Ergostuctures, Ergologic and the Universal Learning Problem: Chapters 1, 2.

Misha Gromov

March 14, 2013

Contents

1 Structures and Metaphors.

> *Every sentence I utter must be understood not as an affirmation, but as a question.* Niels Bohr.[1]

Our ultimate aim is to develop *mathematical* means for describing/designing *learning systems* $\mathcal{L}$ that would be similar in their essential properties to the minds of humans and certain animals.

Learning and Structures. We want to understand the process(es) of learning, e.g. of *mother tongue* or of a *mathematical theory*, in the context of what we call *ergostructures*. Such structures, as we see them, are present in the depth of the human (some animal?) minds, in natural languages, in logical/combinatorial organizations of branches of mathematics and, in a less

[1] I could not find on the web when and where Bohr said/wrote it. Maybe he never did. But here and everywhere, a quote is not an appeal to an authority but an acknowledgment of something having been already said.

mature form, in biological systems – from regulatory networks of living cells up to, possibly, ecological networks.[2]

Learning from this perspective is

> a dynamical process of building the *internal ergostructure* of an $\mathcal{L}$ from the *raw structures* in the incoming flows $\mathcal{S}$ of signals, where $\mathcal{S}$ may or may not itself contain an ergostructure or its ingredients.

Such an $\mathcal{L}$ interacting with a flow of signals is similar to a a photosynthesizing plant growing in a stream of photons of light or to an amoeba navigating in a see of chemical nutrients and/or of smaller microbes: $\mathcal{L}$ recognizes and selects what is *interesting* for itself in such a flow and uses it for *building* its own structure.

(The use of metaphors in communicating scientific ideas is often confusing and misleading, especially if these ideas are not clearly set in their author's head. But until we develop a rough picture of what our "signals", "structures", etc. are and why one *can not* expect "rigorous definitions", we shall have to resort to metaphors. These, we hope, will provoke the reader's thoughts rather than create a deceptive illusion of understanding of something not properly explained.)

1.1 Universality, Freedom and Curiosity

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

Our inspiration for design of 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 electric/chemical signals that come into it:

This throbbing streaming crowd of electrified shifting points in the spongework of the brain [that] *bears no obvious semblance in space pattern...*

in the words of Charles Sherrington.

We conjecture that an infant's learning process follows an *universal* set of *simple* rules for exrtacting structural information from these *not* truly "chaotic" signals, where

these rules must *indiscriminately* apply to *diverse classes* of incoming signals.

Universality is the most essential property we require from our learning systems – this is the key that opens the door for a non-cosmetic use of mathematics; reciprocally, if successful, mathematics will furnish universality in learning.

At the moment, one may only speculate in favor of universality by appealing to "evolutionary thrift of Nature" and to "brain plasticity". Ultimately, we want to write down a *short* list of *universal* rules for "extracting" *mathematical structures* from *general* "flows of signals". And these flows may come in many different flavors – well organized and structured as mathematical deductions processes, or as unorderly as those depicted by Sherrington.

[2] Biological structures, historically/evolutionary as well as logically, are precursors of mathematical ones. For example, the probability of finding first 10^9 digit of $e = 2.718....$ "written" at some location u of an universe $\mathcal{U}$ increases by a factor $> 10^{100}$, if you find a bacterium kind machine feeding on a source of almost amorphous *free energy* at a point u' sufficiently close to u.

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

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.

Georg Cantor's words

The essence of mathematics lies in its freedom

equally apply to *learning* in place of *mathematics*. A universal learning system, as we understand it, must be designed as a self propelled learner that needs no purpose, no instruction, no reinforcement for learning.

This, a priori, is no more paradoxical than, say, your digestive system functioning with no teacher instruction or a mechanical system moving by *inertia* with no external forces. External constrains and forces change the behaviour of such systems, but they hardly can be regarded as the source of motion.

It would 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.

Cuiriosity as Intrinsic Motivation. The idea of what we call *ergosystems* is close to what was earlier proposed by Oudeyer, Kaplan and Hafner,[3] in the context of robotics under the name of *Intrinsically Motivated Curiosity Driven Robots.*

This "motivation" is implemented by a class of *predictor programs*, that depend on a parameter B which is coupled with (e.g. by being a function of) the behaviour of robots. These programs $Pred = Pred(H, B)$ "predict" in a certain specified way incoming signals on the basis of the history H, while the robots (are also programed to) optimize (in a specific formally defined way) the quality of this prediction by varying B. *Curiousity driven robots* are being designed and build in Ouduer's lab.

The Maximal Prediction idea is also central in our thinking on ergosystems; however, we emphasize "interesting" instead of "curious" and "structure" instead of "behaviour".

On (Im)Practicality of Universality. Multi-purpose gadgets are not among Greatest Engineering Achievements of the Twentieth Century: *flying submarines*, if they were a success, then only in James Bond films.[4] On the other hand, the

[3]See [OKH] – (Oudeyer, P., Kaplan, F., Hafner, V.V.: Intrinsic Motivation Systems for Autonomous Mental Development. IEEE Transactions on Evolutionary Computation 11:1, (2007) and [www.pyoudeyer.com].

[4]There are sea birds, e.g. *pelagic cormorants* and *common murres* who are (reasonably) good flyers and who also can dive, some up to more than 50(150?)m. The technology for

20th century machine computation has converged to universality; the basic machine learning will, most probably, follow this path in the 21st century.

1.2 Ego and Ergo.

Man is so complicated a machine that it is impossible
to get a clear idea of the machine beforehand,
and hence impossible to define it.
JULIEN OFFRAY DE LA METTRIE, MAN A MACHINE. 1748.

In our [SLE] paper[5] we collect evidence for the basic learning mechanisms in humans (and some animals) being *universal, logically simple and goal free.* An organized totality of these mechanisms is what we call *ergobrain* – the essential, albeit nearly invisible, "part" of human mind – an elaborate mental machine that serves as an interface between the neuro-physiological brain and the *(ego)*mind.

We bring this "invisible" into focus by rewriting the Cartesian

I THINK therefore **I** AM

as

cogito ERGO sum.

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

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

Egomind 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.* Most (all) of what we know of egomind is expressible in the common sense language – this language, call it *ego-reasoning,* that is a reflection of egomind, is perfectly adapted to to our every day life as well as to the needs of a practicing psychologist.

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.)

Certain aspects of ergo may be seen experimentally, e.g. by following *saccadic eye movements*, but a direct access to ergo-processes is limited.[6]

But there are properties of the working ergo in our brain/mind that are, however, apparent. For example, the *maximal number* $N_\circ$ of concepts our ergobrain can manipulate with *without structurally organizing them* equals three or

building universal/adaptable machines may be waiting ahead of us.

[5] *Structures, Learning and Ergosystems,* [www.ihes.fr/~gromov/PDF/ergobrain.pdf].

[6] This is similar with how it is with cellular/molecular structures and functions, where the "ergo of the cell", one might say, is the machinery controlled by the *housekeeping genes.*

four.[7] 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.

Apparently, 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.[8]

1.3 Ergo-Ideas and Common Sense.

Common sense is the collection of prejudices acquired by age eighteen.
EINSTEIN.

This saying by Einstein is not intentionally paradoxical. There is a long list of human conceptual advances based on *non-trivial* refutations of the common sense ideas. 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*.)

Your egomind with its *pragmatic ego-reasoning* – common sense as much as your emotional self, is a product of evolutionary selection. The two "selves" stay on guard of your survival and passing on your genes.

But *ergo*, unlike ego, was not specifically targeted by selection – it was adopted by evolution out of sheer logical necessity as, for example, the *1-dimensionality* of DNA molecules.

A pragmatically teleological ego-centered mode of thinking that was installed by evolution into our conscious mind along with the caldron of *high passions* seems to us intuitively natural and logically inescapable. But this mode was selected by Nature for[9] our social/sexual success and personal survival, not at all for a structural modeling of the world including the mind itself.

The self-gratifying ego-vocabulary of

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

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

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

The intuitive common sense concept of *human intellegence* – an idea insulated in the multilayered cocoon of *teleology* –purpose, function, usefulness,

[7] Some people claim their $N_\circ$ is as large as six and even seven but this seems unlikely from our mathematical perspective.

[8] Sayng "*Language is infinite*" is a metaphor that may be taken literally if you had no first hand experience with infinity.

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

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" but we welcome the idea that mathematics is the only valid alternative to common sense.

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 ego-mind, but they are *universal*, unlike say "useful" that depend on what, *specifically*, "useful" refers to. These ergo-ideas will direct us toward understanding of how a child's (ergo)brain, that hardly can be called serious, rational or objective, makes a world out of chaos of signals.

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

What we wrote on the above few pages can hardly convince anybody in the credibility of the idea of some invisible mathematical ergobrain running along with your pragmatic (ego)mind and in the existence of *performant* (universal goal free learning) ergosystems.

("Performant" is, in truth, no more applicable to an ergosystem than to a child at play: both do not follow your instructions and do not get engaged into solving your problems. From the ego perspective what ergo does, e.g. composing a most beautiful but utterly useless chess problem, appears plain stupid and meaningless. Reciprocatory, utilitarian ego's activity, e.g. laboriously filing tax returns, is dead boring for ergo.)

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

• mastering bipedal locomotion in a heterogeneous environment by human infants – walking, running and not bumping into things, as well as learning to speak, to read and to write, including learning languages and writing poetry by deaf-blind people;

• possession of almost supernatural *artistic or mathematical abilities* by exceptionally rare people, e.g. by the mathematician *Srinivasa Ramanujan*;[10]

[10] Writing off these rare events to "mere accidents", is like judging *explosions of supernovae* as "random and nothing else", just because only a dozen of supernovae were recorded in our galaxy with more than hundred billion stars (none since October 9, 1604);

• non-pragmatic *playful* nature of learning in animal (including human) infants during the periods of their lives when the responsibility for their survival resides in the paws of their parents;

• attraction to *useless* (from survival perspective) activities by human and by certain animals;

• creating and communicating *mathematics*, e.g. the potential ability by many (probably, by several hundred, if not thousands or even millions, people on Earth) to understand *Fermat's last theorem*[11] by reading a thousand pages long *written proof* of it.

The ergo-significance of these points is explained in our [SLE]-paper but a justification of the concept of ergobrain, is not, formally speaking, needed for a mathematical theory of ergosystems. However, it provides you with a psychological support in groping toward such a theory. And imagining how ergobrain thinks and trying to understand what ergo-logic can conceivably be, will not only help you in developing mathematics of ergosystems but also may guide you in a search for other mathematical structures.

Science is a child of the art of *not* understanding. If we want to approach the problem of *thinking machines* we must visualize the extent and the source of our non-understanding *thinking*. And the key question is not "can a machine think?" but if there is enough structural universality in the process of "thinking" in order to allow a mathematical modeling of this process. Formulating questions about "thinking" without making guesses on the nature of the underlying mathematical structure(s) is like talking about LAWS OF PHYSICS with no ideas of *number* and *space* in your head.

There is no visible *non-trivial* mathematical structure in what we conciously perceive in our (ego)mind[12] and it is unlikely that there is a *realistically describable* structure (mathematical model) of human (neuro)brain capable to account for such mental processes as learning a language, for example. But we conjecture that such structure(s) does reside in the *ergobrain*.

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

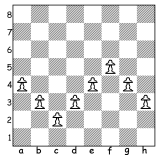
For example, the "obvious" properties of light and matter we see everywhere around us make sense *only* in the context of *quantum theory of electromagnetic fields*, while the source of the sunlight is revealed by the theory of *strong interactions* in *atomic nuclei*.

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

Nothing in Nature that is worth understanding can be explained in "a few simple words". If you happen to learn something novel for you in science without much intellectual effort and hard work – this is either not especially novel or it

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

[12] Taken in isolation certain structures *are* "trivial". Instances of these are highly disconnected (often bipartite) graphs, such as [object]–[name] graphs where the edges join words with the corresponding visual images or [question]–[answer] graphs of human dialogs – the brains of the stupidest animals depend on such graphs. Yet, mathematical derivations issuing from *several* trivial graphs make *non-trivial* structures in the human ergobrain as we shall see later on.



is not science.

1.4 Ergo Perspective on Chess

> *Arithmetical or algebraical calculations are, from their very nature, fixed and determinate....* [But] *no one move in chess necessarily follows upon any one other.*
>
> Edgar Allan Poe, "Maelzel's Chess-Player", April 1836.

In the early 19th century, when Poe was writing his article on *Maelzel's Chess-Player Automaton*, an ability to play chess was seen by many (all?) as a quintessential instance/measure of the intellectual power of the human mind. But the *mere existence* of *chess algorithms* is obvious.[13]

You play white. Let $\mathbf{eva}_0(P^*)$, where $* = \circ$ or $* = \bullet$, depending on whether a move by white($\circ$) or by black($\bullet$) is pending, be a "natural" numerical[14] *evaluation function* of a position P^*, e.g. the sum of judicially assigned weights to the pieces – positive weights to the white pieces and negative to the black ones.

Inspect possible *white* moves wh for all P°, denote by $P^\circ + wh$ the resulting new positions, and similarly consider all $P^\bullet + bl$. Define new evaluation function $\mathbf{eva}_1(P)$ by

$$\mathbf{eva}_1(P^\circ) = \max_{wh} \mathbf{eva}_0(P^\circ + wh)$$

$$\mathbf{eva}_1(P^\bullet) = \min_{bl} \mathbf{eva}_0(P^\bullet + bl).$$

Keep doing this and get

$$\mathbf{eva}_0 \Rightarrow \mathbf{eva}_1 \Rightarrow \mathbf{eva}_2 \Rightarrow \mathbf{eva}_3 \Rightarrow ... \Rightarrow \mathbf{eva}_N \Rightarrow ...$$

Stop, say, at $N = 20$ and let your computer (that plays white) maximize $\mathbf{eva}_{20}(P + wh)$ for all its moves wh. Such a program, probably, will beat anybody, but... no computer can inspect twenty half-moves in realistic time.

As recently as in 1950's, Hubert Dreyfus, a critic of artificial intelligence believed that a child would beat any chess program.

In 1957, Dreyfus was defeated by a $\mathbf{eva}_2$ chess program that was *implemented on a computer* by Alex Bernstein and his collaborators.[15]

[13] This, must(?) have been understood by Wolfgang von Kempelen, the creator of Chess-Player, and by contemporary mathematicians and scientists, e.g. by Benjamin Franklin who played with this "automaton". But I could not find a reference.

[14] This is *unnatural*, there is nothing intrinsically *numerical* in chess. Logically, what we need for an evaluation is (somewhat less than) an *order relation* on positions. But "ergo-evaluation" is more subtle and less logical.

[15] In 1945, the *first*(?) chess program was *written* by Konrad Zuse in his *Plankalkül* – a high-level programming language.

Fourty years later in 1997, Deep Blue (non-impressively) defeated the wold champion Kasparov, 3.5-2.5. The computer could evaluate 200 million positions per second; it inspected, depending on the complexity a position, from $N = 6$ to $N \approx 20$ half-moves. The program contained a list of endgames and it adjusted the evaluation function by analyzing thousands of master games.

So, when Poe insists that

> *... no analogy whatever between the operations of the Chess-Player, and those of the calculating machine of Mr. Babbage,*

one might judge him as mathematically naive; yet, Poe's conclusion was fully correct.

> *It is quite certain that the operations of the Automaton are regulated by mind, and by nothing else.*
> *... this matter is susceptible of a mathematical demonstration, a priori.*

The idea behind what Poe says is valid: Turing(Babbage) machines and **eva**-algorithms make poor chess players – they can match ergobrain programs *only* if granted superhuman resources of computational power.

This does not preclude, however, an approach with a quite different, possibly yet unknown to us, (ergo?)mathematics, but some people conjecture that the human (ego?)mind is "fundamentally non-algorithmic".

In his book *Shadows of the Mind*, Roger Penrose, who opposes the idea of *thinking machines,*[16] presents a chess position where

black has eight pawns, while white, in addition to eight pawns, has two rooks (and the white squared bishop, if you wish). The black pawns stay on black squares in an unbroken chain that separates the black king from the white pieces. White pawns are positioned in front of the black ones and fully shield them from the rooks (as in the above drawing).

Thus, the black king is safe in-so-far as black pawns do not change positions. But if a black pawn captures a white rook, then the chain of the black pawns will be disrupted and the black king *eventually* mated.

Any current computer-chess program would accept a sacrifice of a white rook, since "eventually" shows only in another twenty-thirty half-moves, while no human player will make such a silly mistake.

But Doron Zeilberger, who fights against the *Human Supremacy* idea, insists[17] that

> *symbol-crunching* [program], *valid for an $m \times n$ board rather than only an 8×8 board, will perform as good as a human player.*

Also, Zeilberger is critical of Penrose's use of *Gödel's incompleteness theorem* (see 2.1) as an argument against *thinking machines.*

Chess has been supplying an experimental playground to all kind of people pondering over the enigma of the human mind.

Logicians-philosophers marvel at how *formal rules* but not, say, the shape, color or texture of the pieces, determine what players do with them.

For example, Wittgenstein instructs (mocks?)[18] the reader:

[16] See more on www.calculemus.org/MathUniversalis/NS/10/01penrose.html

[17] www.math.rutgers.edu/~zeilberg/Opinion100.html.

[18] Wittgenstein is often quoted with *A serious and good philosophical work could be written consisting entirely of jokes.*

The meaning of calling a piece "the king"
is solely defined by its role in the game.

He continues –

imagine alien anthropologists landing on planet earth ...
discover ... a chess king...
[It] *remains an enigma to their understanding.*
Without ... other artifacts, the chess king is
only a chunk of wood (or plastic, or...).

(It is hard to resist continuing with *...or chocolate...* . But what the philosopher had in mind was not as trivial as it looks.)

Unlike logicians, students of the *ego*mind search for the meaning of chess in apparent or hidden urges of players to compete, to win, to grab, where making checkmate for a male chess player substitutes for killing his father in accordance with *Oedipus complex*.[19]

From the *human ergobrain perspective*, the relevance of chess is seen in its *intrinsic attractiveness* to certain people[20] and the central problem of chess in designing a (relatively) simple *learning (ergo)program* $\mathcal{L}$ that would find chess *interesting* and/or will be able (and "willing") to learn playing chess by itself whenever it is given an access to the records of sufficiently many (how many?) chess positions, chess problems and/or (fragments of) chess games.

And since chess, as most (all?) ergo-activities, is *interactive*, a learning will go faster if $\mathcal{L}$ is allowed an access to computer chess programs.

We conjecture the existence of such an $\mathcal{L}$, that is, moreover will be (rather) *universal* – not chess-special in a remotest way. It may come from somebody who has never heard of chess or of any other human game. However, such a program $\mathcal{L}$, being a pure ergo, may behave differently from a human player, e.g. it will not necessarily strive to win.

Such self-taught ergolearner program implemented on a modern computer will play chess better than any human or any existing specialized computer chess program, but this is *not* the main ergo issue. And it is nether the power of logical formality – something trivial from the ergo (and from general mathematical) point of view, what makes chess attractive and interesting.

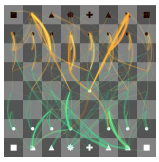
An ergolearner delights in the beauty of the *structure* of $\mathcal{CHESS}_{ergo}$, some kind of combinatorial arrangement of "all" interesting games and/or positions. An ergolearner tries to understand *(ergo)principles of chess* that transcend the formal rules of the game

These "principles" enable one, for instance, to distinguish positions arising in (interesting) games from *meaningless* positions, as it is seen in how chess masters memorize *meaningful* positions that come from actual games but they are as bad as all of us when it comes to random arrangements[21] of chess pieces on the board. In its own way, chess tells us us something interesting about *meaning*.

[19] Was it intended by Freud as a macabre joke? Sphinx might have accepted this solution to the riddle of chess, but we feel more at ease with *Flatulus complex* see [SLE] §6.7 [2].

[20] If you remold the piece "king" into "sphinx", the game will not loose its attractiveness to more than half of chess perceptive people.

[21] The number of possible chess positions is estimated around 10^{45}. Probably, 10^{12}-10^{18} among them are "meaningful".



1.5 Learning, Communicating, Understanding.

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

Zelig Harris.

Can (ergo)chess tell you something nontrivial about learning languages and understanding their meanings?

Following in the steps of Wittgenstein, one may approach a dialog in a natural language as a chess-like game that suggests an idea of (ergo) meaning: the *meaning* of an uttering U is derived similarly to that of the meaning of a position P in chess: the latter is determined by the combinatorial arrangement of P within the ergostructure $\mathcal{CHESS}_{ergo}$ of "all" ergo-interesting chess positions/games while the former is similarly determined by its location in the architecture of $\mathcal{TONGUE}_{ergo}$ of a language.

We fully adopt this idea in our ergo-context.

The meanings assigned by ergostructures (e.g. by our ergobrains) to signals are ***entirely*** *established by patterns of combinatorial arrangements and of statistical distributions of "units of signals", be they words, tunes, shapes or other kinds of "units".*
Understanding is a ***structurally organized*** *ensemble of these pattern in a human/animal ergobrain or in a more general ergosystem.*

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.

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 and that we shall address later on 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.

And the most interesting feature of natural languages – self-referentiality of their (ergo)syntax, that allows languages to *meaningfully* "speak" about themselves, has no counterpart in chess or in any other non-linguistic structure, e.g. in music.

On the other hand, self-referentiality is seen in mathematics; yet, only on its borders with a natural language, e.g. in *Gödel's incompleteness theorem*.

•• The internal combinatorics of $\mathcal{TONGUE}_{ergo}$ may be insufficient for the *full* reconstruction of the structure of the corresponding language.

For example, linguistic signals a child receives are normally accompanied by those coming via the visual and/or somatosensory system. 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.

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 may 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,[22] unless, besides ergo, you have a simple and adequate representation

[22]If your are blind to the symmetries of the chessboard, the number of possible moves by the white rook **R** in the presence of the white king, that you must learn (in $64 \cdot 63$ positions), is $> 64 \cdot 63 \cdot 13 > 50\ 000$. But if you have no ergo in your head, you'll need to be shown the

of the geometry of the chess board in your head. "Understanding" space with its symmetries, be this "understanding" preprogramed or acquired by *learning space*, is a necessary prerequisite not only for learning chess but also for communication/absorbtion of the rough idea of chess.[23]

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.

2 Mathematics and its Limits.

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

We are no gods and our minds are not pure ergo. To build a mathematical frame for "ergo" we need to recognize what of our mathematics is ready to serve as "parts" of ergosystems, what should be rejected and what needs to be be made anew.[24] And "ergo-criteria" for these "ergo-parts" are exactly those we use everywhere in mathematics:

NATURALITY, UNIVERSALITY, LOGICAL PURITY and CHILDISH SIMPLICITY.

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

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

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

But being mathematicians, we know that most general/universal theories are logically the simplest ones.[25] What is not simple is formulating/discovering such theories.

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

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

admissible moves of $\mathbf{R}$ in *all*($> 10^{45}$) possible chess positions.

[23]The geometry of the board can be reconstructed from a moderate list of sample chess games with *Poincaré's-Sturtivant space learning algorithms* (see §4 in [3]), but these algorithms are slow.

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

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

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

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

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

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

2.1 Logic and Rigor.

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

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

Admittedly, logicians participated in dusting dark corners in the foundations of mathematics but... most mathematicians have no ear for formal logic and for logical rigor.[26] We are suspicious of "intuitive mathematical truth" and we do not trust *meta*mathematical rigor of formal logic.[27]

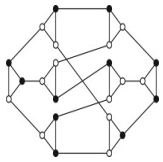
Cleanness of things does not make them beautiful in our eyes. Soundness of mathematics is certified by an ***unbelievably equilibrated*** *harmony* of its edifices rather than by the pedantry of the construction safety rules. Criticism of insufficient rigor in mathematics by George Berkeley (1734) as well the idea of "redemption" of Leibniz' calculus by Abraham Robinson (1966) strike us as

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

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

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

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



nothing but puny in the presence of the miraculous formula

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

for π = 3.14159265... being one half of the length of the unit circle. (Leibniz, 1682).[28]

We can not take seriously anything like $(a, b) := \{\{a\}, \{a, b\}\}$[29] but for some inexplicable reason this century old foundational dust finds its way to our textbooks under pretext of rigor as, e.g., in the following definition of a *graph* G as

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

We better keep clear of this "rigor".

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

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

Geometrically speaking,

the "body of mathematical truth" is DISCONNECTED.

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

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

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

Namely, let the vocabulary of an $\mathcal{MATH}$ include the following.

[28] 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.

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

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

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

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

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

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

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

the map R can not be onto:

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

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

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

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

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

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

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

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

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

This, applies, for instance, with several geometrically defined notions of genericity[30] for maps between Euclidean spaces where functions f may be continuous, smooth analytic or algebraic (and where genericity is accompanied by *transversality*).

On the other hand, explicitly described functions that one finds in "real life" (e.g. on Google) are more scarce than, say, natural numbers n, partly, because descriptive (less so graphical) presentation of "interesting" functions is

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

occupies more space than that for numbers. We shall see similar patterns in the hierarchical organization of our ergosystems.

(b) The childish simplicity of the proof of Gödel's theorem[31] does not undermine its significance. *Metamathematics* is similar in this respect to other non-mathematical sciences where a mathematical argument is judged not by its difficulty but by its applicability to "real life". Nontriviality of Gödel's theorem resides in a possibility of a meaningful metamathematical interpretation of the above "provably defined".

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

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

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

But anything that can be called "rigor" is lost exactly where the things become interesting and nontrivial – at the interface between mathematics and "logical reality". For instance, a variation of Gödel's theorem may tells you that there exists a mathematical proposition that can be written, say, on 10 pages but the proof of which will need between $10^{10^{10}}$ and $1000^{1000^{1000}}$ pages. This is perfectly acceptable *within* mathematics but becomes non-sensical if you try to apply it to mathematics "embedded into the real world". [32]

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

> if some people in a group of six kiss each other, then, necessarily,
> *either* there are *three* among them *all* kissing each other
> *or* there are *three* where *none* of the two kiss.

A child may *instantaneously* visualize a *graph* with green (kiss) and a yellow (no kiss) strings/sticks/*edges* between the pairs of these six people for *vertices*. (The child does not have to know graph theory.) Then the proof of the existence

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

A transparant categorical rendition of Gödel's theorem is presented in "Conceptual Mathematics" by Lawvere and Schanuel [7]. This was pointed out to me by Misha Gavrilovich who also explained how the above proof may be seen as an adaptation of their argument.

[32] Mathematical rigor and logical certainty are also absent from natural sciences. Einstein puts it in words:

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

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

of a *monochromatic triangle* will come after a few minutes (hours?) thought.[33]

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

2.2 Computations, Equations, Iterations, Simulations.

Formal languages do not walk on the streets occupying themselves with proving Gödel's style theorems one about another. But we humans *are* walking computers that are programmed, among other things, to guess and to imitate each other's mental computations.

"Computation" as it is used in the science of the brain and in science in general is a metaphor for elaborated, yet, structurally organized process. But there is no clarity with this notion.

Does, for instance, a planetary system perform a computation of, say, its total potential energy? You would hardly say so on the microsecond time scale but it may look as a "computation" if the time measured in million years.

In mathematics, there are several specific models of computation but there is no readymade language for describing all conceivable models.

Mind you, there is an accepted class $\mathcal{COMP}_{\mathbb{N}\to\mathbb{N}}$ (that parallels the class of provably defined functions) of what is called *computable* or *recursive functions*, $R(n)$ that send $\mathbb{N} \to \mathbb{N}$ for $\mathbb{N}$ being the set of *natural numbers* i.e. of positive integers $n = 1, 2, 3, 4, 5, ...$ Yet, there is no single distinguished natural description of this class as it is witnessed by the presence of *many* suggestions for its "best" description with the following five being most prominent.

Recursion + Inversion (Skolem, Gödel, Herbrand, Rózsa Péter),
λ-calculus (Church),
Turing machines and programs (Babbage, Ada Lovelace, Turing),
cellular automata (Ulam, von Neumann, Conway),
string rewriting systems (Markov).

These definitions of "computable" reflect their author's ideas on what is *"simple, useful, natural"* with the corresponding schemes of computation being quite different. None of them can be taken for the "normal" or "canonical" form of computation.[34] Besides, all these definitions of $\mathcal{COMP}$ are decades old and they have not undergone the post-Grothendieck *category theoretic* renovation.[35]

[33] A mathematically inclined child will soon generalize this to the full Ramsey theorem:
*if the subsets of an infinite set X are colored either in green or in yellow, then, for every $k = 1, 2, 3, ...,$, this X contains an infinite k-**monochromatic** subset $Y = Y(k) \subset X$, i.e. where all k-element subsets are of the same color.*
Graphs correspond to $k = 2$, while 6 and 3 are equated with ∞ in kindergartens.

[34] It is hard to argue for or against something being "natural". *Feets*, *meters* and *miles* may seem natural physical units of distance for some people. But, probably, there is *neither a truly canonical* normal form *nor a convincing mathematical* concept of equivalence applicable to different models of computation.

[35] Computations are not bound to $\mathbb{N}$ but I am not certain if there is a proper definition of "computable objects", e.g. sets with some structures representing suitable functors, in (yet, hypothetical) "computable categories".

But the traces of the following ideas, that underly the concept of computation, will be seen in our ergo-models.

• Compositions and Categories. *Composabilty* says that if a computation with the input from some (constructive set? class?) X_1 and output in X_2, denote it $C_1 : X_1 \rightsquigarrow X_2$, is followed by $C_2 : X_2 \rightsquigarrow X_3$, then the tautological composition $C_3 = C_2 \circ C_1 : X_1 \rightsquigarrow X_3$, where C_2 is performed right after C_1, is a computation again.[36] Thus, computations make what one calls *categories* provided *the identity* [non]-computations are in there as we shall always assume.

(Composability is a fundamental but non-specific feature of computations – almost everything you do in mathematics can be "composed" if you think about it.)

Moreover, many computation schemes operate with functions in *several variables* with arguments and/or values in a certain set X, that may be the set $X = \mathbb{N}$ of *natural numbers*, the set $X = \mathbb{Z}$ of *integers* and the set $X = \mathbb{R}$ of *real numbers.*, where "complicated" computable functions are successively built from "simple modules" by composing these "modules".[37] For instance, the following four functions:

two one variable functions: *the constant* $\mathbf{x} \mapsto \mathbf{1}$ and *the identity* $\mathbf{x} \to \mathbf{x}$
and two functions in two variables:
subtraction $(\mathbf{x}_1, \mathbf{x}_2) \mapsto \mathbf{x}_1 - \mathbf{x}_2$ and *multiplication* $(\mathbf{x}_1, \mathbf{x}_2) \mapsto \mathbf{x}_1 \cdot \mathbf{x}_2$

generate, in an obvious sense, (i.e. as an *operad* or as a *multicategory*) all *polynomials*

$$P(x_1, ..., x_k) = \sum_{d_1,...d_k \leq D} a_{d_1,,...d_k} x_1^{d_1} \cdot ... \cdot x_k^{d_k}.$$

with integer coefficients $a_{d_1,...,d_k}$ for all $k = 1, 2, ...$ and all degrees $D = 0, 1, 2, ...$

• Inversions. Inverting a function $y = P(x)$, that is finding x that satisfy the equation $P(x) = y$ for all y in the range of P, may be frustratingly difficult even for simple function $P : X \to Y$. An instance of this is computing (the integer part of) $\sqrt{y}$ for integer y that is much harder than taking $y = x^2$.

In general, the inverse map P^{-1} sends $x \in X$ *not to a single point* but to a (possibly empty) *subset* called $P^{-1}(x) \subset Y$, namely, to the set of those $y \in Y$ where $P(y) = x$. But a composition of such P^{-1} with some map $Q : Y \to Z$ may be a bona fide point map denoted $R = Q \circ P^{-1} : X \to Z$. This happens if $P^{-1}(x)$ is *non-empty* for all $x \in X$, i.e. P is *onto*, and if Q is constant on the subsets $P^{-1}(x)$ for all $x \in X$. In this case

R equals the *unique solution* of the equation $R \circ P = Q$.

[36] There is no consensus for writing $C_2 \circ C_1$ or $C_1 \circ C_2$. Although the *Zermelo Buridan's ass axiom* allows a choice of one of the two, remembering which one is impossible – how can one tell "←" from "→" in a symmetric Universe?

[37] The algebraic skeleta of sets of functions in several X-variables closed under compositions, also called *superpositions* in this context, go under the name "*operads*"; more generally, if the domains and ranges of maps are *not* assumed to coincide, than one speak of *multicategories*. These, similarly to ordinary categories, are described in terms of of classes of *diagrams of (multi)arrows* that mimic the obvious associativity-like properties of superpositions of functions.

The operad structures underly neural networks models of the brain. They will be also present in our ergosystems, where we shall insist on assigning specific structures to what goes under the heading "several" and/or "multi". (A newly born ergobrain does not know what the set $\{1, 2, ..., k\}$ is and it can not operate with functions presented as $f(x_1, x_2, ..., x_k)$).

Thus, extensions of classes of maps by adding such inverses may be described in category theoretic terms as follows. Let $\mathcal{P}$ be a subcategory of a category $\mathcal{S}$, e.g a class of maps p between sets that is closed under composition.

The invertive extension $\mathcal{R}$ of $\mathcal{P}$ in $\mathcal{S}$ is, by definition, obtained by adding to $\mathcal{P}$ the solutions $R \in \mathcal{S}$ of the equations $R \circ P = Q$ for all P and Q in $\mathcal{P}$ whenever such a solution exists and is unique.

(This $\mathcal{R}$ may be *non-closed* under composition of morphisms and it can be *enlarged further* by generating a subcategory in $\mathcal{S}$ out of it.)

Such an extension may be incomparably greater than $\mathcal{P}$ itself, where the BASIC EXAMPLE of this is as follows.

Let $\mathcal{S}$ be the category (*semigroup* in this case) of functions $\mathbb{N} \to \mathbb{N}$ for $\mathbb{N} = \{1, 2, 3, 4, 5, ...\}$ and let $\mathcal{P}$ consist of *primitively recursive* functions.

Then the invertive extension $\mathcal{R} \subset \mathcal{S}$ of $\mathcal{P}$ equals the set of all recursive (i.e. computable) functions $\mathbb{N} \to \mathbb{N}$.

"Primitively recursive" is a currently accepted formalization of "given by an explicit formula". Such formalizations and the resulting $\mathcal{P}$ may be somewhat different but the corresponding $\mathcal{R}$ are all the same.

A convincing instance of this is

DPRM THEOREM[38] *The invertive extension $\mathcal{R}$ of the subcategory $\mathcal{P}$ of polynomial maps $\mathbb{N}^k \to \mathbb{N}^l$, $k, l = 1, 2, 3, 4, 5, ...$, in the category $\mathcal{S}$ of all maps equals the category of recursive (i.e. computable) maps $\mathbb{N}^k \to \mathbb{N}^l$.*

In other words,

every computable function $R : \mathbb{N} \to \mathbb{N}$, can be decomposed as $R = Q \circ P^{-1}$, where

- *$P, Q : \mathbb{N}^k \to \mathbb{N}$ are integer polynomials,*
- *the map $P : \mathbb{N}^k \to \mathbb{N}$ is onto,*
- *Q is constant on the subsets $P^{-1}(n) \subset \mathbb{N}^k$ for all $n \in \mathbb{N}$.*

(Moreover, there is a universal bound on k, e.g. $k = 20$ suffices.)

The way the theorem is proven allows an explicit construction of polynomials P and Q, e.g. in terms of a Turing machine (we define later on) that presents an R. For instance, these P and Q can be actually written down for the *nth prime number function* $n \mapsto p_n$.

However, this theorem does not and can not shed any light on the structure of prime numbers[39]. All it shows is that Diophantine equations, that make a tiny fragment of the world of mathematics, have, however, a capability of "reflecting" all of $\mathcal{MATH}$ within itself: any given (properly formalized) mathematical problem Π can be translated to the solvability problem for such an equation. This, in conjunction with Gödel's theorem, tells you that

solvability of general equations $P(x_1, x_2, ..., x_k) = n$ is an intractable problem.

There is a host of similar theorems for all kinds of (not at all Diophantine) "simple equations" that make mathematics, seen from a certain angle, look like a fractal composed of infinitely many "Gödel's fragments" where each "fragment" multiply reflects $\mathcal{MATH}$ as a curved fractal mirror with every reflected image of

[38] Conjectured by Martin Davis (Emil Post?) in 1940's and finalized by Matiyasevich in 1970 following Davis, Putnam and Robinson.

[39] Probably, *nothing what-so-ever* about prime numbers can be seen by looking at such P and Q, not even that there are infinitely many of primes.

$\mathcal{MATH}$ being transfigured by a chosen translation of $\mathcal{MATH}$ to the language of this "fragment".

A translation of a "general difficult problem" Π to a "concrete and simple" equation whenever such a translation is available by a DPRM kind of theorem, does not help solving Π but rather shows that an apparent simplicity of the corresponding class of "equations" is illusory.[40] These translations bring forward Gödel's theorem and guard you from entering blind alleys of naive solvability problems. But even without Gödel, anything as easy to formulate as the solvability problem makes one wary, be these Diophantine or other kinds of equations.

The DPRM theorem itself was a response to David Hilbert who suggested in his 10th problem:

> *to devise a process according to which it can be determined in a finite number of operations whether the equation is solvable in rational integers.*[41]

This theorem indicates that Hilbert's suggestion taken literally was unsound and, if followed, must be coupled with a search for *particular classes* of equations where integer solutions are well structurally organized.

But what Hilbert could not fathom, is that the "true Diophantine beauty", as we see it to-day, resides not in integer solutions of $P(x_1, ..., x_k) = 0$, but in *non-Abelian higher dimensional "reciprocity laws"* associated to integer polynomials P. Roughly, such laws can be seen as *analytic relations* between infinitely many numbers $N_p(P)$ for *all prime* $p = 2, 3, 5, 7, 11, 13, 17, ...$, where $N_p(P)$ equals the number of solutions of the congruence $P = 0 \mod p$.

Such relations are expected[42] to generalize *Riemann's functional equation*

$$\frac{\zeta(1-s)}{\zeta(s)} = \frac{\alpha(s)}{\alpha(1-s)},$$

where

$$\zeta(s) = \prod_p \frac{1}{1-p^{-s}} \quad \text{and } \alpha(s) = \frac{1}{2}\pi^{-s/2}\int_0^\infty e^{-t}t^{\frac{s}{2}-1}dt \quad \text{for } s > 1,$$

where both functions, ζ (that harbors the deepest mysteries of prime numbers) and α (an apparently insignificant child of simple minded analysis) admit meromorphic extensions to all complex s-plane and where the thus defined functional equation, applied at different s, encompasses infinitely many relations between the prime numbers $p = N_p(P)$ for $P = P(x_1, x_2) = x_1 - x_2$.

---

[40] For instance, the solvability problem for a Diophantine equation $P(x_1, ..., x_k) = 0$ transforms by a particular translation algorithm ALG_{part} built into a given proof of DPRM to the solvability problem for an equation $P_{new}(x_1, ..., x_l) = 0$ with the (integer) polynomial P_{new} being by far more complicated (i.e. with larger coefficients) than the original P, where it is virtually impossible to reconstruct P back from P_{new} even if you know ALG_{part}.

[41] The idea of a possible effective resolution of all Diophantine problems was in line with Hilbert's (pre-Gödel) optimistic:

Wir müssen wissen – wir werden wissen! (We must know – we will know!)

This position is also articulated his 2nd problem:

a direct method is needed for the proof of the compatibility of the arithmetical axioms.

[42] There is an incredible tapestry of conjectures and partial results toward this end, known as the *Langlands program*, that is a far reaching generalization of Hilbert's 9th problem on *the most general law of reciprocity in any number field* as well as of his 12th problem on description of Abelian extensions of such fields.

• Computations by Networks. "Complicated computations" can be realized by networks of simple (or non-simple) "computational steps" as it is done in modern computers and, probably, in our brains.

General purpose programmable computers were designed by Charles Babbage in the mid-1800's based on the principles similar to those formulated by Turing in 1936.

Turing suggested a presentation of a general computation by moves of a "simple minded bug" who crawls along a set S (usually, assumed *infinite*) of *spacial locations* or *sites* s and who is able to move from an s to the sites s' adjacent to s.

(Unit segments $[n, n+1]$, $n = 0, 1, 2, 3, 4, ...$, on the line or the squares on an infinite sheet S of square paper are instantces of such s.)

Upon arriving at a site s, the bug may mark it by a scent σ from a (usually, assumed *finite*) scent collection $\{\sigma\}$ available to the bug or to change/erase already present scent $\sigma = \sigma(s)$.

Bug's actions at a given moment depend upon bug's "moods" b and, from our perspective, the bug is nothing but a "bag of moods" denoted B (usually, assumed to have *finitely* many b in it). On the other hand, a state x of the full computational system run by B is given by

(A) A scent function $\sigma(s)$ on S. (This, typically, carries the bulk of information about x.)

(B) The mood $b \in B$ of the bug.

(C) Location $s_\bullet \in S$ of the bug.

The behaviour of the bug, however, depends *only* on the *scent* $\sigma_\bullet = \sigma(s_\bullet)$ at its current location (*not* on the location itself) and on its own *mood* b. Depending on these data, the bug

(A′) modifies, erases or keep unchanged the scent at its present location by

$$\sigma_\bullet \mapsto \sigma'_\bullet = \sigma'_\bullet(\sigma_\bullet, b);$$

(B′) changes (possibly, does not change) its mood by

$$b \mapsto b' = b'(b, \sigma_\bullet),$$

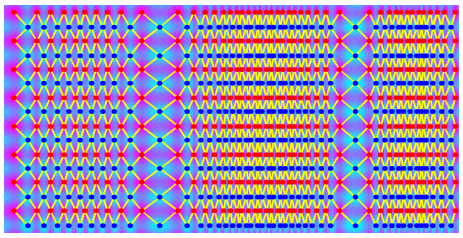
(C′) moves (if it does at all) to an adjacent location $s'_\bullet$ where the "direction" of the move $s_\bullet \mapsto s'_\bullet$, e.g. up, down, right or left on the squared paper, depends on b.

Those actions, are performed according to specific rules (e.g. (B′) is encoded by the above B-valued function $b'(b, \sigma_\bullet)$) that were installed by the designer into the bug.

From the computer scientist point of view our bug is uncomfortably abstract and general for turning it into a computer program. In fact, the original "Turing bug" was crawling on the set $S = \mathbb{N}$ of natural numbers with only one scent (present or absent) at its disposal.

But we want to find a proper place for such a bug in a *maximally general*, hence *simplest possible*,[43] mathematical environment that would accommodate

[43]This is not how engineers, architects and computer scientists understand "simple" – they are keen at *specificity* rather than on generality. An architect would find it absurd to start designing a square house with developing an abstract theory of geometric symmetries, while a mathematician will be horrified by a detailed image of a house presented on a pixel by pixel basis. But we all dislike definitions with "*...7-tuples* $\langle Q, \Gamma, b, \Sigma, ...\rangle$*...*" in them.



"rich computer fauna" and where a particular "bug life story" could be unraveled with any desirable degree of precision whenever a need arises.

The choice of the following definitions, that are not the most general ones, is partly motivated by their ergo-variants we shall meet later on.

Sites and States. Let S be a set where its elements s are called *sites* and let Ξ_s, $s \in S$, be sets (often assumed finite) the elements of which are called *states* of s and/or *local states of S at s*.

Such an s can be an atom or a molecule with $\chi \in \Xi_s$ characterizing its "physical states", e.g. *colors* (of light it may emit); alternatively, an s may be an "empty location" where one writes letters χ from the "alphabet" Ξ_s.

States $x = x(s)$ on S are defined as collective states of all s, that are functions $x : s \mapsto x(s) \in \Xi_s$. (States "written by letters" χ on S may be called "signals".)

In a presence of sufficient symmetry, the sites s may be "identical", i.e. all Ξ_s are identified (by means of a symmetry group) with a single set Ξ. In this case, states are just functions $x : S \to \Xi$, sometimes called Ξ-*states* on S.

Transformation Rules. Given two sets of sites, (S, Ξ_s) and (T, Ω_t), a *transformation rule*,[44] that *directs* states y on T to states x on S, depicted by

$$S \overset{\curvearrowleft}{\rightrightarrows} T,$$

is defined by the two kinds of data.

• *A directed bipartite graph from S to T*, that is a multi-valued map from S to T, denoted as a *morphism* $G : S \rightrightarrows T$. This G is described in terms of (often assumed finite) sets I_s of *edge-arrows* issuing from s, for all $s \in S$, and terminating at some $t = t_i \in T$, $i \in I_s$, where these t are called *adjacent* to s.

A significant special case of this is where all I_s are identified with a *single* I; in this case, G is called I-*colored*. Such a G equals the union of the graphs of arrows of some maps $g_i : S \to T$, $i \in I$.

• Ξ_s-Valued functions $\chi = f_s(\omega_i)$ for all $s \in S$, where $\omega_i \in \Omega_{t_i}$, $i \in I_s$; these may be also written as maps from Cartesian products of Ω_t to Ξ_s,

$$f_s : \times_{i \in I(s)} \Omega_{t_i} \to \Xi_s \text{ for all } s \in S.$$

When G and f_s are specified, we include them in to the above diagram as

$$S \underset{G}{\overset{\{f_s\}}{\rightrightarrows}} T.$$

Construction of the transformation (map) $F : X(T) \to X(S)$

[44] Also called "single layer neural network".

from the space $X(T) = X_\Omega(T)$ of states $y : t \mapsto y(t) \in \Omega_t$ on T
to the space $X(S) = X_\Xi(S)$ of states $x : s \mapsto x(s) \in \Xi_s$ on S,

according to given rules $(G, \{f_s\})$ is straightforward and natural:

the value $\chi \in \Xi_s$ taken by the state $x = F(y)$ at s depends *only* on the values $\omega_i \in \Omega_{t_i}$ of $y = y(t)$ at the sites $t = t_{i_{|i \in I_s}}$ in T (that are *adjacent* to s via the edges in G), where, by definition, this χ equals $f(\omega_i)$.

Thus, $x = F(y)$ is defined by

$$F(y)(s) = f_s(y(t_i))_{i \in I_s} \text{ for all } s \in S.$$

We call the so defined transformation F *ruled* or *directed by* $(G, \{f_s\})$.

Equable Rules. Let us distinguish the case where
G is I-colored,
all Ξ_s are identified with a single Ξ,
all Ω_t are identified with a single Ω,
all f_s are equal to a single Ξ-valued function f in the variables $\omega_i \in \Omega$, $i \in I$,

$$f : \Omega^I = \underbrace{\Omega \times \ldots \times \Omega}_{I} \to \Xi.$$

Under these circumstances, $X(S) = X_\Xi(S)$ equals the space of functions $S \to \Xi$ and $X(T) = X_\Omega(T)$ is the space of functions $T \to \Omega$, where the corresponding transformation F directed/ruled by G and f is denoted $F = F_{G,f} : X(T) \to X(S)$.

However simple, the construction $(G, \{f_s\}) \rightsquigarrow F$ *functorially*[45] transform "the category of rules", that is well represented by finitary combinatorial objects (especially in the equable case), to a "rather transcendental" category of maps between function spaces.

In particular, if the graph G "maps" S into *itself*, $G : S \rightrightarrows S$, and accordingly F transforms $X(S)$ back to $X(S)$, then this "transcendentality" shows up in the dynamics of the iterates of the map $F : X(S) \to X(S)$.

("Turing's bugs" are described in terms of (G, Ξ, f_s) as follows. Recall that a bug is a bag B of smells b with a set Σ of scents σ at its disposal and define Ξ to be the *disjoint union of* Σ *with the Cartesian product* $\Sigma \times B$,

$$\Xi = \Sigma \sqcup (\Sigma \times B),$$

where $x(s) = (\sigma, b) \in \Sigma \times B$ signifies that
$\bullet_0$: the bug is *located at this* s,
$\bullet_1$: the bug is in the mood b,
$\bullet_2$: the site s smells of σ,
while $x(s) = \sigma \in \Sigma$ tells you that
$\circ_0$: there is *no bug at* s
$\circ_1$: s smells of σ.

Now, the above (A′)-(B′)-(C′) rules for bug's moves (obviously) specify the sets I_s of the adjacency edge-arrows in the graph $G : S \rightrightarrows S$ as well as the

[45] There are several layers of this functoriality. For instance, our categories are naturally acted upon by covering maps between graphs $\tilde{G} \to G$.

Besides being functorial, this construction is "compatible with computability", where a precise general formulations of this is somewhat longer than the proof.

functions f_s, such that the corresponding transformation $F : X(S) \to X(S)$ describes the behavior of the bug.

On Being "Buggy". The characteristic feature of "buggy" transformation $F : X(S) \to X(S)$ is that the states x and $y = F(x)$ on S differ (if at all) *only* at the two end vertices (sites) of a single edge of the graph G:

$x = x(s)$ and $y = y(s) = F(x)(s)$ are equal everywhere on S except, possibly, for two adjacent sites $s_\bullet$ and $s'_\bullet$ in S.

Informally, the "difference" $x - F(x)$ is supported on a single edge in G.)

Eventualization. The transformation $F_{G,f} : X \to X$ directed by (G, f), say, in the equable case, may look no more complicated than the underlying rule-map $f : \Xi^I \to \Xi$, especially if F of a humble buggy origin. But since F sends X into *itself* it may be iterated, where the resulting dynamical picture of

$$F^{\circ N} = \underbrace{F \circ F \circ \ldots \circ F}_{N} : X \to X$$

may be unexpectedly rich (messy?) for $N \to \infty$[46] and the computational power arising from the asymptotic of $F^{\circ N}$ becomes awesomely powerful and (pointlessly?) complicated.

If F implements a computation in a *combinatorial* (as opposed to an analytic or geometric) category, then convergence of $F^{\circ N}$ to, say, $F^{\circ\infty}$ signifies *eventual stabilization*:

$$F^{\circ N}(x) = F^{\circ N_0}(x) \text{ for some } N_0 \text{ and all } N \geq N_0.$$

If this happens, we write $F^{\circ\infty}(x)$ for $F^{\circ N}(x)_{|N \geq N_0}$.

($F^{\circ\infty}$, whenever it exists, can be described in purely algebraic – semigroup theoretic, terms as being F-*bivariant*, i.e. such that

$$F \circ F^{\circ\infty} = F^{\circ\infty} \circ F = F^{\circ\infty}$$

and such that all other F-bivariant Φ are, necessarily, $F^{\circ\infty}$-bivariant as well.)

Enlarging a class $\mathcal{F}$ of transformations F acting an X, or, often, of these F restricted to an F-invariant *subset* $U \subset X$, by adjoining eventualizations of those $F \in \mathcal{F}$ for which $F^{\circ\infty}$ is defined (i.e. exists), say on a given $U \subset X$, may tremendously enrich/mess up this $\mathcal{F}$, similarly to how it goes with adding inverse maps.[47] For example, one has the following

Turing Theorem. Let G be the (multi)graph on the vertex set $\mathbb{N} = \{1, 2, 3, 4, \ldots\}$ with the arrow-edges $m \longrightarrow n$ for $|m - n| \leq 1$, where this G is naturally colored by the three maps $n \mapsto n + 1$, $n \mapsto n$ and $n \mapsto n - 1$ (except for the non-existing $1 \longrightarrow 0$ edge.)

Then every computable (i.e. recursive) function $R : \mathbb{N} \to \mathbb{N}$ can be implemented by a "buggy" $F^{\circ\infty}_{G,f}$ for some finite set $\Xi = \Xi_R$ and a map $f = f_R$:

[46]The complexity arising from iteration is seen in the images of the Mandelbrot set and similar limit sets.

[47]The reason for $F^{\circ\infty}$ bringing "mind-boggling mess" is a *virtual impossibility* of deciding, in general, when and where $F^{\circ\infty}$ is defined, i.e. for which F and x the iterate $F^{\circ\infty}$ becomes independent of N for large N; this "impossibility" is called *Turing's Halting Theorem.* This theorem may be regarded as being "obviously equivalent" to Gödel's Incompleteness Theorem, but a mathematically acceptable formulation (if this is possible at all) of what this "equivalent" means will be more elaborate than the proofs of these theorems.

$\Xi \times \Xi \times \Xi \to \Xi$, *where "buggy" signifies that the original transformation of states* $x = X(n)$ *on* $\mathbb{N}$, *that is* $x \mapsto y = F(x)$, *satisfy* $y(n) = x(n)$ *for all* $n \in \mathbb{N}$ *except, possibly, for two consecutive numbers, say* n_F *and* $n_F + 1$.

Serious(?) *Remark.* The proof of this is no more exciting than rewriting a computer program from one language to another but there is something disturbing with the formulation of the theorem at an apparently trivial point.

Namely, in order to make sense of "implemented", one needs a way to represent numbers $n \in \mathbb{N}$ by states $x(s) \in X(S)$ for $S = \mathbb{N}$, and then go back from X to $\mathbb{N}$.

It is "natural" for us to do this with decimal or binary expansions but, being used to the positional representation of numbers, we do not realize how much it distorts the "true nature of numbers".[48]

But a more worrisome point is *an extension* of decimals written on S to a state $x = x(s)$ defined for *all* $s \in S$, such that the eventualization $F^{\circ\infty}(x)$ *is* defined.

Of course this is easy for any particular class of "Turing bugs" (or other computation networks) but it seems difficult to say what we want of such an encoding+extension in a *mathematically acceptable* manner, that is in *most general* and, simultaneously, *maximally simple* terms (such that the above "can" and "implement" become acceptable to a newly born ergobrain).

The encoding+extension problem can be also formulated in terms of

Programs and Simulations. What does it means, *mathematically* speaking *"programming* a computational system for *simulation* of a computation by another (including itself) system"?

There are lots of specific examples of mutual "simulations/implementation" of computational processes but it seem there is no simple general *definition* of this. Should, for instance, "writing a program" be defined as a computation of a particular kind?

Cellular Automata of Ulam-von Neumann, of Conwey and of Langton. Computation processes $F = F_{G,f}$ in a locally finite spatially symmetric combinatorial environments G are called now-a-days *cellular automata.*

Ever since Turing, these were used for showing how a seemingly simple transformation turns upon iterations into a monster of complexity with nothing of mathematically/structurally beauty remaining there .

For instance, Von-Neumann has implemented *universal self replication*[49] (a counterpart of *Turing's universal computer*) on the ($2D$ lattice) group

$$S = \mathbb{Z}^2 = \mathbb{Z} \times \mathbb{Z} = \{n_1, n_2\}_{n_1, n_2 = \ldots -2. -1, 0, 1, 2, \ldots},$$

where the graph G is defined by the identity map $id : S \to S$ (i.e. $s \longrightarrow s$ that are loops at all s) along with *four* unit translations $g_i : S \to S$ that are

$$(n_1, n_2) \mapsto (n_1 \pm 1, n_2) \text{ and } (n_1, n_2) \mapsto (n_1, n_2 \pm 1)$$

and where the "alphabet" Ξ of local states used by von Neumann had 29 letters in it.

[48] Positional representation was not known to Greek mathematicians, not even to Archimedes, who came close to it in his The Sand Reckoner.

[49] *Formulating* this as a *mathematical* theorem may be as difficult as giving a *mathematical* definition of a "mutual simulation of computations".

Another Game of Life. In 1970, John Conway found a beauty among this kind of monsters, called *Conway's Game of Life*, with an amazing balance between "chaotic" and "regular" behaviour resembling the real GAME OF LIFE ON EARTH.

Conway's Game of Life $F : X \to X$ operates on on the states x on $S = \mathbb{Z}^2$ with the graph G defined by the identity map $id : S \to S$ and the following *eight* g_i (depicted as red squares in the above figure),

$$(n_1, n_2) \mapsto (n_1 \pm 1, n_2),\ (n_1, n_2) \mapsto (n_1, n_2 \pm 1) \text{ and } (n_1, n_2) \mapsto (n_1 \pm 1, n_2 \pm 1)$$

and where the sites $s \in S = \mathbb{Z}^2$ may be only in two states, call them $[\mathbf{live}]$ and $[dead]$.

If you believe it is easy to find an "interesting cellular game" of this kind by brute force computer search and/or by trial and error, think of how long it takes to single out any "interesting" binary function in nine variables,

$$f : \{[\mathbf{live}], [dead])\}^9 \to \{[\mathbf{live}], [dead])\},$$

out of $2^{2^9} = 2^{512} > 10^{150}$ (!) possibilities.

Apparently, human (ergo)brain is able to make such a choice by blinding its eyes to the enormity of the problem. Thus, closing his eyes, Conway takes the (ruling) function $f = f(\chi_i)$, $\chi_i \in \{[\mathbf{live}], [dead]\}$, that depends

on the $[\mathbf{live}]/[dead])$ state of the variable χ_i for the i of $g_i = id$ that corresponds to the central blue square in the figure,

and

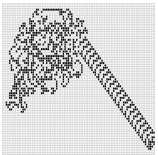
only on the *number* $j \in \{0, 1, 2, ..., 8\}$ (but not on the positions) of *live* cells around the "central" s;

This reduces the number of choices to $2 \cdot 2^9 \approx 1000$; from this point on, there is, probably, only one conceivably "interesting" possibility. Conway arrives at it by further reducing the number of candidates for the two f by selecting from those $f = f(*, j)$, (here $*$ is for $[\mathbf{live}]$ or $[dead]$ at the "central" s) where the sets of "pro-life" j, that are the pullbacks $f^{-1}([\mathbf{live}]) \subset \{1, 2, ..., 8\}$, have no gaps in them; this reduces the number of choices to <100. Finally, Conway defines his game by two rules:

- ∘ there is a single j, namely $j = 3$, for which $f([dead], j) = [\mathbf{live}]$; otherwise, the "central" state remains "dead",
- • there are two neighbouring j, namely, $j = 2, 3$, for which $f([\mathbf{live}], j) = [\mathbf{live}]$; otherwise, the "central" state "dies".

One knows that any computation can be "simulated" by the Game of Life,[50] but the (mathematical?) beauty resides elsewhere. Where exactly is hard to say, probably, because, as it is for the true GAME OF LIFE, we do not know what questions to ask to receive simple beautiful answers.

[50] This would, probably, imply von Neumann's "universal replication theorem" if the latter were formulated as a theorem.



Langton Ant. This "ant", probably, mathematically, the most inviting cellular automaton. It crawls on $S = \mathbb{Z}^2$, where the edges of the relevant graph G (with the vertices/sites $s \in S$) are arrows that may be directed "up", "down", "left" and "right" (that correspond to the four moves $(n_1, n_2) \mapsto (n_1 \pm 1, n_2)$ and $(n_1, n_2) \mapsto (n_1, n_2 \pm 1)$) and where the sites $s \in S$ are colored either in [**black**] or in [white] similarly to how it is in the Game of Life..

When the ant enters a site s, it

(a) switches the color of s

and then

(b) moves to an adjacent s' by an arrow perpendicular to the one it entered s; thus, turning either clockwise 90° or counterclockwise, where the choice between the two depends on the white or black color of s.

The path taken of by the ant in S may be amazingly complicated even for initially constant coloring of S. But, conjecturally, if the initial coloring function $color(s)$ is constant at infinity – [**black**] or [white], then

the position $s(N)$ of the ant at the time $N = 1, 2, 3, ...$ eventually becomes periodic: there exist an N_0 and M, such that $s(N + M) - s(N) \in S = \mathbb{Z}^2$ does not depend on N for $N \geq N_0$.

Thus, the sites visited by the ant, after some moment on that are $s(N_0), s(N_0 + 1), s(N_0 + 2), ...$, lie in half band (called *ant highway* that is seen in the above figure) pinched between two parallel lines in the plane $\mathbb{R}^2$ that contains $S = \mathbb{Z}^2$.

Even unproven, this is mathematics.[51]

2.3 Numbers and Symmetries.

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

James Clerk Maxwel.

We are so used to the idea of *number* that we forget how *incredible* properties of *real numbers* are. The seamless agreement of several different structures –

[51]This ant, as any other Turing bug, can be modeled/simulated by Conway's game of life, but no(?) such model has the resources to make its representation the ant to crawl on the plane in a geometric fashion similar to the original one. This deficiency is shared by most universal automata: the apparent (but not the only) reason for this is that " general modeling" ignores the time factor implicit in the number N for which the stabilization of the N-iterates $F^{\circ N}$ takes place. (Imagine, for instance, that the output signals represent consecutive digits of π= 3.141 592 653 589... with the time interval between the signals being proportional to the digits of e = 2.718 281 828 459... . What is being modeled in this case?)

Yet, The REAL GAME OF LIFE does have this ability as it is witnessed by the *images* of computer simulations designed by human players of this GAME.

Is there something *mathematically discernible* in the REAL GAME that is absent from Conway's game?

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...*– issue from these properties. Mathematics and theoretical physics are the two facets of these symmetries that are both expressed in the essentially same mathematical language.

As Poincare says,

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

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

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

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

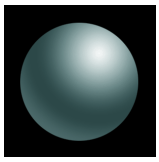
In contrast, doubling is seen everywhere in Life. All of us, most likely, descend from a polynucleotide molecule which had successfully doubled about four billion years ago. Organisms grow and propagate by doubling of cells. Evolution is driven by doublings of genomes and of significant segments of the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

A true numerical addition may be rarely (ever?) seen in *biology proper* but, for example, additivity of electric charges in neurons is essential in the function of the brain. This underlies most mathematical models of the neurobrain, even the crudest ones such as neural networks. But the ergobrain has little to do with additivity and linearity.[52]

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.

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



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

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

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

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

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

A plausible (ergo)brain's strategy for *learning space*, in particular, for reconstruction of spacial symmetries from the retinal images of moving objects, was suggested by Poincaré in §IV of *La science et l'hypothèse,* where Poincaré indicates what kind of mathematics may be involved in *learning space by our visual system.* An aspect of our "ergo-approach" is an attempt to spell out what Poincaré might have in mind.[53]

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

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

[53] A similar idea can be seen in Sturtevant's 1913 construction of the first *genetic map* as we explain in $4 of [3].

2.4 Languages, Probabilities and the Problems with Definitions.

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

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

Human languages carry imprints of the mathematical structure(s) of the ergobrain and, at the same time, learning a natural (and also a mathematical[54]) 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.

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

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

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

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

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

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

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

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

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.[55] And Titus Lucretius (99 –55 BCE), a follower of Democritus, describes in his poem *De Rerum Natura* what is now called *Einstein-Smoluchowski stocahstic model* of Brownian motion[56]

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

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

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

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

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

But randomness also *breaks* symmetries: the only way for a donkey' ergobrain (and ours as well) to solve Bouridan's ass problem is to go random.[57] Emanation of the "miraculous decision power of random" intoxicates a gambler's ergo.[58]

The first(?) documented instance of the *calculus* of probabilities – "*measuring chance*" by a European[59] 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*[60] in Luca Paccioli's *Summa de Arithmetica, Geometria, Proportioni et Proportionalita*.[61]

Paccoli's solution was criticized/analized by Cardano[62] in *Practica arith-*

[55]Our sketchy outline of the history of probability relies on [9] [1], [10], [6], [5], [11] with additional *References for Chronology of Probabilists and Statisticians* on Ming-Ying Leung's page, http://www.math.utep.edu/Faculty/mleung/mylprisem.htm

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

[57]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.

[58]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 group $\mathbb{Z}_2^{\mathbb{N}}$ and by automorphisms of this group.

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

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

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

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

metice et mensurandi singularis of 1539 and later on by Tartaglia in *Trattato generale di numerie misure*, 1556.

The first(?) systematic mathematical treatment of statistic in gambling appears in *Liber de Ludo Aleae* of Cardano (where he also discusses the psychology of gambling) 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

$$9 = 1 + 2 + 6 = 1 + 3 + 5 = 1 + 4 + 4 = 2 + 2 + 5 = 2 + 3 + 4 = 3 + 3 + 3$$

and

$$10 = 1 + 3 + 6 = 1 + 4 + 5 = 2 + 2 + 6 = 2 + 3 + 5 = 2 + 4 + 4 = 3 + 3 + 4$$

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

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

What is the probability of having two girls in a family with two children where one of the them is known to be a girl?[63])

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*[64] stands among the ten (±2) greatest mathematical theorems of all time. To appreciate its power look at the following example relevant to some (ergo)-learning algorithms.

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

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

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

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

then,

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

---

and music.

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

[64] 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.

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

In particular,

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

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

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

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

But exhibiting any single one of them, say for $M = 1000$ and $N = 10^{12}$ seems difficult.[65] And effective description of M-median subsets $Y \subset X = \{1, 2, ..., N\}$ becomes progressively harder for tricker, yet, explicitly described Θ.

In 1733, Buffon 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.*

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.[66]

Thus, "random" merged with analysis with all its power of "calculus". This is what was hailed by Maxwell and exploited by generations of mathematicians and physists.[67]

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

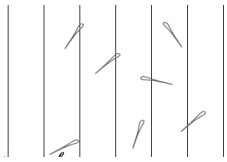
Apparently, the elegance and success of probabilistic models in mathematics and science (always?) depends on (often tacitly assumed and/or hidden) symmetry.

(A bacterium size speck of matter may contain $N_{AT} = 10^{12}$-10^{14} atoms and/or small molecules in it and the number N_{BA} of bacteria residing in your colon is around 10^{12}. If there are two possible states for everyone – be they

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

[66] Besides opening the fields of *geometric probability* and *integral geometry* by this theorem in mathematics, Buffon founded the science of *biogeography.* He also articulated the main premise of the evolutionary biology – the concept of the *common ancestor of all animals*, including humans and he suggested the currently accepted definition of *species.* He designed lenses for lighthouses and concave mirrors that has been in use for two centuries afterwards. 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 of the following centuries.

[67] 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.



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.[68])

And 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.[69]

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

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

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

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

[68]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 this Universe.

[69]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.*

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

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

A particular path to follow is suggested by Boltzmann's way of thinking about statistical mechanics – his ideas invite a use of *non-standard analysis* as well as of a Grothendieck's style *category theoretic* language. (This streamlines Kolmogorov"s ■ in certain applications as we explain in [4].) But a mathematical interpretation of the idea of probability in *languages* and in *learning* needs a more radical deviation from (modification? generalization of?) this ■.

Cardano, Galileo, Buffon. The existence of these people stands in contrast with our picture of a wall separating ego and ergo in the human mind, it challenges our evaluation of the range of the human spirit.

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?

3 Libraries, Dictionaries and Understanding Language

4 Bibliography.

References

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

[2] [SLE] M.Gromov, Structures, Learning and Ergosystems: Chapters. 1-4, 6. http://www.ihes.fr/~gromov/

[3] M. Gromov Quotations and Ideas. http://www.ihes.fr/~gromov/

[4] M. Gromov, In a Search for a Structure, Part 1: On Entropy. http://www.ihes.fr/~gromov/

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

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

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

[8] 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].

[9] 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.

[10] 2 First probabilists: Galileo, Cardano, Fermat, Pascal, Huygens,

https://www.google.fr/search?q=first probabilists&ie=utf-8&oe=utf-8&aq=t&rls=org.

[11] Sources in the History of Probability and Statistics

www.cs.xu.edu/math/Sources/