words into "parts of speech".

Then the third partition may divide content words according to their "overall meanings", e.g. nouns according to whether they represent "physical objects" or "abstract concepts", etc.

2: Classifications by Coordinates. These are given by several *coordinates* that are functions on objects we try to classify, where determination of

[220]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 "quasi-poetic bridges" between the two classes.

On the other hand, there are more – about a dozen – different class-words spelled "bat", that are, essentially, subclasses of [**bat-hits**].

each coordinate does not depend on the values of the rest of the other coordinates.

The 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 age group (3-6, 7-11, 12-...) of a child who is able to understand the phrase u.

In general, coordinatizatin "imbeds" a potentially large set U of units to the *coordinate space* that is the Cartesian product of several small sets.

Our goal is formulating *universal* classification rules *a priori, applicable to all kinds of strings* s well as differently structured signals that would be as good, eventually better, than classifications based on "meaning".

31 Clustering, Biclustering and Coclustering

.

There can be no isolated sign. Moreover, signs require at least two Quasi-minds.[221]
CHARLES SANDERS PEIRCE.

Suppose certain pairs of points in a set V of "units" are connected by edges, graphically $\circ\!\!-\!\!\!-\!\!\bullet$, that represents certain *resemblance* between these points regarded as vertices of a graph on the vertex set V.

Simple Clustering of such a V is a partition of V into subsets, called *clusters* such that that (the numbers of) the connections between members of clusters are stronger (more numerous) than interconnection between different clusters.

The archetypical clustering is division of a graph into its *connected components*, but, in general, there is no mathematical clustering recipe applicable to all graphs – after all, many graphs are non-clusterable.

On the other hand, many "resemblance graphs" in life admit more or less non-ambigous dissection into clusters.

Simple clustering often applies to sets V with a distance (or distance like) function $d(v_1, v_2)$, e,g, to subset in the n-space with the *ordinary Euclidean* metric.[222]

Biclustering is more intersting than simple clustering. For instance, imagine a language that admits a simple general definition of *word-unit* and where there is 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.)

[221]Mystifying but inspiring.
[222]This allows one to blissfully use precooked formulas from the book.

Biclustering is a clssification of words according to their *functions*: 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 then you rewrite it as

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

Now, it points toward an iterative process (algorithm) that transforms a prelimenaty clusterization to a more advanced one.

It is often difficult to define and/or to 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.[223]

Both, "*close together*" and *"often"* are variable, where the latter must be adjusted to the former: what is regarded *often* for coming next to each other will be considered rare for the simultaneous presence of these words on the same page in a text.

Granted a specification of *"close"* and *"often"* one arrives at what is called the *co-occurrence graph* G on the set W of words,[224] where w_1 is joined with w_2 by an edge if the two "*often come close together*".[225]

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

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

On terminology.

Division of "objects" into classes is called *categorization* in linguistic and in psychology, but we prefer to call this *classification*.

A classification achieved by some reduction $G \to \underline{G}$ 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".[226]

We use the term "coclustering" where functional cooperation may involve more than two units and reserve "biclustering" to the above case of binary cooperation.

Biclustering applies not so much to graphs but rather to *functions* in two variables, $G(u, v)$ where the domains U of u and V of v do not have to be equal.

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

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

[225] Since we are now firmly on a mathematical ground, ambiguity of "*often come close together*" , or vagueness of terminology on a preliminary stage of exposition, in general, poses no danger. Firstly, because it is *not disguised* as being something precise, and secondly because mathematics promotes crystallization of valid ides into logically solid ones.

This is different from what happens to almost all speculative considerations fuelled by "raw intuition" unaided by mathematics.

[226] 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.

(Graphs are reperesent by 2-valued *edge/no-edge*, or $\{0,1\}$ for brevity's sake, functions.) Namely,

⋆ *reduction of* $G(u,v)$ to a function $\underline{G} = \underline{G}(\underline{u},\underline{v})$ that is defined on a pair of smaller, often significantly smaller, sets $\underline{U}$ and $\underline{V}$ is a pair of maps from U onto $\underline{U}$ and from V onto $\underline{V}$ say $P^{\downharpoonleft} : U \to \underline{U}$ and $Q^{\downharpoonleft} : V \to \underline{V}$ written as

$$u, v \rightrightarrows \underline{u} = P^{\downharpoonleft}(u), \underline{v} = Q^{\downharpoonleft}(v),$$

such that the composition, sometimes called *superposition*, $G^{\downharpoonleft}(u,v)$ of the functions $\underline{G}$ and $R^{\downharpoonleft}$, that is

$$G^{\downharpoonleft}(u,v) = \underline{G}(\underline{u},\underline{v}) = \underline{G} \circ P^{\downharpoonleft} \& Q^{\downharpoonleft}(u,v) = \underline{G}(P^{\downharpoonleft}(u), Q^{\downharpoonleft}(v))$$

provides a *"good approximation"* to the function $G(u,v)$.

(If $U = V$ and $G(u,v) = G(v,u)$, then one may take $\underline{U} = \underline{V}$ and $P^{\downharpoonleft} = Q^{\downharpoonleft}$. But if $G(u,v) = G(v,u)$ then $\underline{U}$ and $P^{\downharpoonleft}$ are not necessarily equal to $\underline{V}$ with $Q^{\downharpoonleft}$ even if $U = V$.)

⋆ *Clusters* (classes) are the subsets in U and in V corresonding to $\underline{u}$ and $\underline{v}$ via $u, v \rightrightarrows \underline{u}, \underline{v}$, i.e. the subsets of those u in U and v in V for which $P^{\downharpoonleft}(u) = \underline{u}$ and $Q^{\downharpoonleft}(v) = \underline{v}$.

"*Approximation*" here is different from the above positional closeness of words. It depends on what kind of function G is, where its values lie. In our examples, besides being a two valued function it may be *three valued*, saying whether u and v cooperate *strongly, weakly* or *not at all.* Also it may be a *number valued* functions with $G(u,v)$ being the relative frequency of co-occurence of u and v.

In either case, the range of the function, denote this range by I, must be equipped with a *metric* measuring the distances between different values.

If I equals a set of positive numbers, then one takes the absolute value $|i_1 - i_2|$ for such a distance, and if I is an "abstract" two or three point set, one may implement these by numbers, say by $\{0,1\}$ and $\{0,1,2\}$ in and use the distance $|i_1 - i_2|$ again.

And then *closeness* may be defined relative to the *Hamming distance* (also called l_1-*distance*)

$$dist(G, G^{\downharpoonleft}) = \sum_{u \in U, v \in V} |G(u,v) - G^{\downharpoonleft}(u,v))|$$

To see how this works, let

∘ $U = V$ be a set that is comprised of 100 000 words,

∘ u and v in W be regarded as *"cofunctional"* if v goes *right after* u.

∘ "*often*" means *"at least ten times"*.

The function G corresponding to this can be seen as a (non-symmetric) 100 000 × 100 000 matrix with $\{often, rare\}$ entries. A reliable evaluation of the values $G(u,v)$, that are the entries of this matrix, needs a library of more than $10^{11} = 10 \cdot 10^{10}$ words in it.[227]

But it may happen, and it does often (albeit approximately) happen in "real life", that this huge matrix is (approximately) determined by something mach

[227]If you check one pair of words per second - eight hours a day - five days a week, it will take more than 10 000 years to go through such a library.

smaller, say by a 300×300 matrix, where you need only 90 000$< 10^5$ entries to fill in description of which needs only

$$90\ 000 + (2\log_2 300) \cdot 10^5 < 2 \cdot 10^6 \text{ bits}$$

instead of the original 10^{10} bits.

And biclusterization serves for achieving such simplification by reduction of your big matrix/function G to an $\underline{G}(\underline{u},\underline{v})$ that is defined on a set $\underline{U} \times \underline{V}$ with $300 \times 300 = 90\ 000$ elements in it.

Quasi-Uniqueness of $\underline{G}$ and $P^{\downharpoonleft}\&Q^{\downharpoonleft}$.

The existence of a reduction $G \rightsquigarrow \underline{G}$ is, a priori, *extremely unlikely* even if we do not require $G^{\downharpoonleft}(u,v) = \underline{G} \circ P^{\downharpoonleft}\&Q^{\downharpoonleft}(u,v)$ to be an especially fine approximation of $G(u,v)$.

Therefore, even if certain $G^{\downharpoonleft}(u,v)$ delivers only a rough approximation to $G(u,v)$ the corresponding $\underline{G}(\underline{u},\underline{v})$, $P^{\downharpoonleft} : U \to \underline{U}$ and $Q^{\downharpoonleft} : V \to \underline{V}$ will be (essentially) *unique with an overwhelming probability.*

But if the sets $\underline{U}$ and $\underline{V}$ are small, say consisting of 2-4 elements, then there may be several candidates competing for the roles of $\underline{G}$ and $P^{\downharpoonleft}\&Q^{\downharpoonleft}$ and one has to select the "best ones". A preferred choice here is where the function $\underline{G}$ is the farthest from the most probable one. (This may be sometimes formulated as *minimization of a kind of an entropy* of $\underline{G}$ as we shall see below.)

Completion/Extrapolation of the Matrix G.

You easily keep in you memory 100 000 words and you (subliminally) remember a few million occurrences of some pairs of them coming close together.[228] But, certainly, this never comes anywhere close to 10 000 000 000 – the number of entries in the matrix G: most of G in your head is filled with question marks.

On the other hand, a couple of million examples is not so little when it comes to the matrix $\underline{G}$ that has only 90 000 entries, And if you have these, you replace $G(u,v) = ?$ by $G(u,v) = G^{\downharpoonleft}(u,v) = \underline{G}(\underline{u},\underline{v})$.

As a result, when necessity arrives, you will not hesitate to accept or to reject as implausible pairs of relatively rare words coming next to each other, such as "intellectually posterior", "hydraulically superior", "superior posterior", "intellectually hydraulically" , "corticofugally inhibited" , "intellectually candied", etc.[229]

Additive and Probabilistic Biclustering

Let all u in a set U carry *weights*, that are positive numbers denoted $|u|$. Then an *additive reduction* of such a *weighted space* is a map that add up the weights. Namely, if, say, $P^{\downharpoonleft}$ maps U onto some $\underline{U}$,

$$U \overset{P^{\downharpoonleft}}{\to} \underline{U},$$

then this $\underline{U}$ is endowed with weights $|\underline{u}|$, that are sums of the weights of all u that go to $|\underline{u}|$,

$$|\underline{u}| = \sum_{P^{\downharpoonleft}(u)=\underline{u}} |u|.$$

[228]An average book contains about 100 000 words, and there are other sources of words besides books,

[229]In truth, you need several coclustering mechanisms working in parallel, e.g. to reject offhandedly "clusterizations scries" and alike.

Similarly if the entries of a matrix (function) $G = G(u,v)$ are positive numbers, then the *additive reduction* $\underline{G}$ of G under

$$U \overset{P^{\downarrow}}{\to} \underline{U}, \text{ and } V \overset{Q^{\downarrow}}{\to} \underline{V}$$

is where the weights of the entries from G add up to the corresponding entries in $\underline{G}$,

$$|G(\underline{u},\underline{v})| = \sum_{P^{\downarrow}(u)=\underline{v},Q^{\downarrow}(v)=\underline{v}} |G(u,v)|.$$

Normalization and Probability. A positive weight function on a set U, $v \mapsto |v|$ is called *normalized* if these weights add up to one,

$$\sum_{u \in U} |u| = 1.$$

Then these weights $p_u = |u|$ are interpreted as "*probabilities of the events* u"

In general, e.g. when the weights represent frequencies of occurence of u, one normalizes by setting

$$|u|_{prob} = \frac{|u|_{freq}}{\sum_{u \in U} |u|_{freq}},$$

thus turning $(U, |...|_{freq})$ to a *probability space* $(U, |...|_{prob})$.

CLUSTERIZATION OF LETTERS: **VOWELS & CONSONANTS.**

Additive probabilistic clustering is practical for small sets W, e.g. where A equals the set of 26 letters in the English alphabet, where the roughest such clustering divides the alphabet A into two classes, denote them $\circ$ and $\bullet$, according to relative frequencies of pairs of letters in English.

The weighs $|a|$ and $G(a,b)$ we use here are normalized frequencies of a and ab in a given text and we search for the (mathematically most natural) reduction, where the four weights

$$\begin{matrix} |\underline{G}(\circ,\circ)| & |\underline{G}(\circ,\bullet)| \\ |\underline{G}(\bullet,\circ)| & |\underline{G}(\bullet,\bullet)| \end{matrix}$$

of the entries of the 2×2 matrix $\underline{G}$ have *minimal relative entropy* with respect to the matrix of the products of the weights in $\underline{A} = \{\circ, \bullet\}$ that are

$$\begin{matrix} |\circ| \cdot |\circ| & |\circ| \cdot |\bullet| \\ |\bullet| \cdot |\circ| & |\bullet| \cdot |\bullet|. \end{matrix}$$

Most likely, – this is certainly known, but I did not check it – this minimal entropy division of A into the classes $A_\circ$ and $A_\bullet$ coincides for the most part of it with the division of letters into vowels and consonants.

Graphically, G (approximately) "reduces" to the two vertex graph $\circ\!\!-\!\!\bullet$ by dividing the vertex set A into two classes/clusters

A=**vo**wels & **co**nsonants

$$\overset{\mathbf{vo}}{\circ} - \overset{\mathbf{co}}{\bullet}.$$

Observe that this partition of A does not depend on any a priori knowledge of the "nature" of letters, but only on the relative frequencies of letters and pairs;

the idea of *meaning* we attribute to these classes is not the source, but *a product* of the proccess of mathematical structuralzation where multiple biclusterization plays an essential role.[230]

GEOMETRIZATION OF G.

To be specific, let a function $G(u,v)$ in two variables take values in the set $\{0,1,2\}$ (standing for *nohing*, *something*, *much*) and let

$$U \overset{\mathcal{G}}{\to} \{0,1,2\}^V \text{ defined as } u \overset{\mathcal{G}}{\mapsto} f(v) = g_u(v) = G(u,v)$$

be the tautological map from the domain U of u to tne space of $\{0,1,2\}$-valued functions $f(v)$ on the domain V of v, that is the Cartesian product space of copies of $\{0,1,2\}$ indexed by v from V,

$$\{0,1,2\}^V = \underbrace{\{0,1,2\} \times \{0,1,2\} \times ... \times \{0,1,2\}}_{V}.$$

For instance, let U be the set of 100 000 words and V a particular subset of say, 30-100 words selected by some preliminary mathematically defined process. For instance, these may be

• 100 most frequent words,

or, more interestingly,

★ 100 most frequent words from some class obtained by another biclustering algorithm,

such as

○ a representative group of function words,

∗ 100 most common verbs.

✤ list of 30 common four legged animals

☞list of 30 common professions.

The space $\{0,1,2\}^V$ comes with many distance-like functions where a prefered one is the Hamming distance

$$dist_V(f_1(v), f_2(v)) = \sum_{v \in V} |f_1(v) - f_2(v)|,$$

that passes to U via $\mathcal{G}$.

Then (bi)clustering of U according to possible coworkes v of u, may be achieved by simple clustering relative to such a "distance" on U.

RECIPROCAL CLUSTERING.

If you choose a subset V of words in U on random, then there will be no preferred clustering of U for the distance coming from $dist_V$. On the other hand some exceptional V would lead to "clean" clusterings of U. Such special V are make tight knit group of similar words, such for instance as: {*the*, *a*} or

{***Red***, ***Green***, ***Yellow***, ***Orange***, ***Purple***, ***Pink***, ***Brown***, ***Black***, ***Gray***, ***White***}.

COMBINATORIAL CLUSTERING: WHY PIGS DO NOT FLY.

Let again G be a graph on the vertex set V. Then the vertices v of this graph can be classified/clustered according to the combinatorics of subgraphs comrpised of vertices and edges in the vicinity of v, where the simplest characteristic of such "vicinity" is the *valency of d* that is the number of edges attached to v.

[230] Phonetically more accurate clustering needs tracking *triples* (quadruples?) of letters that will allow distinguishing certain pairs, such as "th" and "wh", for instance.

Thus, for instance, one may first divide V into two parts V_{small}, where this valency is small, and V_{large}, where it is large, and then subdivide further according to the values of *pairs of numbers* of edges from v to V_{small} and to V_{large}.

Pigs, Pigeons and Sparrows. "Birds fly" on 14 000 000 Google pages, *"Pigs fly"* on 3 000 000 pages, *"pigeons fly"* on 1 000 000 pages and "sparrows fly" on 500 000 pages.

What distinguishes pigs from birds is not the sheer numbers of occurrences of sentences with "pig&fly" or "bird&fly" in them, but the numbers of combinatorial structures displayed by these sentences: there are by far less types of sentences with the words *"pig"* and *"fly"* in them, than of those with *"pigeon"* and "sparrow" replacing *"pig"*.

This is similar to the use of prepositions in English, e.g. *under* and *in*, that may be accompanied by *different* kinds of nouns and/or verbs; yet the geometries/combinatorics of their vicinities in the *"network of short English sentences"* look, nevertheless, quite *similar* that brings all(?) preposition to the same cluster.

On Biclustering Algoritms.

Detection of clusters of "natural units" u in a U, e.g. of words, may not need the full knowledge of $G(u, v)$ at all (u, v) but only for v taken from special *small subsets* in V. For instance, "*the*" divides other words into two groups according to their systematic occurance just before ot just after "*the*".

This does not work for general G and biclustering of sets say of cardinalities 100 000, associated to reductions $U \to \underline{U}$ and $V \to \underline{V}$ to sets $\underline{U}$ and $\underline{V}$ of cardinalities of order 300 seems, generically speaking, computationally unfeasible.

On the other hand there is a variety of heuristic algorithms that work pretty well for finctions G coming from "life".

Words in Contexts: Biclustering and Triclustering.

Biclustering may be applied to the function $G(w, x)$ where w are words and the variable x represent a context, e.g a book from some collection X.

The natural function G encodes (frequent) presence/absence of a w in x and biclusterization serves to classify books by topics according to their "key words" while the words themselves become classified by topics they frequently used in, such as: *chemistry of plants, animal foods*, etc.

Structurally more informative classification, e.g. with an organisation of classes as *trees* with several branches, may be achieved with tri-clusterization for $G(w_1, w_2, x)$ recording pairs of words (w_1, w_2) that appear in the same book x.

In general, however, it is unclear how to proceed with tri-clustering, partly, because there is no convincing counterpart to the above "geometrization" $\mathcal{G}$ of G. On the other hand, most(?) multiple interactions appear as "combinations" of binary ones and multi-clustering reduces to several biclusterings.

Conclusion.

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 but rather 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: reductions of $U\&V$ that lead to (approximate) clusterizations of U and/or V are (not always) composable and combinatorics of systems of (not quite) commutative diagrams of reductions represents an interesting "higher order structure" in $U\&V$.[231]

And besides mere classification a more subtle structure of a language may be extracted from a "distance" on V induced by the above $\mathcal{G}$ from a space of function on some auxiliary set V, where the essential properties of such a "distance" are encoded by (not-quite) *category* of approximate partial isometries of U with respect to this "distance".

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.

32 Bibliography.

References

[1] Bhaskar Acharya, The Ecology of the Dhole or Asiatic Wild Dog (Cuon alpinus) in Pench Tiger Reserve, Madhya Pradesh.

[2] Amato: http://www.popsci.com/science/article/2013-02/when-brain-damage-unlocks-genius-within

[3] E. Anderesen, A.Dunlea, The impact of input: language the visually impaired, First Language, 13 (1993) pp 23-49.

http://fla.sagepub.com/content/13/37/23.full.pdf

[4] Inge Auerbacher, Finding Dr. Schatz: The Discovery of Streptomycin and A Life it Saved, 2006

[5] 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). http://WilliamCalvin.com/1980s/1983JTheoretBiol.htm.

[6] Lennard Bickel, Florey: The Man Who Made Penicillin

[7] A. Bovykin, Brief introduction to unprovability,

https://www.cs.umd.edu/ gasarch/TOPICS/largeramsey/bovINTRO.pdf

[8] Kevin Brown, Penicillin Man: Alexander Fleming and the Antibiotic Revolution, 2005

[231] Heuristic algorithms for iterative clusterization are customary performed in terms of *neural networks*. Alternativly, combinatorics of superpositions of functions can be described in the language of *operads* and *multicategories*.

[9] Cephalosporins and Penicillins: Chemistry and Biology edited by Edwin H. Flynn

[10] J-P Cheng, The Origin of Probability and The Problem of Points, www.math.rutgers.edu/~cherlin/History/Papers2000/cheng.html

[11] Heather L. Van Epps. René Dubos: unearthing antibiotics,

http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2118194/

[12] Michael Dunnill, The Plato of Praed Street: the life and times of Almroth Wright 2000, Royal Society of Medicine Press

[13] First probabilists: Galileo, Cardano, Fermat, Pascal, Huygens, *www.bricol.net/downloads/misc/HistProb/*02.*First_Probabilists.pdf*

[14] Simon E. Fisher, Gary F. Marcus, The eloquent ape: genes, brains and the evolution of language, Nature Reviews Genetics 7, 9-20 2006.

http://www.psych.nyu.edu/gary/marcusArticles/Fisher%20Marcus%202006.pdf

[15] A Hald, A History of Probability and Statistics and Their Applications before 1750, Wiley Series in Probability and Statistics, 1990.

[16] In Memoriam, H. W. Florey: an Episode - Microbiology mic.microbiologyresearch.org/content/.../10.../00221287-61-3-289?...

[17] R. Johnson and S. Liddell, *Toward a Phonetic Representation of Signs: Sequentiality and Contrast*, Sign Language Studies Volume 11, Number 2, (2010) pp. 241-274. (Only the first page of the article is freely accessible.)

[18] F. Lawvere and S. Schanuel, Conceptual Mathematics, Cambridge University Press, 1997.

[19] The Handbook of Linguistics, Mark Aronoff, Janie Rees-Miller editors, Blackwell Publishing 2002-2007.

[20] G. Macfarlan Alexander Fleming, The Man and the Myth

[21] G. Macfarlan Howard Florey, The Making of a Great Scientist, Oxford University Press.

[22] 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.

[23] Carol L. Moberg, Zanvil A. Cohn, Launching the Antibiotic Era: Personal Accounts of the Discovery and Use of the First Antibiotics

[24] 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].

[25] Jason Padgett Struck by Genius: How a Brain Injury Made Me a Mathematical Marvel.

[26] Sources in the History of Probability and Statistics www.cs.xu.edu/math/Sources/

[27] Robert D Potter, The race to get enough penicillin (1944)

http://clickamericana.com/eras/1940s/the-race-to-get-enough-penicillin-1944

[28] J. Schmidthuber, Formal Theory of Fun and Intrinsic Motivation and Creativityhttp://www.idsia.ch/~juergen/

[29] Selman Waksman: the Father of Antibiotics

http://www.jbc.org/content/279/48/e7

[30] Selman Waksman and Antibiotics,

http://www.acs.org/content/acs/en/education/whatischemistry/landmarks/selmanwaksman.html

[31] G. Shafer and V.Vovk. The Sources of Kolmogorov's Grundbegriffe, Statist. Sci. Volume 21, Number 1 (2006), 70-98.

[32] Teaching and Learning: International Best Practice, edited by Dennis M. McInerney and Arief Darmanegara Liem.

[33] Darold A. Treffert,Savant Syndrome: An Extraordinary Condition A Synopsis: Past, Present, Future, http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2677584/

[34] Extraordinary People : Understanding Savant Syndrome (2006)

[35] Selman Waksman, My life with the microbes.

[36] Selman Waksman and Antibiotics commemorative booklet produced by the National Historic Chemical Landmarks program of the American Chemical Society in 2005 (PDF).

[37] Who Saved the Most Lives in History - Science Heroes News

http://www.scienceheroes.com/index.php?option=com_content&view=article&id=258&Itemid=27

[38] Wolpert, Daniel M., Diedrichsen, Jörn; Flanagan, J. Randall "Principles of sensorimotor learning". Nature Reviews Neuroscience Nature Reviews Neuroscience 12, 739-751 | doi:10.1038/nrn3112 http://www.nature.com/nrn/journal/v12/n12/abs/nrn3112.html

[39] Qin Zhua, Geoffrey P. Binghamb, Human readiness to throw, Evolution and Human Behavior 32 (2011) 288-293.