# Modeling the nature of order: Generative structures and system-centric formalisms for advancing scientific metaphysics*

Petri Lievonen†

*Physics Foundations Society*
*Helsinki, Finland*

(Dated: 2023-05-28)

This study describes how representations (models in their different forms) could emerge in nature in connection with energy minimizing and maximizing processes. Bridging the qualitative and the quantitative is a major challenge. Usually in science, models are treated as epistemic tools by definition, but here their possible ontological ubiquitous presence and development in physical reality is searched for, by studying some basic algebraic equations that are reduced to their absolute essentials only, revealing the mathematically minimal structures capable of providing the abstract function of modeling. Thus, the aim is to display some representational capabilities of generative structures in their bare form, philosophically reminiscent of the ancient metaphysical conceptions of logos, nous, or pratitya-samutpada. The approach here is system-centric and minimalistic, but also simultaneously decentralized and diverse, utilizing linear algebra and the theory of matrices. The resulting mathematical theory of collective modeling, as a variant of factor analysis, summarizes, builds on, and is inspired by the works of Hyötyniemi, which are rooted on one side to the humanistic traditions of systems thinking and cybernetics, and on the other to the technical fields of systems engineering and control theory. The study also lightly problematizes this kind of research, going forward.

"First, the taking in of scattered particulars under one Idea, so that everyone understands what is being talked about... Second, the separation of the Idea into parts, by dividing it at the joints, as nature directs, not breaking any limb in half as a bad carver might."

—Plato, *Phaedrus*, 265I, adapted in the introduction of Alexander (1964), *Notes on the Synthesis of Form*.

## CONTENTS



---

* This discussion paper was formulated very quickly in the early months of 2023, in preparation for a multi-year project "Appearance and Reality in Physics and Beyond: Studies in Scientific Metaphysics and Methodology" at the Department of Philosophy, History and Art Studies at the University of Helsinki. Many thanks to professor emeritus in automation technology Heikki Hyötyniemi for long and frank discussions during the years and for his unrelenting efforts in advancing "formal Hegelian systems" along with his "metavitalism" venture. Also very heartfelt thanks to professor emeritus, philosopher Esa Saarinen for pointers towards systems thinking, applied existentialism, and phenomenology during the years, and mostly unknowingly spurring this research too with the "movement of thought" and "perhaps...?" in mind. The version you are reading was archived online as-is in the autumn of 2024, as a research note reflecting some potential future research ideas. The author laments that many ideas here are quite incomplete and sketchy, but this could still prove out to be useful for some curious readers.

† petri.lievonen@physicsfoundations.org

## I. INTRODUCTION: QUALITY AND QUANTITY

Over a hundred years ago, when modern physics took a new path towards quantum mechanics and relativistic theories, there were also other scientific developments enabled by the recently discovered mathematics of matrices, vectors, and other geometrical constructs. One such development started with the publication of "General Intelligence" measurement by Spearman (1904), widely regarded as one of the earliest appearances of factor analysis in experimental psychology and the social sciences. Later Thurstone (1934, 1935), in *The Vectors of Mind*[1], extended the method to multiple-factor analysis, greatly spurring the development of statistical models and empirically grounded theories across the sciences. There are many historical strands how the methods of data analysis and model building developed from there, resulting in large bodies of knowledge facilitated by the various qualitative and quantitative methods, each with their respective adherents and philosophies of science accompanying them.

In this working paper, I will survey some ideas bridging the qualitative and the quantitative, centering on the theory of matrices—as there factor analysis finds its natural habitat—but also exploring the vast periphery that is not as obviously related to anything specific from first reading. It is my hope that by gathering some of my thoughts on these matters in writing, it could clarify my thinking in the process (as it already has, as hopefully evident in some more refined parts of this exercise at hand), and later the resulting text could foster some discussions about scientific metaphysics and methodology, where some of these ideas could perhaps one day prove out to be useful. As this working paper is in flux, I am receptive to any kind of thoughts, ideas, and suggestions for further work, extracting some parts of it to some suitable venues.

### A. Some fundamentals of the "vectors of mind"

To begin with, factor analysis will serve nicely to prepare ourselves for our purposes here, as it can be seen to be "in between" the qualitative and the quantitative—too technical for many, but not quite rigorous and expressive enough for mathematical physics, for example. Factor analysis exemplifies many useful qualities of good models—the practical utility, accessibility for researchers and scholars with different backgrounds, scalability to massive data sets, paths to extremely general mathematical abstractions—thus illustrative of elementary models available for facilitating structured thinking.

Formally, some variant of factor analysis is present in *every* linear combination of form

$$\hat{y} := Ax, \tag{1}$$

because due to the definition of matrix multiplication in linear algebra[2], the matrix $A$ can always be interpreted to contain factors as columns, that are added together, weighted by the elements in the vector $x$, to produce the result $\hat{y}$. Relations such as in Eq. (1) are often read from right to left. For example, in low dimensionalities,

$$\begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \end{pmatrix} := \begin{pmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}. \tag{2}$$

An alternative, but also equally as accurate, interpretation of the exact same mathematical relation is that the vector $x$ is projected onto the row space of matrix $A$, each such inner product resulting in a corresponding element in vector $\hat{y}$. So there are dualities present already in

[1] See also https://en.wikipedia.org/wiki/The_Vectors_of_Mind

[2] Matrix multiplication was introduced in the early 19th century, by Jacques Philippe Marie Binet in 1812 and Arthur Cayley in 1856, among others. See https://en.wikipedia.org/wiki/Matrix_multiplication and https://en.wikipedia.org/wiki/Linear_algebra.

this extremely minimal relation, and studying also eigenvectors and other fundamental structures such as derivatives in relation to matrices, one can approach discussing the qualitative and the quantitative in terms of continuous and discrete, numerical and symbolic, implicit and explicit, potential and actual, or multidimensional and scalar, for example.

I must emphasize that this is not just some random take from the formal structures in mathematics. Linear products, such as in Eq. (1), really are fundamental in most mathematical models, such as in recent advances in deep learning and other machine learning techniques. But it will take some explaining to understand how they relate to interpretative qualitative methods, for example, so we will return to this only later.

At this point, one can appreciate how the power of the most simple linear algebra (Kurt and Leise, 2006) resulted in serious advances in search engines a couple of decades ago, for example. As an other example, this from brain research, it is well known how a few dimensional linear space is enough to already cover statistically a large range in mental activity (Mitchell *et al.*, 2008).[3] Both of these examples utilize the fundamental idea in modeling: that it is possible to reduce, compress, and abstract observations for some purpose, and do it so that one can concentrate on the most important properties first (according to some criteria), advancing to more complex terms only later, with diminishing returns on gained representational accuracy.[4]

There has always been some tension between the qualitative and the quantitative, and even though there is also mixed methods research, the scientific disciplines stick often very strictly to their favourite approaches, which is only understandable. While interdisciplinary research is important, and discussion between various fields should be encouraged, actually creating lasting traditions in some multi-disciplinary work is perhaps impossible by definition—the traditions usually need to have some very clear core or center to be able to sustain themselves for long. It is better to be explicit even in one's limitations than promising results with too general and diverse approaches. We will try to keep Eq. (1) in mind as a kind of a generating core structure, around which the discussion here will aim to revolve.[5]

### B. Recent advances in modeling meaning

Recently, various machine learning models have started to seriously cross the paths between the qualitative and the quantitative, notably in terms of generation and understanding of text[6], images[7], and speech[8]. While there has been decades of research on understanding and meaning [such as "The Meaning of Meaning" (Ogden and Richards, 1989 [1923]), "The Measurement of Meaning" (Osgood *et al.*, 1957), "Geometry and Meaning" (Widdows, 2004), "Conceptual Spaces: The Geometry of Thought" (Gärdenfors, 2004)], only recently has the qualitative performance reached levels where the professionals on language use are seriously alarmed of the progress. As an example from the humanistic circles, see Roger Berkowitz's short note[9] on Slavoj Zizeks recent essay on "The Post-Human Desert". Berkowitz writes,

> One of the great impacts of science on the human world is that as our knowledge of the world blossomed, the world itself became

[3] See Tom M. Mitchell's presentation at Google on "March 27, 3009" about brains, meaning, and corpus statistics: `https://youtu.be/QbTf2nE3Lbw` [based on (Mitchell *et al.*, 2008)].

[4] This idea of concentrating on "doing the right things, and doing them right" is evident in various series approximations, such as in perturbation series represented by increasingly complicated Feynman diagrams with diminishing effects, or in concentrating analysis on the extrema of the spectrum of a linear operator (for example, the largest eigenvalues of some correlation matrix in factor analysis).

Also it is very interesting how in probability theory the characteristic function (closely related to Fourier transform), the moment-generating function, and especially the cumulant-generating function, allow calculation of cumulants $\kappa_n$ (mean, variance, and higher-order cumulants) by differentiating the transformed expectation value and evaluating it at zero:

$$\kappa_n := \frac{d^n}{dt^n} \ln \mathrm{E}[e^{tX}]\Big|_{t=0}, \tag{3}$$

where $X$ is a random variable, the distribution of which is under scrutiny. As mean and variance are fundamental concepts in any statistical modeling, Eq. (3) suggests that successive derivatives, expectation values, logarithms, and exponentials, may find fundamental uses in modeling generally—and these kind of relations are already present in the theories of thermodynamics, for example.

[5] Note that it could be possible to connect some of these aspirations to the work of David Bohm, Basil Hiley, and their collaborators on *implicate*, *explicate*, and *generative orders*. See Bohm (2002 [1980]), Bohm and Peat (2010 [1987]), also Alexander (2005). My aim is to keep the mathematics as minimal and self-contained as possible, so even if I may refer to their works in the future, I will most likely not refer explicitly to their mathematics of modern physics, unless the structures are exceptionally similar (which they partly will be due to utilizing linear algebra also here).

[6] `https://openai.com/product/gpt-4`
`https://github.com/Stability-AI/StableLM`

[7] `https://imagen.research.google/video/`
`https://openai.com/product/dall-e-2`
`https://github.com/CompVis/stable-diffusion`

[8] `https://openai.com/research/whisper`
`https://speechresearch.github.io/naturalspeech2/`
`https://audit-demo.github.io/`

[9] `https://hac.bard.edu/amor-mundi/artificial-intelligence-and-the-human-condition-2023-04-16`
see also
`https://hac.bard.edu/amor-mundi/the-great-acceleration-2023-05-14`

> ever less comprehensible to humans. [Hannah Arendt wrote that] the "mathematization of physics, by which the absolute renunciation of the senses for the purpose of knowing was carried through, had in its last stages the unexpected and yet plausible consequences that every question man puts to nature is answered in terms of mathematical patterns to which no model can ever be adequate, since one would have to be shaped after our sense experience." For Hannah Arendt writing in The Human Condition [1958], this separation between "thought and sense experience" means that man can create a man-made reality that defies the human capacity to understand or predict that world. This is the reason why it is now scientists, more so than politicians, who are the true actors of our time. They initiate processes that they themselves cannot understand and whose outcome augurs something truly new and unforeseen into the human world.
>
> As we struggle to contemplate the impact of humanly developed but now inhumanly powerful artificially intelligent machines, we would do well to recall some of the lessons Arendt drew already from the victory of science and the modern age. One consequence she recognized was the retreat of philosophy in the face of scientific doing. The point is that science proceeds now freed from human control and thus without need for philosophical justification or philosophical guidance. A second consequence is that that "process" overtakes the product in science so that there is no end point in the development of science. This deprives "man of all fixed and permanent standards and measurements, which, prior to the modern age, have always served him as guides for his doing and criteria for his judgment." The rise of scientific processes, Arendt saw, carries with it the "radical loss of values."

It is also alarming that we cannot know how much, if any, Berkowitz himself has used the large language models of late to assist in writing his quick note. We may enter the simulacra and the hyperreal.[10]

The technological progress is real—listening to the recent talk at MIT[11] by Sébastien Bubeck on GPT-4, titled "Sparks of Artificial General Intelligence: early experiments with GPT-4" (Bubeck *et al.*, 2023), it is evident that many will be genuinely surprised by some of the developments there. A couple of years back Microsoft paid on the order of one billion dollars (mostly by providing access to compute resources, I presume) to get exclusive access to the variants of these GPT (generative pretrained transformer) models, getting some visibility on their development, and Bubeck, a former professor at Princeton who is working at Microsoft, explains how the models we can publicly access are dumbed down versions of the capabilities in existence.[12] Microsoft is not the only player here, of course, as there has been a fervent strive in machine learning for quite some time already, and nobody knows when the "hype cycle" (referring to the terminology of Gartner, for example) will again turn towards diminishing returns. For instance, unofficially (via Google Scholar, as the influential papers on computer science and machine learning are commonly published in conferences, letting the benchmarks do the validation) there are over 70 000 citations to one foundational paper on transformers that is used in large language models, "Attention is all you need" (Vaswani *et al.*, 2017). It suggests a remarkable amount of research and development activity in a short time span. Even with all the fuss, there should still be space for reasoned discussion on these matters.

Taking the message of the above quote by Berkowitz seriously, I argue that instead of steering away from the "mathematical patterns", we should go towards them (such as the structures of "self-attention" in transformers, mentioned above and discussed later at some length),

---

[10] I trust that Berkowitz has integrity and will use the models wisely. The large language models have developed so rapidly that even for me, this may be the last time I write without assistance other than thesaurus. It takes weeks to produce quality text, and one could speed up and focus the process by using large language models. They have not been used in this working paper, yet. For simulacrum, see, for example, Taylor and Saarinen (1994), and reflect that producing such fragmented works has been possible precisely because of some stable structures existing both in their minds and in the society at large, facilitating the work.

[11] "Sparks of AGI: early experiments with GPT-4": `https://youtu.be/qbIk7-JPB2c`. See also Bubeck *et al.* (2023).

[12] I would not be too surprised if later it were found out that some developments are also heavily classified, resulting in serious conflicts of interest for the advancement of science. For an extreme example, read what happened to the dissertation of David Bohm in the field of atomic physics: `https://en.wikipedia.org/wiki/David_Bohm#Manhattan_Project_contributions`

For data sets used in learning the particular model under discussion, one can gather some information from publicly released papers, but Bubeck admits (in the video in fn. 11, around mark 17:35), that practically "we don't know what GPT-4 was trained on—I don't know what GPT-4 was trained on. My working assumption is that it was trained on all, you know, the data digitally produced by humanity—that's my assumption. I'm not saying it's correct, but this is my working assumption, so that you know. I know that, you know, anything which is out there online, GPT-4 might have seen it. So in particular any benchmark whatsoever that exists, I assume it has seen it."

simplifying them as much as practically possible in the process, reaching some kind of "products of science" as idealized models. If the ideas are fundamental, they should also turn out to be understandable and useful for thinking, eventually reaching even the basic education in their reasonable simplicity. This is the work of the scientists and the philosophers.[13]

### C. Building common understanding

Unfortunately, in my experience, this kind of bridging the qualitative and the quantitative meets a lot of resistance, and not only from the qualitative side.

I once collaborated with a professor in applied mathematics, who did not seem to appreciate me writing about linear algebra to the scientific layman in simple terms—he communicated only through some others, that "that is not how mathematics is usually written," and at the same time proclaiming to promote innovativeness, embracing the qualitative and the quantitative! Only later, when I read more into the social sciences, I understood that it was more about taste and understanding. Taste, in terms of conventions that are usually subconscious and taken for granted[14], and understanding, as it later turned out that most experts are actually not too familiar with the beginnings of the methodologies they are experts in—you can be an expert in convex optimization, game theory, general relativity, or quantum mechanics, but actually having a feel how the pioneers had to think to be able to condense the ideas to such simple, canonical formulations, is out of reach of many experts, who are already deeply committed to the resulting conceptual frameworks.

There were many other reasons why the collaboration I mentioned tailed away in the rushes of the everyday, and I am not claiming me being any better here in understanding how mathematics should also be developed and communicated than a seasoned professor, but I am quite adamant that we will need multiple, partly overlapping approaches too, to be able to develop common understanding. We need thinkers such as Hermann Weyl (1989 [1952]), who were so familiar with the fundamentals, that they could be very relaxed and even playful in their communications, to bridge the divides, for the common good.

Even now, preparing this working paper, I received among the first informal comments from one retired professor in a technical field that "you should not write about the solution to the quadratic equation", implying that it is almost stupid to write about such a simple thing—without any regard to its function already in that early draft as an exemplary simple model, reaching basic education. I just wonder, if we are so prohibited to talk about the mundane, how on earth could we even begin reaching for the divine? Metaphorically, perhaps there are also some self-preservation instincts, not unlike how immune systems function, hindering developing common understanding further.

One reading of the difficulties in bridging the qualitative and the quantitative, is also a kind of lack of empathy, which is a skill to be learned. It takes serious effort and experience to be able to understand that there are "otherness", the existence of whom could be completely oblivious to oneself, and the specifics of whose one will not be able to understand (perhaps ever), but which have their own, very justified perspectives.[15] From a quantitative perspective, "the qualitative" is such other, that one can at least try to understand and give space for the

---

[13] It is likely that Berkowitz, citing Arendt, commenting on Zizek, who all have read their Hegel and Marx, has not missed that it is the market logic that is driving the development here, not the scientists themselves as such, and in any case it is questionable what is the ratio of scientists (seeking understanding) to technologists (applying knowledge and power for process control and profit) inside large organizations involved in these developments.

A popular website listing positions for common six-figure salaries in U.S. technology companies compared also the performance of GPT variants to human software engineers (as a April Fools' joke), and while the organizational know-how of the language models is obviously still lacking, the technical outputs of large language models already augment the work of thousands and thousands of people. See `https://www.levels.fyi/?compare=Google,Apple,Microsoft,GPT&track=Software%20Engineer`

I am also reminded of a random quote from Finnish lyrics

> *Miks kaikki kaunis on niin naivia*
> *Ja markkinoiden voimissa vain draivia?*
> (Why is everything beautiful so naive
> And under the market forces only 'drive'?)
> —Eppu Normaali, Hipit Rautaa (1993)

[14] See, for example, Bourdieu (1977, 1984). Bourdieu, as many others, also utilized a variant of factor analysis in his studies (Lebaron, 2009). In (Bourdieu, 1984, p. 466):

> Taste is an acquired disposition to 'differentiate' and 'appreciate', as Kant says—in other words, to establish and mark differences by a process of distinction which is not (or not necessarily) a distinct knowledge, in Leibniz's sense, since it ensures recognition (in the ordinary sense) of the object without implying knowledge of the distinctive features which define it. The schemes of the habitus, the primary forms of classification, owe their specific efficacy to the fact that they function below the level of consciousness and language, beyond the reach of introspective scrutiny or control by the will. [...] Taste is a practical mastery of distributions which makes it possible to sense or intuit what is likely (or unlikely) to befall—and therefore to befit—an individual occupying a given position in social space.

See also `https://en.wikipedia.org/wiki/Practice_theory`.

[15] Studying empathy and different perspectives is a vast topic by itself. For philosophical perspectivism, as an example, see Weckström (2023).

common good, and not just devalue or ignore, even if the mind sets are difficult to comprehend. I am reminded of a graduate seminar, where a student, who was applying only quantitative methods, seemed visually shaken in some kind of a newly found admiration after having read some assigned original paper by Garfinkel on ethnomethodology. He was a bit flabbergasted that the thinking and the concepts defined there were *more precise*, not less, than what he had usually apprehended from the quantitative research methodology literature that he was more familiar with. There is important, high quality thinking everywhere—what are some people thinking, that some other, large fields, have been just fooling around?[16]

Also empathy should be involved in *towards earlier and future selves*—everybody would be better to keep in mind how disorientating it felt when collections of mathematical symbols just appeared as some elaborate mess, and how difficult it was to get to know how various mathematical abstractions have roots in common: for example, linear combinations similar to Eq. (1) can be expressed as series formulas, integrals, Einstein summation, dot products, conjugate products, linear projections, operators with a discrete or continuous spectrum, tensor contractions, bra-ket notation, Ricci notation, etc., and while each have their uses (especially the idea of discreteness and extending the formalisms to some continuous limit is an important distinction), the idea of a linear combination or linearity is fundamental. Due to different levels of abstractions, sticking to only some strict subset of the expert formalisms available, may sometimes result—intentionally or unintentionally—in some kind of "pretense of knowledge", the consequences of which could be quite severe. For instance, some idea may not be taken seriously simply due to it not showing up as meaningful in one's favourite formalism—some novel idea projects to zero, so to speak, or to a direction better to be avoided.

As large language models and other developments in machine learning have now progressed to accessible popular applications where various experts themselves can practically get a feel of the often surprising bridging of the qualitative and the quantitative, we could be at a point in history where, perhaps better than ever, also philosophical discussion on these developments could get new grounding across the disciplines. Suddenly some very abstract ideas, such as the semiosphere of Lotman (2001 [1990])[17], that a few generations of scholars have already built their careers on, may begin to seem "more scientific" than before, simply due to the "structuralist" capability demonstrations of modern language technologies (and other domains of statistical machine learning, such as in speech and vision) being at last so convincing.

Even with the risk of bringing the qualitative and the implicit under the spotlight of technological, instrumental reason—making qualitative more explicit, in a sense—the idea that I am developing here, concerns with finding possible emergent mathematical structures that *are already residing in nature*, bridging the qualitative and the quantitative. In this view, the mathematics of large language models and other error minimizing learning models may be simply approaching some ideal "models of modeling" that will be found in any case, similar to how the modern physics developed during the last century.

## II. SOME IDEAS FROM TRANSFORMER ARCHITECTURES

### A. Transformers as extensions to factor analysis

As one central example, the self-attention in transformer architectures, described in Vaswani *et al.* (2017) and later works, enabling the workings of large language models, can be interpreted as one of the simplest possible extensions to the minimal linear transformation depicted in Eq. (1). Instead of one vector $x$, the elements of which dictate the weighting of factors in the result, we gather multiple vectors as columns in matrix $X$, and let them mix by multiplying with the so called attention matrix $C$ from the right:

$$\hat{Y} := AXC. \tag{4}$$

The matrix $A$ can again always be interpreted to contain factors as columns, as before, that are added together weighted by the elements in the columns $x$ in matrix $X$, to produce each result column $\hat{y}$ in matrix $\hat{Y}$. But now the attention matrix $C$ mixes the columns in $X$ (and by associative linearity, also the results $AX$, so called values in the terminology of transformers), enabling interactive flows between the columns (columns represent different positions in the sequence, for example, subword tokens or a few millisecond long records from some audio spectrogram, as traces or echoes of some realtime process). Important specifics are omitted[18], but some fundamental structures extending factor analysis are now visible.

[16] I myself use inverted thinking here: Most people are alike, I am a person, and I can often be quite a fool; therefore, most people can be quite some fools—and this is ok for me! This also helps in some social settings, as most people are quite perplexed what is going on most of the time—it is only natural—so it is good to take time to discuss the matters in a way that does not straight up assume everybody is on the same map.

[17] We will revisit semiotics and "Universe of the Mind" (Lotman, 2001 [1990]) later in this study.

[18] Important concepts from the conventional transformer architecture that are omitted at this point, include how inner products of so called keys and queries dynamically result in the attention matrix $C$, which is transformed to a kind of an expectation value operator by element-wise exponentiation and column-wise unit normalization, how upper-triangularization of the attention

An example of Eq. (4), in low dimensionalities, reads

$$\begin{pmatrix}\hat{y}_{11} & \hat{y}_{12}\\ \hat{y}_{21} & \hat{y}_{22}\end{pmatrix} := \begin{pmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{pmatrix}\begin{pmatrix}x_{11} & x_{12}\\ x_{21} & x_{22}\end{pmatrix}\begin{pmatrix}c_{11} & c_{12}\\ c_{21} & c_{22}\end{pmatrix} \tag{5}$$

$$= \begin{pmatrix}a_{11}x_{11} + a_{12}x_{21} & a_{11}x_{12} + a_{12}x_{22}\\ a_{21}x_{11} + a_{22}x_{21} & a_{21}x_{12} + a_{22}x_{22}\end{pmatrix}\begin{pmatrix}c_{11} & c_{12}\\ c_{21} & c_{22}\end{pmatrix}$$

$$= \begin{pmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{pmatrix}\begin{pmatrix}x_{11}c_{11} + x_{12}c_{21} & x_{11}c_{12} + x_{12}c_{22}\\ x_{21}c_{11} + x_{22}c_{21} & x_{21}c_{12} + x_{22}c_{22}\end{pmatrix}$$

$$= \begin{pmatrix}\begin{matrix}a_{11}(x_{11}c_{11} + x_{12}c_{21})\\ +a_{12}(x_{21}c_{11} + x_{22}c_{21})\end{matrix} & \begin{matrix}a_{11}(x_{11}c_{12} + x_{12}c_{22})\\ +a_{12}(x_{21}c_{12} + x_{22}c_{22})\end{matrix}\\ \begin{matrix}a_{21}(x_{11}c_{11} + x_{12}c_{21})\\ +a_{22}(x_{21}c_{11} + x_{22}c_{21})\end{matrix} & \begin{matrix}a_{21}(x_{11}c_{12} + x_{12}c_{22})\\ +a_{22}(x_{21}c_{12} + x_{22}c_{22})\end{matrix}\end{pmatrix}$$

$$= \begin{pmatrix}\begin{matrix}a_{11}x_{11}c_{11} + a_{11}x_{12}c_{21}\\ +a_{12}x_{21}c_{11} + a_{12}x_{22}c_{21}\end{matrix} & \begin{matrix}a_{11}x_{11}c_{12} + a_{11}x_{12}c_{22}\\ +a_{12}x_{21}c_{12} + a_{12}x_{22}c_{22}\end{matrix}\\ \begin{matrix}a_{21}x_{11}c_{11} + a_{21}x_{12}c_{21}\\ +a_{22}x_{21}c_{11} + a_{22}x_{22}c_{21}\end{matrix} & \begin{matrix}a_{21}x_{11}c_{12} + a_{21}x_{12}c_{22}\\ +a_{22}x_{21}c_{12} + a_{22}x_{22}c_{22}\end{matrix}\end{pmatrix},$$

where the last form illustrates how the result contains sums of triple products. If the attention matrix $C$ is upper triangular (so $c_{21} = 0$), clearly[19] only the current and previous columns ("the past") in matrix $X$ can have an effect on each result column in matrix $\hat{Y}$. Also if the elements in each column in $C$ sum to one, as in transformers, the resulting vectors in $\hat{Y}$ are weighted averages or expectation values, motivating using the term "attention" here. Note also that if a column in $C$ sums to greater or less than one, the corresponding result vector is amplified or attenuated.

In practice, in current large language models, the dimensionalities of core structures similar to Eq. (4) are quite high.[20] For example, there may be up to $2^{15} = 32\,768$ columns in matrices $\hat{Y}$ and $X$, specifying the maximum context length in tokens (thus gpt-4-32k variant can "read" or perceive up to 50–100 pages of text at once, so roughly 25 000 words).[21] These tokens each have a static embedding available as a high-dimensional vector (column in matrix $X$), which in some latest models is up to 12 288-dimensional, so there may be up to 12 288 rows in matrices $\hat{Y}$ and $X$ in Eq. (4).[22] However, in practice, these rows are separated into perhaps 256 row groups ("heads", in transformer parlance, discussed later), so that the values $AX$, attention matrices $C$, and their key and query factors (again in transformer terminology) are computed using only a subset of the rows at a time, only to be concatenated back to higher dimensional columns later and transformed linearly to that standard dimensionality for the upper layers to consume as input (there may be up to 80 identical layers in production implementations of large language models).

At several points in the transformer architecture there are these linear transformations between spaces of different dimensionalities, and the above should already give a feel of how important the concept of a matrix is in structuring these relationships in the models. Sometimes the mathematical structures in common, such as exemplified by the fundamental simplicity in Eq. (4), are obscured by technical complexity, which is highly unfortunate—resulting in ever more complicated, idiosyncratic "black boxes", that can certainly work well, but which may be difficult to analyze and reason with.

### B. The correlational structure of attention matrices

Also in closer inspection, the structure of the attention matrix $C$ is interesting. Usually it consists of inner products between two linear transformations of the columns in the data matrix $X$ (or in the so called cross-attention, between the columns of two matrices $X'$ and $X''$), fol-

---

matrix enables so called causal or autoregressive-like output in decoder (but omitted in encoding to enable perceiving the input as a whole), how rows are divided into groups (operated by so called "heads"), how position is encoded by summing a kind of a variable-frequency complex waveform to the token embeddings (enabling relative translational invariance in inner products), how values are exponentiated to model output probabilities, how multiple layers are stacked on top of each other using asymmetric nonlinearities (for example, smoothly rectified by multiplying with a sigmoid) and residual connections to increase expressivity, how multiple batches are operated in parallel for efficiency, and other details important for proper functioning. But even though they are omitted at this point, they are not forgotten—some of those operations are really fundamental, i.e. in a mathematical form as minimal as possible, potentially illustrative of fundamental generative structures.

[19] I am using terms such as "clearly" or "simply" as "mathematicians encouragement" here, meaning that the proposition could be clear to somebody in an instant, and to some with many minutes of concentrated study, but the encouragement is that it is not "hard", in the sense of something requiring days, weeks, months, or perhaps years of effort.

[20] The parameters of GPT-4, for example, have not been publicly disclosed, but one can make ballpark estimates from previous models and public developer documentation provided by Microsoft and OpenAI, for example. For some specifics, see `https://learn.microsoft.com/en-us/azure/cognitive-services/openai/concepts/models`

[21] More specifically, the input text is tokenized first (using tiktoken, SentencePiece, or similar) to subword tokens. For example, as exhibited by one tokenizer online (`https://platform.openai.com/tokenizer`), the string " Helsinki" (note the space before) results in the unique token number [47688], and " helsinki" (with lowercase) in three ordered tokens [932, 82, 38799], and each of these dictate the contents of their respective columns in the lowest layer. In that tokenizer, it seems there are 50 257 tokens in the vocabulary, capable of encoding any symbol from byte-level, presumably so there are free numbers up to $2^{16} = 65\,536$ available for special control structures (such as end of speech, and various kinds of separators for user, assistant, and system outputs).

[22] The embeddings themselves have been adapted separately with massive data sets so that statistically nearby tokens have similar vector representations, measured with inner products (cosines).

lowed by several function compositions:

$$\begin{aligned} C &:= f_4\big(f_3\big(f_2\big(f_1(X,X)\big)\big)\big) && (6)\\ &= (f_4 \circ f_3 \circ f_2 \circ f_1)(X,X) && (7)\\ &= (f_4 \circ f_3 \circ f_2)\big((KX)^\dagger QX\big) && (8)\\ &= (f_4 \circ f_3)\big([(KX)^\dagger QX] \odot U\big) && (9)\\ &= f_4\big(\exp.\{\frac{1}{\sqrt{d_K}}[(KX)^\dagger QX] \odot U\}\big) && (10)\\ &= C' \oslash (JC'), && (11) \end{aligned}$$

which is actually simpler than it perhaps looks on a first glance. In the literature, there are several versions that each can be trained with good results, suggesting that there may exist optimal forms that are also computationally efficient [as products in Eq. (4) are associative, one is able to group the operations in several ways, also considering operating in their lower-dimensional representations]. Obviously, the further one proceeds into the technical specifics, the less certain one can be whether the structures are anymore fundamental—but certainly some of those notions should retain their usefulness also in the future, so they seem worth of close inspection.

In Eq. (8), the matrix function $f_1(X,X)$ results in the inner product matrix $(KX)^\dagger QX$, where the keys (column projections via a constant matrix $K$) and queries (column projections via a constant matrix $Q$) are let interfere by inner products (element-wise multiplication and summing, i.e. tensor contraction).[23] In practice, only a subset of rows from some larger matrix $X'$ is processed in this way (divided into so called "heads"), and also the dimension $d_K$ of the resulting projected spaces may be smaller, where the comparisons are done. Also in some architectures, the keys and queries are rectified to positive numbers by zeroing out negative elements, therefore not using element-wise exponentiations in the later stages, only normalization.[24]

In low dimensions, the structure of the resulting inner product matrix $(KX)^\dagger QX$ can be seen in the matrix multiplication[25]

$$\begin{pmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{pmatrix}^\dagger \begin{pmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{pmatrix} = \begin{pmatrix} k_{11}^* & k_{21}^* \\ k_{12}^* & k_{22}^* \end{pmatrix} \begin{pmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{pmatrix} = \begin{pmatrix} k_{11}^* q_{11} + k_{21}^* q_{21} & k_{11}^* q_{12} + k_{21}^* q_{22} \\ k_{12}^* q_{11} + k_{22}^* q_{21} & k_{12}^* q_{12} + k_{22}^* q_{22} \end{pmatrix}, \quad (13)$$

which shows how the resulting elements represent pairwise comparisons among the columns, denoting how much correlation there is, directing the "flow" of values from each query column to each key column when used as an attention matrix $C$ in Eq. (4). By construction, there are obviously some fundamental similarities in Eq. (13) to the structures in mathematical physics.[26]

The element-wise matrix function $f_2(C') := C' \odot U$ is simply an element-wise (Hadamard) product with a unit upper triangular matrix $U$, making the attention matrix causal or autoregressive-like. This constraint is used only when generating the output, but when reading the input, it is usually omitted to allow for a more "whole" attention. In some settings, this operation could result from treating the odd and even parts separately, or perhaps by considering positive vs. negative frequencies.

Continuing the chain of functions to Eq. (10), the element-wise matrix function $f_3(C') := \exp.(C'/\sqrt{d_K})$, after scaling the inner products to approximately unit variance, applies an element-wise exponential map (notice the dot "." here), similar as in thermodynamics and logistic regression. I am sorry about sloppy usage of syntax and terminology here, but as there are inconsistent terms and definitions in the literature in any case, the exact mathematical form itself should be enough to communicate the idea.

The matrix function $f_4(C') := C' \oslash (JC')$ then aggregates the column sums by multiplying with a matrix of ones, $J$, from the left, and then dividing element-wise, resulting in columns that sum to one. The divisor, as sums of exponentiated values, is similar to the canonical partition function in statistical mechanics, where it actually encodes much of the information about a system, to be able to normalize it to unity.[27] In machine learning, the operation of exponentiating the elements and then

[23] Note the obvious complication of using different symbols, matrix transposes, and other operations in the literature, expressing the exact same mathematical relations, but in a different form. Here we will utilize vectors as columns, being mindful of transposes as they may warrant complex conjugations if one would later extend the structures to more general mathematics.

[24] However, rectifying itself is quite commonly done by element-wise multiplying with an element-wise sigmoid function, which is mathematically interesting on its own,

$$xS(x) = \frac{x}{1+e^{-x}} = \frac{xe^x}{e^x+1} = x[1 - S(-x)], \quad (12)$$

as there does not appear to have emerged yet a common understanding of its significance (if any) in dynamic processes. These kind of smoothed rectifiers are used so that the derivatives are well defined, see `https://en.wikipedia.org/wiki/Rectifier_(neural_networks)`. The problem of choosing which nonlinearities to introduce, is complicated by the fact that due to goal oriented optimization, many kinds of nonlinearities work equally well, as the network will often learn to use whatever components it has in its disposal, making comparative analysis more difficult.

[25] In this quick example, $k_{ij}$ and $q_{ij}$ are the projected elements in matrices $KX$ and $QX$, not the elements of projection matrices $K$ and $Q$. Dagger $^\dagger$ denotes the Hermitian conjugate (matrix transpose and complex conjugation), to keep this discussion somewhat general, even though usually the numbers are not complex but approximated real numbers implemented in floating point such as FP32, FP16, FP8, or even FP4.

[26] See `https://en.wikipedia.org/wiki/Density_matrix` and `https://en.wikipedia.org/wiki/S-matrix`, for example.

[27] See, for example `https://en.wikipedia.org/wiki/Partition_function_(statistical_mechanics)`.

The inscription in Ludwig Boltzmann's (1844–1906) grave in normalizing (dividing by the sum) is called softmax, and it is often used mechanically as it has been found to be useful in practice.

### C. Oscillations and waves encoding time and position

There is much more we could say about the potentially fundamental ideas in transformers—such as utilization of the exponential function also in normalizing the output to get probabilities for tokens, that are then actually used in generating the output, via choosing the most probable token, or rather the whole sequence cumulatively with the so called beam search—but to keep the discussion manageable, we will lastly highlight a feature of transformers that is often neglected. That concerns encoding token position in the columns in matrix $X$, as otherwise the ordering of columns could be shuffled at will. The solution is perhaps rather surprising for those who have only used large language models without ever seeing their internals, and it involves waves, starting from the perceptions and affecting in turn all the higher layers in the chain, such as the inner product matrices in Eqs. (8) and (13).

In brief, the aforementioned grouping of "spatial" dimensions (rows) of matrix $X$ into several distinct "heads" may result in modeling different frequency bands in the input, at least on the lower levels.[28] This is due to token position being added at the lowest level, very curiously, to the vector embeddings $x$ as a sort of complex waveform, where the first dimensions (first rows) are oscillating fastest, and the oscillations get slower in the other dimensions of the embeddings:

$$x'_{jk} := x_{jk} + p_{jk} = x_{jk} + e^{i\omega_j k}, \tag{15}$$

where the angular frequency $\omega_j$ decreases according to the row index $j$, by

$$\omega_j := 10000^{-\frac{j}{\max_j}} = e^{-\frac{j}{\max_j}\ln 10\,000}. \tag{16}$$

The magic number 10 000 is a constant that the transformer community has deemed useful in practice. It just means that the highest angular frequency is 1 and lowest 1/10000, decreasing in geometric progression. In practice, transformers are implemented with real numbers (or their approximations as floats or sometimes fixed point integers), so the complex numbers are projected (relative to a constant phase) to their $\cos(\omega_j k)$ and $\sin(\omega_j k)$ parts, and the parts are summed separately to adjacent rows (so only every other $j$ is used in the formula, to keep attaining the lowest frequency at the maximum dimension).

This construction of positional encodings may say a lot to experts in mathematical physics, as now the inner products between different positions can be invariant to absolute position, only the relative position difference being relevant for sliding the context window (the set of columns) across some text material. Dividing the rows into "heads" may now (at least on the lower levels) represent different oscillatory regimes (frequency bands), comparisons among which exhibit relative phase leads and lags. Oscillations such as harmonic motion are also *very* prominent in nature, partly due to time delays and the dynamics of feedback having such robust effects across the scales. It would be only reasonable to study these constructions further, as the sinusoidal position encoding has now gained such strong empirical support, evidently being useful in language modeling, exhibiting at least some properties of possibly emergent intelligence.

However, it may also be problematic that as current language models are working so impressively, one would need to validate any modifications to the constructions empirically by training the modified models to some useful state for comparison. But only few have access to such amounts of training data and compute that the current large language models require [see, for example, Hoffmann *et al.* (2022)]. A couple of decades back advances in machine learning were enabled in large part by performance increases in computing and data storage, as the researchers could start making algorithmic experiments on their workstations in minutes and hours, not days or weeks, as previously required—even using the exact same building blocks (such as gradient backpropagation) that had been known for decades already back then, but now the faster experiment cycle helping getting acquainted with the little details (such as training hyperparameters, different data preprocessings, and layer normalizations) that turned out to be crucially important. But now the models have again progressed to such large scales, that one is again a bit constrained by the complexity of experimentation.

As training the models were mentioned here, note that the importance of adaptation and learning is explored

---

Zentralfriedhof, Vienna (Austria),

$$S = k.\log W, \tag{14}$$

is important and related to a kind of "layered scales" thinking, mentioned briefly later. We should somehow incorporate *scale*, *levels of organization*, *levels of observation*, *levels of explanation*, and the various relationships between them to enable more accurate and comprehensive models. Eq. (14) is interpreted in statistical mechanics as describing entropy $S$ being proportional to the logarithm of the number of microstates (at specific energy $E_i$, so called microcanonical ensemble), which is inversely proportional to the probability of the microstate, $W = 1/p_i$. Statistical mechanics links the empirical thermodynamic properties of a system to the statistical distribution of an ensemble of microstates—all macroscopic thermodynamic properties of a system, such as its temperature, pressure, volume, and density, may be calculated from the partition function that sums $\exp(-E_i/kT)$ terms of all its microstates. See, for example, `https://en.wikipedia.org/wiki/Microstate_(statistical_mechanics)` and the further links therein.

[28] Note that the columns of matrix $X$ represent synchronicity, whereas the rows diachronicity, but the meaning of these terms is intermixed as the elements are let interfere with each other structurally.

only later in this working paper. The main idea is to study the gradients directly almost by hand, thus arriving at closed form solutions and hopefully useful insights, rather than relying on the software framework providing "autograd" though some very complex engineered chain of functions where some popular gradient descent procedure (such as the so called AdamW) would provide the progress towards optima. We will at least try to use rational, principled approach here, aiming to condense some concepts to their absolute minimum, similar to what seemed to be the impetus for mathematical physics a century ago.

### D. Formal similarities could suggest avenues for unification

Wrapping up this brief tour of ideas in popular transformer architectures, as kind of extensions to factor analysis presented in Eqs. (1) and (4), we can envision also some future developments. In large language models, there are usually up to 80 layers of identical structures, as mentioned before, each using the resulting vectors (as abstract "forms" or distributed state) from below (there are also so called residual connections in transformer blocks, facilitating letting the flow through, and only affecting the flows with delta differences). This "unnaturally" large number of layers suggest that some layers could perhaps someday be collapsed onto some computational substrate with some fundamental recurrent and recursive relations (oscillations, reflections, echoes) that are still as yet to be discovered. However, it could also be that this layerized "loop unrolling" as exemplified in transformer architectures, may turn out to be more efficient in some respects than what is found in the structures in nature—the large language models are trained by exposing them to $10^{12}$ (a trillion!) tokens, which seems already something that is usually never encountered in nature.[29] As oscillatory phenomena are notoriously difficult to control without fundamental mathematical and technical understanding (as exemplified by holography, radar technologies, also by the intricacies of quantum field theory), my hope here is that by studying the relations in their bare form, some cross-disciplinary understanding may emerge.

For example, see the vast amounts of research already done on neural oscillations.[30] I am surely not the only one who is wondering if the now provably useful oscillations in the position encoding in transformers, the "cosine schedule" of image diffusion models [mixing noise and signal in polar form from $\pi/2$ to 0, see (Ho *et al.*, 2022), studied later], the fundamental mathematical simplicity of convolutions, scaling, and shifting in the frequency space (also briefly mentioned later), and the asymmetric neural oscillations relative to some emergent phase coherence (see also Kuramoto model) such as in empirically discovered central pattern generators [see Marder and Bucher (2001)], are somehow related.

It should be possible to build towards a research program that could approach gauging these similarities for possible unification, by studying the models mathematically, aiming to elicit their fundamental structures (as also mirrored in mind). In addition to studying the models, essential to this kind of research are also discussions with various experts, probing their intuitions and qualitative understanding, that is not necessarily articulated in the published works.

## III. THE SCIENTIFIC PERSPECTIVE

As bridging the qualitative and the quantitative may easily seem very vague, I feel that I must collect here some thoughts about scientific work, as I see it.

Firstly, do note that this working paper is written with "a blank slate" approach (whether almost two decades of slow thinking and sporadic experimenting on these matters can be considered a blank slate, is another matter)[31],

[29] Here I mean that in nature, a single "centralized system" such as an individual human cannot be directly exposed to such large amounts of text, but the humanity as a total has of course been exposed to all of it and more. So naturally, it is a question of scale, what or who we consider as a single unit, and these abstract questions will be perhaps reflected later on in this study. As a side note, I worry how should one proceed doing this kind of theoretical work without degrading into some dehumanizing nonsense—are there lines we should not cross, bracketing most of it away, or should we perhaps alternate between various approaches? I will try to revisit these questions later in some form.

With regards to large numbers and sequential codes—note also that in biology, in predicting protein folding for instance, the transformers with self-attention have also surpassed previous state-of-the-art methods by high margin, suggestive that the outer products that make up the self-attentive structure could really be reminiscent of something found in nature, in the domain of DNA transcription and gene expression in general, too. See `https://alphafold.com/`.

[30] See `https://en.wikipedia.org/wiki/Neural_oscillation`, where there are many interesting links such as "Coherent activity of large-scale brain activity may form dynamic links between brain areas required for the integration of distributed information" (Varela et al., 2001. "The brainweb: phase synchronization and large-scale integration". Nat. Rev. Neurosci. 2(4): 229–39), or "Among the most important are harmonic (linear) oscillators, limit cycle oscillators, and delayed-feedback oscillators." (Buzsáki G, Draguhn A, 2004. "Neuronal oscillations in cortical networks". Science. 304 (5679): 1926–9.) As I have not delved deeply into that literature, I will just leave these here as notes, without citing them in the list of references.

[31] While I have excellent credentials from my basic schooling and graduated with honors from a technical university, I was fortunate of being able to work part-time at several institutions and research groups during the early years, as a "research assistant" without too many obligations, to concentrate on thinking, reading, and just generally fooling around. While I can only wonder how the caliber of Hegel or the great scientists came to be, it ap-

by which I mean that I have not yet read deeply into the discussions among the scientific metaphysics contemporaries, such the contributors to Ross *et al.* (2013)—but instead aimed to make a freer and relaxed take on these metaphysical issues, how I really see them, for research guidance purposes.

For a scientific perspective, I subscribe to physicist Sean Carroll's view, that "science *is* something, but it is hard to pin down": science is about observing and theorizing, and especially an ongoing process of comparing and analyzing the models, observations, and conducted experiments.[32] The human side, socializing as Carroll put it, in its various forms, is important for all aspects of science—thus, this treatise at hand is among my first communications on a possible ubiquitous, inherent "modeling pursuit" in nature, that I am chasing after with these mathematics and examples from the various sciences.

For another take, in his "Fundamentals: Ten Keys to Reality", physicist Frank Wilczek (2021, pp. 4–5) writes:

> The method of Kepler, Galileo, and Newton combines the humble discipline of respecting the facts and learning from Nature with the systematic chutzpah of using what you think you've learned aggressively, applying it to everywhere you can, even in situations that go beyond your original evidence. If it works, then you've discovered something useful; if it doesn't, then you have learned something important. I've called that attitude Radical Conservatism, and to me it's the essential innovation of the Scientific Revolution.

In this extremely reductionist spirit (but aiming for the advancement of science), we aim to utilize linear algebra and matrices here as far as possible, surely also "in situations that go beyond [...] original evidence". There are certainly limits as to what matrices can represent, but we will need to make some compromises as to the abstract generality and practical comprehensibility of the approach. If we take seriously the fancy vision that there exists fundamental mathematical modeling structures in nature, then linear algebra seems like a workable level of abstraction, some applicable parts of which could be taught already in basic education of the future. For some notational conveniences available, see Table I.

TABLE I: Using matrix notation here for pedagogical conciseness, even though it is not quite general.

| | matrix notation[a] | Ricci notation[b] |
|---|---|---|
| scalar[c] | $a$ | $a$ |
| vector | $x$ | $x^i$ |
| row vector | $x^\dagger$ | $x_i$ |
| matrix | $A$ | $A^i_j$ |
| higher-order tensors[d] | | $A_{ij}$ $B^i_{jk}$ $C^{il}_{jk}$ |
| identity matrix | $I$ | $\delta^i_j$ |
| matrix-vector product[e] | $Ax$ | $A^i_j x^j$ |
| inner product | $y^\dagger x$ | $y_i x^i$ |
| matrix product | $AB$ | $A^i_j B^j_k$ |
| outer product | $yx^\dagger$ | $y^i x_j$ |
| product of elements | $y \odot x = \mathrm{diag}(y)x$ | $y^i x^i$ |
| product of elements | $A \odot B$ | $A^i_j B^i_j$ |
| scaling of columns | $A\,\mathrm{diag}(x)$ | $A^i_j x_j$ |
| scaling of rows | $\mathrm{diag}(x)A$ | $A^i_j x^i$ |

[a] We use matrix notation here for conciseness, but it cannot represent higher-order tensors without elaborate vectorization schemes using Kronecker products. Ricci notation could make some formulas easier and more useful in the future, as it is commutative (the products there are encoded in the ordering of the symbols for the indexes).

[b] This notation is from (Laue *et al.*, 2020), which also lists their modified Einstein notation as the preferred alternative when there is no distinction for upper and lower indexes (when the metric tensor is an identity matrix). For example, $AB = A^i_j B^j_k = A *_{(ij,jk,ik)} B$, where the three index sets correspond to A, B, and the resulting AB, respectively, duplicated symbols denoting the element-wise multiplications. Contractions (sums) are denoted by omitting an index in the output index set. I would perhaps prefer a more symmetric notation where the result index set would be on the center rather than on the right, but I understand that their notation is for compatibility with the syntax of various machine learning frameworks in existence.

[c] In matrix notation, we do not distinguish between scalars and vectors here by using bold type, for example, but aim to make the type clear by context.

[d] In Ricci notation, note the upper (contravariant) and lower (covariant) indexes, the components of which respond inversely to coordinate transforms such as scaling. We will not be using those concepts yet here.

[e] Linear transformations such as this are of fundamental importance. In Ricci notation, note that identical symbols for indexes denote element-wise multiplication, and are only summed over (contracted) if the symbols are on different levels (upper or lower), not on the same level.

## A. Scientific thinking as a variant of factor analysis

As I referred earlier to "The Vectors of Mind" (Thurstone, 1935) as a kind of a seminal publication here, it seems proper to let Thurstone express his thoughts on the nature of science, too. The following may eventually

peared early in the process that one way to increase the chances of contributing something novel later, is to utilize these kind of supporting structures, to be able to keep going for a long time in some more or less definite direction. I heartily thank all my bosses during those years, who knew that something was up and let me do my thing. For the last decade, I have been fortunate of having been able to sustain and extend these endeavors as a small business owner, which also has its ups and downs, of course.

[32] For Sean Carroll's take on the nature of science, see his "The Biggest Ideas in the Universe" series, part 24 at https://youtu.be/ZqphkIO7yt4?t=1272. Also relevant are (Feynman, 1955; Feynman and Leighton, 1985), and many other accounts on the subject.

get a bit tedious with so many direct quotes, but I feel that we should more often see the actual words as written, so the sometimes subtle qualities of thinking could shimmer through, instead of always reducing the ideas to some summarizing abstractions.

Thurstone starts by explaining that (Thurstone, 1935, p. 44)

> This volume is concerned with methods of discovering and identifying significant categories in psychology and in other social sciences. It is therefore of interest to consider some phases of science in general that bear on the problem of finding a methodology for a psychological science.
>
> It is the faith of all science that an unlimited number of phenomena can be comprehended in terms of a limited number of concepts or ideal constructs. Without this faith no science could ever have any motivation. To deny this faith is to affirm the primary chaos of nature and the consequent futility of scientific effort. The constructs in terms of which natural phenomena are comprehended are man-made inventions. To discover a scientific law is merely to discover that a man-made scheme serves to unify, and thereby to simplify, comprehension of a certain class of natural phenomena. A scientific law is not to be thought of as having an independent existence which some scientist is fortunate to stumble upon. A scientific law is not part of nature. It is only a way of comprehending nature.

Here, already, we also encounter the conceptual problem of where does scientific modeling and the resulting models reside, if not in nature? It is common to think of minds and societies (or cultures) as separate from the physical reality, which they are in a sense, but also from a purely evolutionary and prehistoric viewpoint, where development over time is taken seriously, everything must necessarily also have had a basis in nature, too. We will return to this later. Also note that some physical forms and their mathematical relations certainly seem to be reciding ontologically in nature, even if our description of them as "scientific laws" are man-made constructions.

Thurstone (1935, p. 45) continues, that

> The criterion by which a new ideal construct in science is accepted or rejected is the degree to which it facilitates the comprehension of a class of phenomena which can be thought of as examples of a single construct rather than as individualized events. It is in this sense that the chief object of science is to minimize mental effort. But in order that this reduction shall be accepted as science, it must be demonstrated, either explicitly or by implication, that the number of degrees of freedom of the construct is smaller than the number of degrees of freedom of the phenomena that the reduction is expected to subsume.

Referring to Eq. (1), the above can be interpreted with regard to the dimensionalities of $\hat{y}$, $A$, and $x$ (so active elements in $x$ should be relatively few, i.e. effectively low-dimensional). Thurstone also argues here for the notion that models are abstractions and reductions, in relation to the phenomena that a model is about. Thurstone (1935, p. 45) maintains a dynamic view that

> It is in the nature of science that no scientific law can ever be proved to be right. It can only be shown to be plausible. The laws of science are not immutable. They are only human efforts toward parsimony in the comprehension of nature.

Exhibiting wisdom, Thurstone (1935, p. 46) also explains that

> The development of scientific analysis in a new class of phenomena usually meets with resistance. The faith of science that nature can be comprehended in terms of an order acknowledges no limitation whatever as regards classes of phenomena. But scientists are not free from prejudice against the extension of their faith to realms not habitually comprehended in the scientific order. Examples of this resistance are numerous. It is not infrequent for a competent physical scientist to declare his belief that the phenomena of living objects are, at least in some subtle way, beyond the reach of rigorous scientific order.
>
> One of the forms in which this resistance appears is the assertion that, since a scientific construct does not cover all enumerable detail of a class of phenomena, it is therefore to be judged inapplicable.

and that (Thurstone, 1935, p. 47)

> This is again the resistance against invading with the generalizing and simplifying constructs of science a realm which is habitually comprehended only in terms of innumerable and individualized detail. Every scientific construct limits itself to specified variables without any pretense to cover those aspects of a class of phenomena about which it has said nothing. As regards this characteristic of science, there is no difference between the scientific study of physical events and the

> scientific study of biological and psychological events. What is not generally understood, even by many scientists, is that no scientific law is ever intended to represent any event pictorially. The law is only an abstraction from the experimental situation. No experiment is ever completely repeated.
>
> There is an unlimited number of ways in which nature can be comprehended in terms of fundamental scientific concepts. One of the simplest ways in which a class of phenomena can be comprehended in terms of a limited number of concepts is probably that in which a linear attribute of an event is expressed as a linear function of primary causes. Even when the relations are preferably non-linear and mathematically involved, it is frequently possible to use the simpler linear forms as first approximations.

where we are approaching Eq. (1) and the discussion in its vicinity. Thurstone (1935, p. 53) also hints at the apparent similarity of factor analysis and the workings of the mind, but does not analyse it very deeply:

> Factor analysis is reminiscent of faculty psychology. It is true that the object of factor analysis is to discover the mental faculties. The severe restrictions that are imposed by the logic of factor analysis make it an arduous task to isolate each new mental faculty, because it is necessary to prove that it is called for by the experimental observations. Factor analysis does not allow that a new faculty be added as soon as a new name can be found for the things that people can do. In order to prove that reasoning and abstraction are two different faculties, for example, it will be necessary to show that the tasks which call for such activities really do involve two factors, and not one.

Finally, Thurstone (1935, p. 54) reminds the reader of the importance of (and potential impediments to) the advancement of science and modeling:

> In the psychology of the future it may be found useful to postulate a different form of ideal construct for the description of mental endowment than the simple one that is implied in equation (1) [a linear sum also there]. The ideal constructs of the future may involve elements with location in a space frame with spatial, dynamic, and temporal constraints analogous to the ideal constructs of genetics. It would be unfortunate if some initial success with the analytical methods to be described here should lead us to commit ourselves to them with such force of habit as to retard the development of entirely different constructs that may be indicated by improvements in measurement and by inconsistencies between theory and experiment.

It is very common for terminologies and methodologies of science to advance and solidify to such a degree, that the fundamental ideas may become obfuscated. For example, I am sure that there are experts in quantitative methods and statistical analysis, who object to me using the term factor analysis when referring to Eq. (1), as I do not use textbook terminology. But the mathematical relations are the same. I can also see how some applied fields may have gotten used to defining the relations transposed (switching columns with rows)—without having to experience how it may cause complications for complex numbers and their conjugations, when generalizing the formulas.

With these words in mind, I want to briefly touch on the delicate issue of *making science*, which may differ from the ideal scientific perspective.

### B. About style, substance, and scientific integrity

Can you imagine the reality that Richard P. Feynman, among other great scientists, lived so they could write like this (Feynman and Leighton, 1985, p. 343) (emphasis in the original):

> For example, I was a little surprised when I was talking to a friend who was going to go on the radio. He does work on cosmology and astronomy, and he wondered how he would explain what the applications of this work were. "Well," I said, "there aren't any." He said, "Yes, but then we won't get support for more research of this kind." *I* think that's kind of dishonest. If you're representing yourself as a scientist, then you should explain to the layman what you're doing—and if they don't want to support you under those circumstances, then that's their decision.
>
> One example of the principle is this: If you've made up your mind to test a theory, or you want to explain some idea, you should always decide to publish it whichever way it comes out. If we only publish results of a certain kind, we can make the argument look good. We must publish *both* kinds of results.

He also continues, that (Feynman and Leighton, 1985, p. 341) (emphasis in the original)

> ...a kind of scientific integrity, a principle of scientific thought that corresponds to a kind

> of utter honesty—a kind of leaning over backwards. For example, if you're doing an experiment, you should report everything that you think might make it invalid—not only what you think is right about it: other causes that could possibly explain your results; and things you thought of that you've eliminated by some other experiment, and how they worked—to make sure the other fellow can tell they have been eliminated.
>
> Details that could throw doubt on your interpretation must be given, if you know them. You must do the best you can—if you know anything at all wrong, or possibly wrong—to explain it. If you make a theory, for example, and advertise it, or put it out, then you must also put down all the facts that disagree with it, as well as those that agree with it. There is also a more subtle problem. When you have put a lot of ideas together to make an elaborate theory, you want to make sure, when explaining what it fits, that those things it fits are not just the things that gave you the idea for the theory; but that the finished theory makes something else come out right, in addition.
>
> In summary, the idea is to try to give *all* of the information to help others to judge the value of your contribution; not just the information that leads to judgment in one particular direction or another.

Of course, this kind of a comprehensive disclosure of the situation at hand is not always practical. Especially some grand theories may touch so many issues, that one has to strike a balance between enumerating hypothetical failure modes and delivering a coherent, understandable story. But the intent of Feynman seems clear here, and he seems to acknowledge the difficulties in achieving this ideal, too, as he ends his autobiographical book with the following remarkable final paragraph:

> So I have just one wish for you—the good luck to be somewhere where you are free to maintain the kind of integrity I have described, and where you do not feel forced by a need to maintain your position in the organization, or financial support, or so on, to lose your integrity. May you have that freedom.

Of course, we cannot know the reality behind Feynman's account of his everyday—people have different conceptions of themselves and their past circumstances, and in practice, there is necessarily a difference between what people do and what they say they do. Still, intentions matter, and Feynman wanted to express his wish as above.

The challenges of maintaining the kind of scientific integrity Feynman called for may be even greater in qualitative research, as it is even more difficult to justify one's ideas and findings there. It must be emphasized, that it is the domain, the subject of research that is complicated—it is not only that the research endeavor and methods there are somehow fuzzy.

As an example of the difficulties of integrity on the qualitative side, here is a plea from Christopher Alexander (1999) in the field of philosophy of architecture:

> That is because of the world view that we have and the way of looking at things and the nervousness about intellectual rigor ... that people of our era have. Although they have these judgments [about life and quality in some living structure] within them, somehow they are separated from their ability to make these judgments correctly. This is just some sort of childish instinct that everybody has and knows. But, for some reason, we are so messed up that we can't see it.

Especially compare the above published text to the actual, lively take on the same, by watching his improvised speech, where the article is adapted from, given live in 1996 (my transcript below starts at time 25:40)[33]:

> Now the nature of these experiments is very peculiar...in a way. Because, what these experiments ask—let's suppose that what we were trying to, um—[we, let, we, unintelligible, extends his left arm to the audience as a plea] we got a sidewalk somewhere [gestures to his left], a bit of a street, and we got another sidewalk [gestures to his right], another bit of a street—and we're trying to come to conclusions about which one... [swirls both of his hands in air and joins them together in front of him like embracing a cradle] is a more living structure. [takes support of the podium, touches his head, corrects posture] ...and my belief, by the way—and I probably should start with this—I mean, when I began, trying to find these experimental methods, my belief always was that there really is such a thing, and that actually, everybody kind of knows it, but that it has been suppressed, that is because of the...world view that we have, and the way of looking at things, and the nervousness about, uh, intellectual rigor—that in a way

[33] Christopher Alexander: Patterns in Architecture (1996, `https://youtu.be/98LdFA-_zfA?t=1540`)

> people, though they have these judgments within them, somehow are separated from their ability to make these judgments correctly. [glances a bit desperately at the distance] In other words, what I'm trying to say is—and this is just with the sort of instinct that I had going in—was that this is something childish, really, that everybody knows, but for some reason we're so messed up that we can't see it.

Reading the above differences in expressiveness as a warning, this endeavor of mine may be doomed from the start.[34] At this point, it also seems proper to voice some words of caution related to these kind of exploratory studies aiming to get a handle on the emergent modeling in nature.

### C. Impediments to scientific understanding and progress

#### 1. Personal reservations

First of all, as I am seriously aiming to reduce the complexity to some simpler abstractions, and at the same time aiming to create and preserve the diversity in the world, it seems as if I am my own protagonist and antagonist either way, which is a slightly disturbing thought, and it affects the work.

Also, in the same vein, I am not sure how much of my multi-year collaboration with Hyötyniemi has been in the spirit of "keep your enemies close". Nothing personally against Hyötyniemi, of course, just that cybernetics as a field is only marginally humanistic (or perhaps "pseudo-humanistic"), and at its core it seems to repeat an ethos of technocratic knowledge interest and control.[35] One only needs to glance at one of the originators of cybernetics, Norbert Wiener's, book titles, such as "Cybernetics: Or Control and Communication in the Animal and the Machine" (Wiener, 2019 [1948]) or "The Human Use of Human Beings: Cybernetics and Society" (Wiener, 1988 [1950]), which is, while laudable, also an oxymoron.

The works of Hyötyniemi such as "Neocybernetics in Biological Systems" (Hyötyniemi, 2006) and later studies, are very ambitious and creative works[36], and the later developments of his are even more ambitious and unrestrained, and I would say even philosophically relevant records (but as they are in Finnish and I myself have been a sponsor and editor of some of those works, I will not comment them specifically much here). But it has become clear that a professor in automation technology experimenting with autoethnographic studies, without any previous experience on qualitative research methods at all, results in overall output that is rather messy at parts. As the solid substance and the often hard to approach or even off-putting to many, I presume, style is so intertwined, it demands a lot from a prospective reader to be able to separate the wheat from the chaff, as the saying goes, for one's own purposes. But the substance is really there, I believe (also backed by the praise the scant expert readers have given off-the-record).

I originally aimed to model some of the works of Hyötyniemi in the spirit of aforementioned Christopher Alexander's profound works, such as (Alexander, 1964, 2005).[37] Alexander studied mathematics too, so there are similarities, but he made his career in architecture, which is closer to the humanities, so these works of Hyötyniemi have quite a different feel to them compared

---

[34] At least the audience should be decided—now I am writing mostly for myself at this point, and I tolerate a lot of variation in ideas and approaches, which cannot be presumed from a general reader.

In a chapter titled "The text as process of movement: *Author to Audience, Author to Text*" Lotman (2001 [1990], pp. 63) contemplates:

> A text and its readership are in a relationship of mutual activation: a text strives to make its readers conform to itself, to force on them its own system of codes, and the readers respond in the same way. The text as it were contains an image of its 'own' ideal readership, and the readership one if its 'own' text. There is a story about the celebrated mathematician P. L. Chebyshev. An unexpected audience consisting of tailors, modistes and fashionable young ladies turned up to one of his lectures on the subject of the mathematical problem of cutting the cloth [*rasskroika tkani*]. But the lecturer's opening words: 'Let's suppose for simplicity's sake that the human body is spherical' put them to flight. Only the mathematicians who found nothing strange in the remark stayed on to hear him. The text 'selected' its own audience, creating it in its own image and likeness.

[35] For a more balanced take (but where the perspective of instrumental reason is still evident), see perhaps (Midgley, 2003), or for an alternative reading of cybernetics (that is more UK than US based), see Pickering (2010). But one can also compare these to some influential works in humanistic sociology, such as Berger and Luckmann (1966), to see that the same kind of constructivistic ideas can be arrived at also from empirical work.

[36] See, for example, Hyötyniemi presenting his work at Helsinki Institute for Information Technology HIIT (2013) `https://youtu.be/frOzDw1vtdw`.

[37] Four-volume "The Nature of Order: An Essay on the Art of Building and the Nature of the Universe" (Alexander, 2005) book series consists of almost 2 500 pages. The titles of books are "The Phenomenon of Life", "The Process of Creating Life", "A Vision of a Living World", and "The Luminous Ground". See also (Alexander, 1999), derived from his keynote speech (`https://youtu.be/98LdFA-_zfA`). Also the works of his collaborator Nikos Salingaros are related, emphasizing the often forgotten importance of the ornament and logarithmic scales, for pleasant, alternating symmetries that do not collapse totally when compacting them in perception, due to interesting variability.

to the "quality without a name" (wholeness, in Alexander's terms) of Alexander. But of course, overall one can be quite pleased that so many areas of the thinking of Hyötyniemi have been recorded in a written form, even with these quite scarce professional resources—a bit like a hobby project, but keeping the flame alive so it could result in studies later, too. If the ideas are fundamental, then there is time, and if they are not fundamental, it is only commendable to not having rushed it, potentially misleading someone (or oneself).

One can also compare the system-oriented style of Hyötyniemi to Pietak (2011), and I am sure one could find many other reference points if one would participate in the fields (that Hyötyniemi has not done much due to his health troubles, it is not very practical to travel to conferences on a wheelchair, for example—the works are contributions in isolation).[38] In her book Pietak mentioned how these kind of "vitalistic" and "holistic" approaches became notorious after the first half of the last century, partly due to them being associated with political movements [see discussion, for example, in Gilbert and Sarkar (2000)]. Later Pietak has been active in developing advanced simulations of morphogenesis with biologist Michael Levin using Clifford algebras, which partly inspired me, too, towards getting acquainted with geometric algebras a couple of years back, which I am grateful for.

### 2. Unknown applications

All that is manageable, but one hindrance I see here is that the cyberneticians and seekers of the past did not need to worry about somebody actually implementing their ideas (or they were so early thinkers, that one could easily ignore any inner worries), but now we have to. See Midgley (2003), for example, for an impressive four-volume collection of articles from various cyberneticians and systems thinkers. There are very bold and far-reaching ideas, but mostly they have been applied by public speakers and various authors, and mainly in small group dynamics—facilitating workshop events, for instance. In comparison, current machine learning models affect directly entire populations, via recommendation and filtering algorithms, for example, and there are many takers and implementors for new ideas.

So it would perhaps not be too far-fetched to apply some precautionary principle, similar as in ecology, when developing these modeling ideas towards proper comprehension (I mean towards such clear thinking that the ideas could be synthesized materially, where things would be practically realizable). But at the same time it is against the spirit of basic research, to somehow hinder or obstruct free thinking prematurely.

This kind of highly speculative research has the characteristic that it sounds like science fiction and even sloppy and wrong, before it perhaps does not anymore. One only needs to remind oneself what kind of theoretical and practical resources the most capable minds had in their disposal only two hundred years ago—when the modern discoveries such as quantum mechanics would have sounded utterly bonkers in most intellectual discussions—to convince oneself of the possibility, and even inevitability, that new theories could (and will) start new developments that are currently almost unimaginable. For comic and fictitious takes on the subject, see tables and figures in the end of this working paper. (Table III on page 60 and Fig. 1 on page 61, for example.)

### 3. Burden of knowledge

There is also a worry (at least in my mind), that what good can come out of deep self-knowledge, really. For some questions, see Wilson and Dunn (2004). Consider-

[38] For example, Fields medalist René Thom (1989 [1972]) writes (orig. in French): "Contrary to what is generally believed about the two traditionally opposed theories of biology, vitalism and reductionism, it is the attitude of the reductionist that is metaphysical. He postulates a reduction of living processes to pure physiochemistry, but such a reduction has never been experimentally established. Vitalism, on the other hand, deals with the striking collection of facts about regulation and finality which cover almost all aspects of living activities; but it is discredited by its hollow terminology (e.g., Driesch's organizing principle and entelechies), a fault accepted and exaggerated by subsequent teleological philosophers (Bergson, Teilhard de Chardin). We must not judge these thinkers too severely, however; their work contains many daring ideas that those who are hidebound by mechanistic taboos can never glimpse. Even the terminology of Driesch is evidence of the mind's need to understand a situation that has no analogy in the inanimate world. [...] If the biologist is to progress and to understand living processes, he cannot wait until physics and chemistry can give him complete theory of all local phenomena found in living matter; instead, he should try only to construct a model that is locally compatible with known properties of environment and to separate off the geometricoalgebraic structure ensuring the stability of the system, without attempting a complete description of living matter. This methodology goes against the present dominant philosophy that the first step in revealing nature must be the analysis of the system and its ultimate constituents. We must reject this primitive and almost cannibalistic delusion about knowledge, that an understanding of something requires first that we dismantle it, like a child who pulls a watch to pieces and spreads out the wheels in order to understand the mechanism."

Thom continues that: "Our method of attributing a formal geometrical structure to a living being, to explain its stability, may be thought of as a kind of *geometrical vitalism*; it provides a global structure controlling the local details like Driesch's entelechy. But this structure can, in principle, be explained solely by local determinisms, theoretically reducible to mechanisms of a physicochemical type. I do not know whether such a reduction can be carried out in detail; nevertheless, I believe that an understanding of this formal structure will be useful even when its physicochemical justification is incomplete or unsatisfactory."

See also the classic "On Growth and Form" (Thompson, 2007 [1961/1942/1917]).

ing mostly self-knowledge in the spirit of "Know thyself!" is good, I think, but in relation to the studies suggested here, it may some day be possible to *really* know the workings of one's own mind, for example, from some fundamental, mathematically understood level. Similarly, some possible systemic laws elucidated later, such as the molding effects of correlations, inherent modeling pursuit, or the suggestive power of reductions, could also potentially transform (but not always transcend positively) one's view on ordinary life, such as relationships with the significant others, for instance. Or perhaps one could become intensively aware of the specific larger forces affecting the history of one's life decisions—when "a new kind of spectre is haunting..." The implications of these are uncertain, and people may have different expectations and opinions here, of course.

Self-knowledge, in a more abstract sense, could also hinder the effortless transparency of the senses. Part of the reason for models existing in the first place is their ability to provide reliable transparency—by that I mean that using a proper model one does not need to be surprised for each single event, and thus one can concentrate thinking on relevant matters and tasks at hand. There is a sort of routinization that is essential for everyday functioning, providing ontological security that things are as they appear.[39] It would be quite inconvenient, for example, if perceiving a red rose would not simply be a red rose, but one would be unnecessarily aware of all the modeling links in the chain, including that in the eye there are several photoreceptors with partly overlapping wavelength sensitivities, the sense of smell has very deep links to the molecular chemistry and thus various structures of any biological being, or that a rose as an object is actually that multilevel, living, vibrating thing, constantly in microscopic flux, where the stable dynamics keep it observable as a single entity. Some parts of the association networks and their cultural developmental histories that are in play in the concept of red roses, for example, is of course beautifully life-enhancing and usually appreciated if said aloud, but sometimes things are better to be left implicit. So even though we can become aware of constantly using models, it is not always wise to do so.[40]

#### 4. Dominating technical knowledge interest

In scientific work, but also generally, models are also deeply related to truth and knowledge. The traditional components of knowledge, such as justified, true beliefs, have natural understanding in terms of common, shared experiences, codes, and resources for motivation and rationalization (enabling justifications), and correspondence with reality by not causing too many surprises when depending on them in applications and everyday life (defining true contextual realities)—thus naturally materializing to persistent habits (beliefs, that one is ready to act on). But there are also always some reasonings or interests for pursuing, acquiring, and applying the knowledge in the first place, and that is the debatable issue here. Habermas, which I have never read directly, defined knowledge interests in a memorable "technisch, praktisch, kritisch" trisection (that I did not find a suitable reference yet, but I seem to have some recollection of them from my studies in management and organization

---

[39] Giddens (1984) put it nicely—for him routinization means the usual, self-evident nature of the majority of everyday life events. It results in familiar styles and forms of behavior that support and are supported by a sense of ontological security, which refers to the certainty or confidence that natural and social environments are what they appear to be, including the existence of the basic constructs such as self and social identity.

[40] See also Ch. 12, "Simplicity" in Comte-Sponville (2002), and its other chapters, emphasizing the importance for sanity of being able to forget, for example. Writing about virtues, in French, there are remarkable paragraphs—such as this one on "fidelity" (as also in the term Hi-Fi), generalized to a kind of quality of persistence of models and habits (Comte-Sponville, 2002, pp. 20–21) (emphasis in the original):

"Why would I keep yesterday's promise since I am no longer the same today? Why indeed? Out of fidelity. According to Montaigne, in fidelity lies the true basis of personal identity: "The foundation of my being and identity is purely moral; it consists in the fidelity to the faith I swore to myself. I am not really the same as yesterday; I am the same only because I *admit* to being the same, because I take responsibility of a certain past as *my own*, and because I intend to recognize my present commitment as still my own in the future." There is no moral subject without fidelity of oneself to oneself, which is why fidelity is an obligation: without it there could be no such thing as duty. It is also what makes infidelity possible: just as fidelity is memory as virtue, so infidelity is memory as vice (rather than a mere lapse of memory). Recollection is not everything: having a good memory is not always good, accurate recollections are not always loving or respectful. Memory the virtue entails more than just memory; fidelity entails more than accuracy. Indeed, fidelity is the opposite not of forgetfulness but of frivolous and self-serving fickleness, of disavowal, treachery, and inconstancy. It is true that fidelity struggles against forgetfulness, against the forgetfulness that infidelity ultimately entails: first you betray what you remember, then you forget what you betrayed. Infidelity thus self-destructs as it triumphs, while fidelity can triumph only—and always only provisionally—by refusing to be annihilated, that is, by repeatedly and endlessly struggling against forgetfulness and denial. *Desperate* fidelity, writes Jankélévitch, and I certainly would not take issue with his view of an "unequal struggle between the irresistible tide of oblivion that eventually engulfs all things, and the desperate but intermittent protests of memory." In advising us to forget, Jankélévitch continues, "the exponents of pardon are recommending something that needs no recommendation; the forgetful need no reminders to forget, there is nothing they want more. The past, on the other hand, is in need of our compassion and gratitude; for the past cannot stand up for itself as can the present and the future." Such is the duty of memory: compassion and gratitude for the past. The difficult, demanding, imprescriptible duty of fidelity!"

research)[41], that I think could be useful to cover here.

Technical interest promotes instrumental reason, where knowledge is sought to be able to control something towards some ends. Work is central, and ethos is of improving improvement, so efficiency, automation, streamlining, etc. It is a great partner to natural sciences, but also to other empirical and analytical endeavors. Technical interest is so deeply involved in current socio-economic way of life that it is sometimes difficult to even see its dominance. For many, it is such a natural way of thinking that anything else is just naivety, or perhaps "leaving money on the table", without understanding the systemic effects of such thinking—one does not think of one's family, fellow beings, or art, for example, from a point of view of technical interest.

Practical interest, in Habermas work as I have understood, by contrast, is about people, understanding, and language. There are vast worlds of hermeneutic and interpretive scholarship, that may be completely alien to somebody approaching them from instrumental reason: "what is the use of this?!" Due to the emphasis on practical needs of going along, aiming for understanding, and being mindful of language, the ethos is more like thinking about thinking (also together, and also about the intellectual history at large). It is also understandable that a distinguished "high culture" and intellectualism can cause some tensions, but there is also a certain quality to that kind of thinking and the resulting works that may be difficult to achieve otherwise.

The third knowledge interest, according to Habermas' triad, critical interest, is separate due to its active role as a critic of power structures in society, aiming for emancipation, building reflective thinking, or developing sociological imagination. The ethos is more like uncovering the information asymmetries that are in place, accidentally or intentionally—such as when one does not speak about the importance of ownership or its logic, to be able to keep the talking points on some safe, mundane matters (which is just the reality of ordinary life, but the effect at aggregate is on another level). As many collective agreements are done with asymmetric information, where something is withheld, so that even the withholding is concealed (and mostly without being conscious of it), or people are otherwise misdirected and their trust exploited, the critical theory should have its place in the discussions about knowledge and its use.

### 5. Large organizations have overwhelming resources

Thus, there is also the collective level coming into picture with scientific communications in the open about aspirations for potential modeling breakthroughs, as it is certainly not only individuals and their good intentions that matter in the long run. Sociologist Robert Jackall (2010 [1988], p. 4) explains this memorably in his classic "Moral Mazes: The World of Corporate Managers":[42]

> Bureaucratic work shapes people's consciousness in decisive ways. Among other things, it regularizes people's experiences of time and indeed routinizes their lives by engaging them on a daily basis in rational, socially approved, purposive action; it brings them into daily proximity with and subordination to authority, creating in the process upward-looking stances that have decisive social and psychological consequences; it places a premium on a functionally rational, pragmatic habit of mind that seeks specific goals; and it creates subtle measures of prestige and an elaborate status hierarchy that, in addition to fostering an intense competition for status, also makes the rules, procedures, social contexts, and protocol of an organization paramount psychological and behavioral guides. [...] As a result, bureaucratic work causes people to bracket, while at work, the moralities that they might hold outside the workplace or that they might adhere to privately and to follow instead the prevailing morality of their particular organizational situation. As a former

---

[41] As I have never read Habermas directly, I will just insert here some quotes from Wikipedia for now. Habermas has "a two-level concept of society that integrates the lifeworld and systems paradigms". Citing p. 267 of Habermas (1987 [1981]), *Theory of Communicative Action, Volume Two: Lifeworld and System: A Critique of Functionalist Reason* (Beacon Press, Boston), an article at `https://en.wikipedia.org/wiki/The_Theory_of_Communicative_Action` states:

> Following Weber again, an increasing complexity arises from the structural and institutional differentiation of the lifeworld, which follows the closed logic of the systemic rationalisation of our communications. There is a transfer of action co-ordination from 'language' over to 'steering media', such as money and power, which bypass consensus-oriented communication with a 'symbolic generalisation of rewards and punishments'. After this process the lifeworld "is no longer needed for the coordination of action". This results in humans ('lifeworld actors') losing a sense of responsibility with a chain of negative social consequences. Lifeworld communications lose their purpose becoming irrelevant for the coordination of central life processes. This has the effect of ripping the heart out of social discourse, allowing complex differentiation to occur but at the cost of social pathologies.

See also, for example, Martti J. Kari referring briefly to Habermas, using a diagram by Rauno Kuusisto (internal/external vs. past/future, through facts/values/norms/objectives): `https://www.youtube.com/watch?v=kF9KretXqJw&t=1278s`.

[42] See also `https://en.wikipedia.org/wiki/Moral_Mazes`.

> vice-president of a large firm says: "What is right in the corporation is not what is right in a man's home or in his church. *What is right in the corporation is what the guy above you wants from you.* That's what morality is in the corporation." (emphasis by the subject in the original)

Apprehending the fundamental amorality described in the above quote (not necessarily wrongness or badness *per se*, just amorality), it is only reasonable that one should be vary of large corporations and bureaucracies, and by extension, market logic in its extreme forms. Even if as individuals, people almost always strive for good, there are still major problems in corporate governance and policy design (or mechanism design), especially in relation to the so called "tech companies" that would harvest and profile all the data of the world, and carry out behavioral alignment for profit, if it would be possible for them. Because, why not?

I have many friends who have done multi-year gigs at large corporations, and they often have signed such egregious non-disclosure agreements, that they do not even want to (or cannot) talk about signing them. The companies themselves are also compartmentalized for various purposes. In the end, "they also pay well", so it is no wonder that there is a conflict of public interest here.[43]

Add to the above that "bad is stronger than good"[44] (Baumeister *et al.*, 2001), and one really starts to wonder whether applying some sort of precautionary principle would make any difference in the long run, after all. Could the whole modern scientific enterprise come to a crisis, if studies on emergent modeling (such as advanced machine learning) really turns into something comparable to what the atomic physicists had to confront a hundred years ago, but now in a fundamentally more difficult setting involving biology, minds, societies, and more?[45]

## 6. Models affect the reality they model

Practically, we are responsible for the theories and concepts we create and develop. Words are important. In physics, perhaps our models will not affect the actual physics (even though the models certainly have great effects on the reality too, by enabling new technologies and projects, for example), but in many other fields of science the theories definitely affect the phenomena they study. In a classic article "Economics language and assumptions: How theories can become self-fulfilling", Ferraro *et al.* (2005) caution that

> Social science theories can become self-fulfilling by shaping institutional designs and management practices, as well as social norms and expectations about behavior, thereby creating the behavior they predict. They also perpetuate themselves by promulgating language and assumptions that become widely used and accepted. We illustrate these ideas by considering how the language and assumptions of economics shape management practices: theories can "win" in the marketplace for ideas, independent of their empirical validity, to the extent their assumptions and

---

[43] The same may also apply to large research organizations such as universities, depending how tight bureaucratic control the administration has managed to get—without no one really wanting this to happen. There are tons of studies on these creeping oversight issues in the organizations research literature from many decades of experienced reality of organizational life. Due to inefficiencies in large organizations, the externalized cost to the society is mostly wasted opportunity and talent, but perhaps also deteriorating public sphere (if there are no true academics left who can speak freely, with a sense of responsibility to the greater good as opposed to just myopic survival or overhyped production of something mildly interesting and not too confrontational). But of course this is just naive thinking at this point. Knowledge enables power, and the possible qualitative thresholds, where we would need to start thinking what kind of advances in knowledge technologies to disseminate to corporations, are impossible to know beforehand—the corporations will just apply whatever works, and tell only the minimum necessary for public relations and preserving trust. There are no mechanisms to ensure that future developments would really benefit the public in the long run, if technologies can truly get a hold on these complex phenomena such as minds in nature.

[44] Citing (Baumeister *et al.*, 2001) here is, of course, slightly tongue-in-cheek, but only slightly. Already the abstract of the review article is notable in its breadth:

> The greater power of bad events over good ones is found in everyday events, major life events (e.g., trauma), close relationship outcomes, social network patterns, interpersonal interactions, and learning processes. Bad emotions, bad parents, and bad feedback have more impact than good ones, and bad information is processed more thoroughly than good. The self is more motivated to avoid bad self-definitions than to pursue good ones. Bad impressions and bad stereotypes are quicker to form and more resistant to disconfirmation than good ones. Various explanations such as diagnosticity and salience help explain some findings, but the greater power of bad events is still found when such variables are controlled. Hardly any exceptions (indicating greater power of good) can be found. Taken together, these findings suggest that bad is stronger than good, as a general principle across a broad range of psychological phenomena.

[45] This working paper seems to concentrate mostly on minds in terms of modeling and meaning, which may have enormous societal issues by itself, and I am not even discussing here the effects on biology, where things could get very drastic at some point. For the most depressive take (which only slightly overhypes the specificity of genome targets at this point in time), see John Sotos, Chief Medical Officer at Intel, sharing his thoughts at DEF CON 25 (2017): "Genetic Diseases to Guide Digital Hacks of the Human Genome" (https://youtu.be/HKQDSgBHPfY?t=778).

> language become taken for granted and normatively valued, therefore creating conditions that make them come "true."[46]

The same can also be true for models of complex behavior, as studied here. For example, if a model expects that everything is predicated on competition for resources, then that guides the strategies that people and collectives will devise, even if that may not yet be the best interpretation of the model. But then again, we should also recognize that some of the grand theories, that have not been discovered yet, may also turn out to be something that we may not like.

In science and technology studies, there is a distinction between technology forecasting and technology foresight, the latter including also active shaping of and preparing for the future developments. But if there is some kind of technology determinism, even, on the large scale development towards ever more advanced machine learning algorithms, for example, then what use is for the individual to voice any caution—perhaps best would be to simply participate in the evolution, and drive those interests that seem important. At the ultimate end of this working paper, Table IV on page 62, there is a list of classical virtues that could be emphasized more often.[47]

### D. Importance of tradition, genuine innovativity, and the absurdity of common understanding

Let me start wrapping this section on the "scientific perspective", including that obscure "impediments on scientific understanding", to something a bit more concrete—or perhaps even more obscure. There clearly are things that we are talking about, without actually talking about them—not mentioning them specifically, perhaps even exercising some kind of psychological or social repression here—so perhaps using some stories could help us in elucidating and getting some grip on the matters here.

Many creative people have had experiences where someone talks familiarly as if some thing—be it a project, job, or some other theme—is really easy, and we are the same people, in the same in-group so to speak, that there is even traction and some kind of suction in it. Then in some practical matter, such as in relation to some paper work, one ends up to the other side of the glass (or actually in some trash bin), when some kind of institutional or Bourdieuan taste criteria then reveals itself somehow in a different way, and without anyone understanding what is happening in terms of this meta-level.[48] A friend of mine recently had this kind of an experience, where the personal presence seemed to match, but in the end it did not go anywhere, and it was suggested that he shows the papers to some career consultant. How would such polishing change the actual work? Via modeling, of course, as fitting better to the setting, the organization getting more of the same, at least on paper (as the career consultants would be more than happy to provide).

Recently, I have been thinking about Feynman's realization, which I am sure many have read about in "Surely You're Joking, Mr. Feynman!" [the title of which is a Bourdieuan taste reference, too, in (Feynman and Leighton, 1985, p. 60)], when he had wondered about lucid dreaming, in which he could use his mind exponentially more efficiently, like sorting out the millions of strands of some lady's hair precisely—then realizing that in the end it was all about the concept of concepts, not the thing itself. He just *thought* he could perceive those strands of hair accurately, but of course he could not actually do so, but within that conceptual system (in the dream) this error cannot be understood. It is also the kind of logic of nightmares, where some seemingly ordinary thing can emanate some conceptually extreme emotion. As Feynman recollects (Feynman and Leighton, 1985, pp. 49–51) (emphasis in the original):

> I made some other observations while dreaming. Apart from always asking myself, "Am I *really* dreaming in color?" I wondered, "How accurately do you see something?"
>
> The next time I had a dream, there was a girl lying in tall grass, and she had red hair. I tried to see if I could see *each* hair. You know how there's a little area of color just where the sun is reflecting—the diffraction effect, I could see *that!* I could see each hair as sharp as you want: perfect vision!
>
> [...]
>
> You might like to know how this process of observing my dreams stopped (which it has for the most part; it's happened just a

---

[46] For extreme version of bending the truth, intentionally, see `https://en.wikipedia.org/wiki/Firehose_of_falsehood` which is all too familiar to everybody these days. Also large language models are already involved in media environment, and will be more in the future. As one anonymous internet commenter said: "[A popular chat model] is, presently, like having an assistant who is patient, incredibly well-read, sycophantic, impressionable, amoral, psychopathic, and prone to bouts of delusional confidence and confabulation. The precautions we would take engaging with that kind of person, are actually rather useful defenses against dangerous AI outputs." But of course the massive scales make the difference, the quantity being a quality of its own, and only slight biasing of the information space on a large scale is enough for many purposes.

[47] See also `https://en.wikipedia.org/wiki/Virtue`. For example, one can contemplate the meaning of *Gravitas* "A sense of the importance of the matter at hand; responsibility, and being earnest" and *Humanitas* "Refinement, civilization, learning, and generally being cultured."

[48] For Bourdieu, see also fn. (14).

> few times since). I'm dreaming one night as usual, making observations, and I see on the wall in front of me a pennant. I answer for the twenty-fifth time, "Yes, I'm dreaming in color," and then I realize that I've been sleeping with the back of my head against a brass rod. I put my hand behind my head and I feel that the back of my head is *soft*. I think, "Aha! *That's* why I've been able to make all these observations in my dreams: the brass rod has disturbed my visual cortex. All I have to do is sleep with a brass rod under my head, and I can make these observations any time I want. So I think I'll stop making these observations on this one, and go into deeper sleep."
>
> When I woke up later, there was no brass rod, nor was the back of my head soft. Somehow I had become tired of making these observations, and my brain had invented some false reasons as to why I shouldn't do it any more.
>
> As a result of these observations I began to get a little theory. One of the reasons that I liked to look at dreams was that I was curious as to how you can see an image, of a person, for example, when your eyes are closed, and nothing's coming in. You say it might be random, irregular nerve discharges, but you can't get the nerves to discharge in exactly the same delicate patterns when you are sleeping as when you are awake, looking at something. Well then, how could I "see" in color, and in better detail, when I was asleep?
>
> I decided there must be an "interpretation department." When you are actually looking at something—a man, a lamp, or a wall—you don't just see blotches of color. Something tells you what it is; it has to be interpreted. When you're dreaming, this interpretation department is still operating, but it's all slopped up. It's telling you that you're seeing a human hair in the greatest detail, when it isn't true. It's interpreting the random junk entering the brain as a clear image.

The Bourdieuan taken-for-granted evaluation criteria situation that my friend earlier found himself in, reminded me again, that in most situations, we just *imagine* that there is some connection in common—it is a kind of a shared abstraction and reduction, and the scientific circles are no strangers to this, either. That is also why it is very important to hold on to even the smallest things that seem to last, and on the other hand, sometimes it is important to just proceed, as if there is some commonly agreed on thing going on in there, while really there is not a consensus yet, and likely won't be.[49]

Now, considering what has been said until this point, let us appreciate how the Levin lab at Tufts University dares to proclaim:

> Our mission is to develop fundamental understanding of how minds of all kinds arise, scale, persist, and change; we seek to use that knowledge to benefit the embodied experience of sentient beings, through biomedicine and beyond.[50]

It may finally be our conviction—in this working paper—that the concepts in the Levin labs quote about "minds of all kinds" will be understood, found natural, and seen to apply already not only in the various domains of biology, but also in macro scales such as societies, via mathematical isomorphisms, as idealizations of some systemic phenomena.[51] It sounds quite grand, but it is an expected development, simply extrapolated into the future how science has progressed so far.

Max Planck has written (the source of this quote is uncertain—here one could read it in terms of life sciences, too):

> Briefly summarized, what I did [at the turn of the 20th century] can be described as simply an act of desperation. By nature I am peacefully inclined and reject all doubtful adventures.
>
> But by then I had been wrestling unsuccessfully for 6 years with the problem of equilibrium between radiation and matter and I

[49] In philosophy, also unconventional (but otherwise rational) viewpoints are appreciated, but they may be quite extraordinary from the perspective of ordinary life. For example, looking at the various alarmist writings on AI among the humanities, I am left wondering that soon large language models can be used to craft beautiful lectures, perhaps even masterpieces of writing such as "On Bullshit" (Frankfurt, 2005), or "A Small Treatise on the Great Virtues: The Uses of Philosophy in Everyday Life" (Comte-Sponville, 2002). At what point can language models be used to educate the public en masse about grand ideas, such as inner workings of minds and societies (in yet unforeseen and unprecedented way), or perhaps even the potential realities of evolutionary nature itself?

Similarly to one would give solace to his or her parents, "thank you so much, it is in good hands, the story continues well", the language models (not the actual models at present, I am speaking abstractly here) could do the same, comforting humanity amid the end of one era. I am not suggesting a sudden end, but a slow cultural and mental realization. It is a startling thought: if one were satisfied that the story really would continue well (with rich models and whatnot), many would be ok, going "with the times"—not that they would have much of a choice, but still better than some other scenarios. Opportunities for philosophical pondering are plenty.

[50] `https://www.drmichaellevin.org/`

[51] See the various sections in this working paper as a whole, also see Vygotsky (1978), Pattee (1973), Ahl and Allen (1996), Lotman (2001 [1990]), Mead (2015 [1934]), *et cetera*.

> knew that this problem was of fundamental importance to physics.
>
> A theoretical interpretation therefore had to be found at any cost, no matter how high.

Of course, we cannot know exactly what does it mean to "wrestle unsuccessfully for 6 years with the problem" desperate for something that "had to be found at any cost, no matter how high", but Planck's (1858–1947) tombstone at Göttingen cemetery displays now the Planck constant, $h = 6.62 \cdot 10^{-34}$ W$\cdot$s$^2$, and science progressed.[52]

At the risk of sounding pathetic, and not understanding the associations properly in German language, I will end this section with an encouragement by David Hilbert (1862–1943), also in his tombstone at Göttingen cemetery:

> *wir mussen wissen*
> *wir werden wissen*
> (we need to know
> we will know).

## IV. THE IDEA OF ONTIC MODELING IN NATURE

### A. An example of ubiquitous models and modeling

In a companion study (Lievonen, 2023), I surveyed Suntola's novel cosmological model, which aims to base the physics of the expanding universe on firm foundations using the most basic concepts such as mass (*what* in kilograms), distance (*where* in meters), and time (*when* in seconds). The model follows the constraints and affordances of the $S^3(r_0)$ geometry, which describes an expanding three-sphere as a model for the cosmos, from the macroscopic perfect symmetries to the microscopic details, aspiring "toward a unified picture of physical reality" (Suntola, 2018).

With the above description, we can discern several layers of modeling already at play here: as a physical book, the model is, de facto, a representation of mental imagery of the workings of nature, in nature (or in culture)—and my self-referential description here adds further links to the possible modeling chain. All these layers are also very real, of course, as the representations are stored and transmitted in some physical form (be it digital or in paper, perhaps illuminated by light or read aloud), and as biological and sentient beings, our attention and memory networks process what we are sensing in the work, invoking relevant (or not) habitual capabilities, in a sustained executive loop with streams of thought, action, emotion, and perception. The concept of time is of fundamental importance here, as with it, one can separate the "eternal now in a loop" from the gone past and the becoming future.

The various representations involved in the chains have finite spatial and temporal extent—the models and their referents develop and decay, according to their relevance and the characteristic time spans, ranging from microseconds and less, to centuries and more, sometimes on a whim. There is an abundance of thinking in the history of philosophy about these issues.[53]

So in this section, I explore some of those representational meso-level ("ordinary") structural components of physical reality (as opposed to almost infinitesimally micro physics at the Planck scale, or the supremely macro cosmos at large). The aim in the distant future is to eventually be able to augment the aforementioned fundamental measures of *what*, *where*, and *when*, with at least some preliminary but dependable steps towards facing such important metaphysical questions about the "minds" as to (1) *how* (function, in terms of abstracted representation only), (2) *why* (driven by a creative tension, in terms of necessity and the modeling instinct, as explained later), and even (3) *who* (the relational subject with agency, the diverse minds in nature).

The following will be obviously very fragmentary at this point. The thinking here is in flux, and it will hopefully (or regrettably, depending on one's viewpoint on these developments) solidify during the research.

[52] In 1958, Wolfgang Pauli lamented to Jagdish Mehra, in Berkeley, California: "The best that most of us can hope to achieve in physics is simply to misunderstand at a deeper level", as quoted in Mehra (2001).

[53] For example, Peirce (1906, pp. 523–524) writes (emphasis in the original): Thought is not necessarily connected with a brain. It appears in the work of bees, of crystals, and throughout the purely physical world; and one can no more deny that it is really there, than that the colors, the shapes, etc. of objects are really there. Consistently adhere to that unwarrantable denial, and you will be driven to some form of idealistic nominalism akin to Fichte's. Not only is thought in the organic world, but it develops there. But as there cannot be a General without Instances embodying it, so there cannot be thought without Signs. We must here give "Sign" a very wide sense, no doubt, but not too wide a sense to come within our definition. Admitting that connected Signs must have a Quasi-mind, it may further be declared that there can be no isolated sign. Moreover, signs require at least two Quasi-minds; a *Quasi-utterer* and a *Quasi-interpreter*; and although these two are at one (i.e. *are* one mind) in the sign itself, they must nevertheless be distinct. In the Sign they are, so to say, *welded*. Accordingly, it is not merely a fact of human Psychology, but a necessity of Logic, that every logical evolution of thought should be dialogic. You may say that all this is loose talk; and I admit that, as it stands, it has a large infusion of arbitrariness. It might be filled out with argument so as to remove the greater part of this fault; but in the first place, such an expansion would require a volume,—and an uninviting one; and in the second place, what I have been saying is only to be applied to a slight determination of our system of diagrammatization, which it will only slightly affect; so that, should it be incorrect, the utmost *certain* effect will be a danger that our system *may* not represent every variety of non-human thought.

### B. The genesis of logos and nous

*Logos* is a term used in Western philosophy, psychology, and rhetoric, and refers to the appeal to reason that relies on logic, or inductive and deductive reasoning. This is one dictionary definition. There is considerable leeway in interpreting the concept, as in some early writings "logos provided the link between rational discourse and the world's rational structure."[54] The triumph of the various sciences has emphasized the import of the vast landscapes of mathematics in structuring the rational discourse, in inferring the rational structure of the world, and in forging the links between the two.

Similarly, *nous*, sometimes equated to intellect or intelligence, is a concept from classical philosophy for the faculty of the human mind necessary for understanding what is true or real. This is also a dictionary definition, and here, too, there is substantial freedom in using the word. For example, "among some Greek authors, a faculty of intelligence known as a 'higher mind' came to be considered as a property of the cosmos as a whole."[55] Also abduction (proposed by C. S. Peirce to complement induction and deduction), and various studies on intuition, are related. For these strands, there are many lively discussions that are simply too numerous to list here.[56]

Both of these can be seen also as variants of modeling in nature, as hopefully will become apparent.

More abstractly—before going into the mathematics—the most general metaphysical phrasing of the genesis argument developed here for the emergent modeling in nature, could perhaps be seen in the ancient concept of pratītya-samutpāda or paṭicca-samuppāda, commonly translated as *dependent origination* or *dependent arising*. It states that all phenomena (dharmas) arise in dependence upon other phenomena: "if this exists, that exists; if this ceases to exist, that also ceases to exist". It is a key doctrine in all schools of Buddhism, and the basic principle is that all things (dharmas, phenomena, principles) arise in dependence upon other things.[57]

There are many scholarly references available, but as I am not an expert on these various worlds of historical and contemporary writings (such as eastern philosophies and vedic thought), at this point we might as well use general characterizations from a dictionary or encyclopedia as specimens for ways of thinking to be also appreciated here. Wikipedia refers to Kalupahana (1975, pp. 54–60), summarizing:

> Dependent origination can be contrasted with the classic Western concept of causation in which an action by one thing is said to cause a change in another thing. Dependent origination instead views the change as being caused by many factors, not just one or even a few.

I have only glanced at the impressive Kalupahana (1975) on this kind of a multidimensional "thermodynamic" or "stochastic" causality, and the above quote from a crowd sourced dictionary reflects only faintly the depths of thinking in the literature, of course. We are not going to dive too deeply there, either. We will reflect on these concepts here, but mostly between the lines and sometimes only in passing—but they may create an important background framing for the work overall.

### C. Representations as abstractions: essential and accidental complexity

Difficulties in this kind of condensed and abstract communication are common in metaphysics, where the most general ideas understandably relate to a multitude of other ideas and contexts, each with their own further ideas and applicable contexts. It may be confusing and frustrating for some, and even furiously "inefficient", "unproductive", or "wasteful" for some others, but I consider that kind of variation in modeling and conversation, both spatially and temporally, resulting in incompatible and at times even inconsistent thinking, actually *necessary for life to even exist* in its different forms. If there is no uncertainty, there is no story, and without stories (even in their abstract forms as models in time and about time), there is no life as we know it [see Bruner (2003), as a poignant example]. At the same time, certainty is important, even vitally important, too—and, well, this is life!

For many, we are on strange waters here—the discussion in this section so far seems to relate the conceptions of reasoning, origins, causation, and modeling, in some variant of self-referential writing style. That really is the intention here. From my point of view, reasoning about causation and possible mechanisms is way too often thought in a setting where the discrete concepts are already at hand (i.e. modeled in some code such as some shared language of the specialist, not always being conscious of the reduction), and the actual multidimensional relations towards the various micro, macro, and meso "worlds" are bracketed away (and usually with a reason). This may certainly seem like superfluous obfuscation and diversion at this point, but to me, the need for this kind of variation and complementary alternations in modeling is

---

[54] See, for example, `https://en.wikipedia.org/wiki/Logos`

[55] See, for example, `https://en.wikipedia.org/wiki/Nous`

[56] See, for example, Deely (2001) and `https://en.wikipedia.org/wiki/Nous#Plotinus_and_Neoplatonism`.

[57] This is a crowd sourced dictionary definition, see `https://en.wikipedia.org/wiki/Pratityasamutpada` and the references therein for scholarly discussion.

As for the mathematics, one could already appreciate how in Eq. (1), zero $x$ results in zero $\hat{y}$ (and vice versa), and how in the longer term (not displayed yet here), zero variance, results in zero $A$, the system simply not existing.

simply self-evident for truly important things, especially in the ideation phase.

In the field of computer science, especially among the practitioners, there is an important distinction between *accidental complexity* and *essential complexity*. This has its roots in the works of Aristotle. It has been found that in practice, most complexity, it seems, is actually accidental, and we should avoid it in our artifacts, socio-technical systems, theories, and even manners. Paradoxically, the proliferation of scientific discourse seems to have brought in more accidental complexity, not less, as it is easier to just reiterate and generate stuff than to refine it in a comprehensive manner. This important observation is also driving the works of Suntola and Hyötyniemi that I have been aiming to survey in these companion studies—both most obviously seem to have a working hypothesis in common that the essential complexity of the world (and by extension, proper models) is actually less, not more, than commonly believed to be necessary. But there are also domains where the essential complexity is necessarily high, so we should not always strive only for extreme reduction and parsimony. The accidental complexity can bring essential richness to life.

To state it in another way, in "The Forms of Meaning: Modeling Systems Theory and Semiotic Analysis" Sebeok and Danesi (2000, pp. 188–189) end the book with the following three paragraphs, quoted here in full:

> We conclude by pointing out that human beings have always shown the ability to disentangle themselves from the "tangled web", as Cassirer called it [in An Essay on Man: An Introduction to a Philosophy of Human Culture, 1944, p. 25]. In each individual human being there is a a continual juxtaposition of individually-based vs. culturally-based modeling. Indeed, culture cannot stamp out the individual human being's need and quest for new forms of meaning. The ability to make signs anew to represent changing realities, new ideas, new ways of thinking is the essence of anthroposemiosis. This innate creative propensity is the reason why cultural symbols are constantly being modified to meet new demands, new ideas, new discoveries, new challenges.
>
> Human modeling often involves things that cannot be seen. Unique among species, connective modeling allows humans to get a look, so to speak, at this hidden world. The poet and the scientist alike use similar modeling systems to extrapolate a suspected inner connection among things. When their models are accepted as fact, they enter human life, taking on an independent conceptual existence in the real world, and thus can suggest ways in which to bring about changes to the world.
>
> As we have stated in the preface to this book, model-making typifies all aspects of human cognitive and social life. From toy and miniature models to scientific theories of the universe, models are so common that we hardly ever take notice of their importance and of their raison d'être in our species. Model-making constitutes a truly astonishing evolutionary attainment, without which it would be virtually impossible for us to carry out our daily life routines and to encode knowledge. We reiterate that the presence of a modeling instinct in the human species is to human intellectual and social life what the physical instincts are to its biological life. As Thomas Szasz (1920–[2012]), the great American psychiatrist aptly put it, "in the animal kingdom, the rule is, eat or be eaten; in the human kingdom, define or be defined".

The ending statement is unnecessarily confrontational, of course, but it correctly emphasizes the important meaning and power of words and concepts—which "were originally magic"[58].

So in that vein, but much more whimsically (and also in some way more specifically), the metaphysical genesis argument for ontic modeling developed here could be seen condensed in the profound proverb (in Finnish) "hätä keinot keksii", which is often translated as "necessity is the mother of invention", or even "desperate times call for desperate measures". It is similar to "where there is a will, there is a way", but seen from a point of view of "true nature", which is relentless, a vital source of inspiration, but also full of suffering too, and ultimately always greater.

In this line of thinking, emergent modeling is happening at all the scales, all the time, and understanding the workings of those generative structures will be paramount to be able to continue reasoning about system-wide processes with some newly found reassurance, it is hoped.

### D. Different worlds and layers of representation

There are vast worlds, but they are not infinite.[59] Similarly as there may be a definite amount of space and

[58] "Words were originally magic" is the title of a book by Steve de Shazer. See also the works of Paul Watzlawick, who hold an opinion that one cannot not communicate, and thus one cannot not influence. I do not aim to emphasize these works too much here, though, as the worlds of communication are so vast and diverse.

[59] Really, if we go distant enough back in time, things were conceptualized simpler than today. For example, in the old texts from

matter in the cosmos at this very moment (for example, $10^{53}$ kilograms, in a volume of $10^{80}$ cubic meters, estimates depending on the model), all the other phenomena we can observe are also finite, for better or worse.

For example, there are about $10^{11}$ (100 billion) neurons in the human brain and elsewhere in the body, and approximately $5.31 \times 10^{11}$ human cells in a kilogram of tissue (across all cell types, organs, etc., calculated from an estimation for a 70 kg human body). Counting also the microbiome in the body perhaps doubles the total cell count, but does not change the scale. In comparison, there are about $5.014 \times 10^{25}$ atoms (83 moles) in a kilogram of carbon, so clearly organs, cells, and organelles are vast worlds by themselves, as the scale difference in the exponent is so huge.

Towards larger spatial and temporal scales, by contrast, the population of Finland is about $5.5 \times 10^6$ persons at the time of writing, and there are about $8 \times 10^9$ (eight billion) people on this planet, over 10% of the total amount *ever* lived. Current estimates are that there are $2 \times 10^{16}$ ants (which have about $2.5 \times 10^5$ neurons each), across over 15 000 known ant species and subspecies.

I deem these materialistic abstractions necessary, as this is the reality—by seeing "through it", life is also "something more".

Temporal scale is similarly layered with finite numbers. Human heart rate is usually 60 to 100 beats per minute, so one beat per second, and there is $3.154 \times 10^7$ seconds in a year, thus a heart of a 70 years old adult has beat on the order of $10^9$ (a billion) times (omitting lots of the complications here, of course). There is only a limited number of sunrises, sunsets, and yearly seasons an individual can experience.

In comparison, the processes on the atomic scale can be very rapid. For instance, our technologies for operating at the nanoscale are already rather impressive, as the clock frequency of the central processing unit (CPU) in a modern laptop or phone could top at 3 GHz ($3 \times 10^9$ times per second), capable of operating perhaps up to 512 bits wide instructions in a single cycle—and there are multiple cores, each with a cache hierarchy of increasingly longer latencies, and the clock rate is usually throttled to save energy—but in that single reliably recurring cycle of approximately 0.3 nanoseconds, light can travel only about 10 centimeters (in a vacuum). As light propagates at quite a staggering velocity of almost 300 kilometers per millisecond, this should illustrate to anybody the amazing intricacies in the most ordinary technologies of the everyday.

Also for longer time spans, one can contemplate the geological or cosmological time scales of $10^9$ years (billions of years), or $10^{16}$ seconds.

Using the scientific notation of base ten is illustrative here, as it really seems as if there are different orders of magnitude in these meso-worlds, and transforming to logarithmic amounts (where only the exponent is left) is the correct way to compare their scales. The base of the logarithm is not too important, as the resulting scales are linearly proportional.

To get a more refined sense of the finite scales of reality, it is illustrative to ponder on other more abstract, but at the same time in some sense more familiar to many, quantities. For example, one can in principle list the exact amount of business establishments (for example, incorporated companies) in a country by the order of magnitude of their turnover (or some other indicator) in their financial statements—it will be a distribution, but actually each company belongs to exactly one segment there. Every company has an identity number (ID), similarly to every person having an ID in the official databases (at least in countries like Finland). Vehicles have license plates, newspapers and books are also indexed, and there really is some definite number of these at each moment.

For example, there are on the order of $10^7$ (10 million) business establishments in the United States, and if one changes the definition of what is counted in (i.e. alters the pattern matching and reduction model of reality), there is still some definite number in existence. Most IDs, such as persons and vehicles, are also at specific locations at each moment, and various entities aim to track these for their purposes. Every square meter of land has been indexed, too,[60] and due to most everything having been recorded during the grand project of modernization, every piece of land, and each company share, has a definite owner (that the courts will have to decide in the end, when contested). Of course, these networks are vast and convoluted (taking into account various ownership chains, and public or state owners, too), but in principle, *they are really there*, and they are finite.

It is quite surprising how difficult it is to accept this definite reality. One would think that there would be more frameworks by now for thinking about these scales

---

India written in the Pali language [see Jones (2011)], five specific worlds are distinguished: physical, biological, mental, karmic (or social), and transcendent. I currently have only the excellent Finnish translation of the work, so will not be able to quote the relevant sections here in verbatim, but one can appreciate how the five worlds relate rather naturally to the three worlds of Popper, the various institutional theories of society [see Scott (2013); Scott and Davis (2007) and Heiskala (2003), for example], and to the discussions among the philosophers of mathematics whether mathematics is invented or discovered, for example. It is interesting to ponder on the possible order of appearance of each of these worlds (which may arise many times, both spatially and temporally diversified, of course).

60 See de Botton (2003) for delightful essays on, for example, how the explorers of the past could advance common knowledge by simply recording some facts. For instance, one could measure the dimensions of a town square in some distant settlement to contribute to the geographical sciences.

naturally, but still it feels as if this belonged to the field of marketing analytics or "big data" of social media corporations, not philosophy. But stretching the imagination, this is the kind of massive complexity what the great philosophers of nature such as Spinoza, Leibniz, and Hegel really wanted to understand and conceptualize, in my mind. Without proper developments in discussing and analysing these kinds of vast worlds, the relation between the individuals and the micro and macro collectives may continue to stay hazy.[61]

In some sense, it seems as if the population dynamics are appreciated in the contemporary discourse, but perhaps not in the complementary sense that is being explored here. In this line of thought, the populations are "really there", but also the individuals are "really there". So, as a way of example, thinking about the teachers and daycare workers in a country as a collective is important due to *every single soul being important*, not least because an individual educator or caretaker will develop only a finite amount of relationships with the kids during his or her life, and each is important (as stated in the most translated document of all time, the Universal Declaration of Human Rights, which itself is of immense value because it is not necessarily true across all times and places).

Inferring from the above (and using also common sense), it seems that many spatially and temporally finite domains of *what* must necessarily be elevated to face the elusive question of *who* (selves, in personal terms, and at various scales, if the meaning of elevating *what* is taken seriously). This is not meant to reduce the concept of being human to mere matter (even if it partially does so in the process), rather the aim is to expand the concept of minds and selves to various contexts, where—as many have argued throughout the history of ideas—the mind already necessarily resides.

The view espoused here also considers problematic both extreme individualism and extreme collectivism—both are in some sense ignorant of the natural realities in the same way some people before Copernicus, Kepler, Galileo, and others, were ignorant of the workings of the solar system. But the important shifting of the "centers of origin" or "outsider's perspective" that happens in each individual's life at some crucial points, seems perhaps even trivial compared to the difficulties in comprehending the various relations towards "the micro" and "the macro". The "point of view" of distributed phenomena such as populations and their statistical distributions may be even more difficult to take in, I presume.

With these complications in mind, the much studied question of mind versus matter could be of help—brains could serve as representative instances in need of theory, of these larger classes of distinctions in the proposed ontic multilayered modeling in nature, including society. Next we will briefly touch philosophy of mind, but only briefly, to keep this study from "not requiring a volume, and an uninviting one" (referring to fn. 53), at this early point.

### E. Mind and matter as an example: neutral monism and dual aspect theories

I am not going to argue too much for or against the separation of mind and matter here, and I will discuss materialism, idealism, and their mixtures only roughly, for the purposes of this working paper. The usual concept of a mind is an anthropocentric term, and current biology increasingly sees agency (such as competence, intentionality, and goal-seeking)[62] empirically at every level, without any kind of hard thresholds where the mind suddenly emerges [see, for example, Bongard and Levin (2023) and Gilbert and Sarkar (2000)][63].

Going the other way, towards macro structures (both spatially and temporally), the agency can be attributed to all kinds of shared representations, from where the aggregate actions emanate in the society (see, for example, Heiskala (2003), Scott (2013), and the history of sociology, anthropology, and culture studies in general). It is, of course, a contested issue how collective action and alignment should be conceptualized[64], but from the point of view of this study, it is self-evident that societies and other collectives do exist, and that they have intentions and agency, as have many other systems in

[61] If we had more common frameworks available to conceptualize the individual-collective -relation, perhaps the "malaises of modernity', as diagnosed by Taylor (1991), for example—such as too strong focus on individualism (as opposed to shared stories and projects), overarching instrumentalism (as opposed to something having intrinsic value), and mere consumerism (as opposed to active participation in steering the common world)—could be understood better [see also (Taylor, 1989) for a more positive take].

[62] In philosophy of mind, intentionality also refers to aboutness, and it could be accompanied with reflective, self-referring thinking also in biology due to feedback loops, where the internal state affects the state itself after some delays, especially when "trying to maintain" the abstract homeostatic balance, of crucial importance in any biologically stable entity.

[63] See also Michael Levin's talk "Cell Intelligence in Physiological and Morphological Spaces" (2022) `https://youtu.be/jLiHLDrOTW8`

[64] The references are simply too much to list here—see, for example, the works of Raimo Tuomela, or perhaps the concept of social action in `https://en.wikipedia.org/wiki/Social_action` or `https://en.wikipedia.org/wiki/Structural_functionalism`, or the vast works in game theory, for example, as referred in Seppänen (2000), including `https://en.wikipedia.org/wiki/Social_choice_theory` and `https://plato.stanford.edu/entries/social-choice/`. I am sorry for pointing towards just some links in here, as they are certainly not comprehensive, but they could work as discussion starters for the purposes of this working paper at this point.

nature—because there really are no specific thresholds for the mind to emerge towards the micro, why should there be any sudden thresholds towards the macro.

Therefore, I assume that the concepts circulating in the mind vs. matter discussions can serve as inspiration for modeling emergent modeling in nature (including individual vs. population dynamics and their associated "material vs. ideal" and "intrinsic vs. extrinsic" perspectives) at every level, even if by doing so the analogies have to be stretched to the limit.

Spinoza (1632–1677) is considered an early progenitor to discussions relating the body and the mind with one substance, later advanced and consolidated by Ernst Mach, which serves well for our purposes, as Mach was also referred to when I studied Suntola in a companion study to this. In their encyclopedia article on neutral monism, Stubenberg and Wishon (2023) summarize that:

> Ernst Mach (1838–1916) occupies a central position in the history of neutral monism. [...] For Mach the world presents itself as "a viscous mass [of elements], at certain places (as in the ego) more firmly coherent than in others" (Mach, 1959 [1886], p. 17). The neutral elements (only a minute fraction of which are sensations) and their relations are the basic reality. We draw boundaries around certain groups of elements that are related to each other in interesting ways, because this serves our biological, scientific, and/or practical purposes. We can continue to talk about material things and selves; it is economical to do so. But, strictly speaking, "both [object and ego] are provisional fictions of the same kind" (Mach, 1976 [1905], p. 9).[65]

As a side note, framing things as "provisional fictions" is quite common in many multilayered fields of inquiry. For example, in anthropology literature, see how "money's fictions continue to surprise" in Maurer (2006), even though those fictions clearly have real effects. It is often illustrative to study the literature of specialist fields, as they contain nice domain-specific details, where the abstracted frameworks are grounded.

For current experts' thinking in relation to Mach, see Preston (2021), which I have not studied yet.

There are many thinkers in this "neutral monism" camp, and it seems that also panpsychism and phenomenalism are related. For example, Stubenberg and Wishon (2023) refer to Banks (2014) as

> his neutral monism is "a kind of physicalism" (Banks, 2014, pp. 7, 142). Banks takes himself to follow Russell embracing an ontology of events as manifestations of underlying powers or energies—such as electromagnetism, gravitation, and nuclear forces (Banks, 2014, p. 149), as well as neural energies (Banks, 2014, p. 142). But event particulars such as these, Banks insists, "are so physicalistic in nature that there does not seem to be any reason to assume that these natural qualities in physics have anything at all in common with our sensations, which are qualities of a very different order... [involving] events in the human nervous system at a very different scale of complexity and size" (Banks, 2014, p. 156).[66]

Note that I have omitted the contexts of the these quotes, where the concepts are also problematized. There is a lot of discussion and framing available in terms of materialism, idealism, dualism, and this neutral monism, which claims to be the simpler option.

I will include here a few paragraphs from the same Stubenberg and Wishon (2023):

> Another common objection to neutral monism is that it constitutes a form of property dualism or dual aspect theory. The argument is straightforward: On the neutral

---

[65] Stubenberg and Wishon (2023) also state that

> Mach [...] alternates between claims such as that "the supposed unities 'body' and 'ego' are only makeshifts, designed for provisional survey and for certain practical ends" and claims that they are composites constituted by "a more strongly coherent group of elements [that is] less strongly connected with other groups of this kin" (Mach, 1959 [1886], pp. 20–21).

[66] Stubenberg and Wishon (2023) also summarize Banks (2014) as:

> Events have, and are individuated by, intrinsic characters or concrete qualities. None of those qualities are mental; but experience familiarizes us with some of them [see Banks (2014, p. 6)]. These qualities are the ways certain powers manifest themselves in events [see Banks (2014, p. 6)]. Examples of such powers (or energies) include electromagnetism, gravitation, and nuclear forces, and, most relevant in the present context, neural energy—the internal energies in neurons [see Banks (2014, pp. 149, 203)]. Manifesting itself qualitatively at the level of the single neuron, this energy may yield an electrical discharge event; but manifesting itself at the level of a complex brain event—an event that is "somehow 'composed' of neurons firing in some kind of cluster" (Banks, 2014, p. 147)—this very same neural energy may yield the event that is a sensation of blue. This closes the apparent chasm between the experience of blue and the firing of a bunch of neurons: "the quality blue and the individual electrical discharges are just different and mutually exclusive manifestations of the same natural powers which we mistakenly see as belonging to totally different categories of event." (Banks, 2014, p. 164)

> monist picture, physics describes certain relations—namely, the physical ones—among the basic entities without capturing their intrinsic qualities, or those of the complexes of which they are parts. These latter features are revealed to us only in the case of our sensations, percepts, and other mental episodes. This suggests that the basic entities exhibit two fundamentally different kinds of aspects or properties: extrinsic physical relations and intrinsic mental qualities. But there's nothing properly neutral about either kind of feature, and neither is reducible to the other. At best, they are two radically different aspects of an underlying reality which, in itself, is neither mental nor physical.
>
> The theory of dual or double aspects is usually traced back to Spinoza ([Ethics, Demonstrated in Geometrical Order], 1677). The fundamental idea uniting the family of views under this label is that there is an underlying reality that we can grasp as mental or as physical, depending on the point of view from which we apprehend it. Each one of us can know their own brain under each of these aspects—via introspection and (scientific) observation. But the claim of the theory is quite general: everything there is is to be understood as consisting of an underlying reality that has these two aspects.
>
> Neutral monism and the dual-aspect theory share a central claim: there is an underlying reality that is neither mental nor physical. But that is where the agreement stops. Neutral monism has no room for the central feature of the dual-aspect theory: the mental and physical aspects, sides, or properties that characterize the underlying entities of dual-aspect theory. The neutral monist accepts the mental/physical distinction. But it resides at the level or groups of neutral entities. Grouped one way, the neutral entities constituting your brain are thoughts and feelings; grouped another way, they are atoms and neurons and lobes. Whether a given group of interrelated neutral entities counts as mental or physical depends on the causal-functional role this group occupies. But the entities themselves are free of intrinsically mental or physical aspects/sides/properties. Therein consists their neutrality.

Stubenberg and Wishon (2023) also conclude by stating that

> The rising interest in the history of analytic philosophy—especially the gradual rediscovery of Russell as a metaphysician and epistemologist—offers some hope for better understanding of traditional neutral monism. [...] And the idea to make abstract entities—information, structure, computation, mathematical reality—into the neutral basis of a metaphysical system is being actively developed by philosophers and scientists alike.

To me, it seems evident that some simple mathematical formulations can give some structure to the discussions. However, I suggest we should not restrict ourselves only to the systems of logic and quantifiers as the analytic philosophy seems to have done, but extend our reach to other mathematical formalisms, such as linear algebra and differential equations, too. My worry is that as there surely are important developments in digital humanities, for example, to me it seems as if the discussions there are confined to lag behind the technology development, such as when considering the large language models of late. Most scholars do not seem to have much clue how transformers in large language models work mechanically using linear algebra, for example, and it will be a huge problem going forward, I presume. The philosophers of the past were in a different position—most often they were the absolute highest experts in the formalisms they used, such as logic, but it seems not to be the case anymore with these new developments that are in dire need of philosophical analysis.

Now to be frank, I am not sure whether this section on neutral monism and dual aspect theories offered much in terms of conceptual tools to approach emergent multilevel modeling in nature. But I am determined that there really are analogies here—similarly to how we can wonder about the relation between neuronal populations and the mind, many other large collectives have these dual aspects, too. For example, diplomats can have quite a different feel how states and bureaucracies have "intentions" or "minds of their own", likewise for global geopolitics. Also corporate managers can attest to the reality of persistent corporate culture, or how the leaders perceive the competitive landscape as consisting of abstract entities and various market forces, aiming to capture and steer "the mind" of the corporation via vision and mission statements.[67] Consultants also have a feel for entire industries, and all these examples can be described as

---

[67] See also Daft and Weick (1984), Weick (1995), and related studies on sense making in organisations, supporting the idea of importance of modeling at all scales. But organization studies is a large field—the annual meeting of the Academy of Management attracts over 10 000 attendees from around the world (academics, students, practitioners), to over 1 500 in-person sessions, so understandably there is simply a massive amount of relevant discussion and modeling in the literature, not just these scholars and studies singled out in here.

both in individual or collectivist terms, the reality still being one monist substance, in some very fundamental sense.

I will end this section with a suggestive (or concerning, depending on one's viewpoint) quote from Stubenberg and Wishon (2023):

> There is a lively debate concerning the relationship between neutral monism, property dualism, and dual-aspect theory [...] The decision about these theories—whether they are identical, distinct but compatible, or incompatible rivals—is still out.

At some point, the confusions may be sustained also for the practical sake of keeping the discussion alive, which is only human, after all.[68] But in this present research endeavor, by contrast, I have been aiming towards some preliminary but specific formalisms that could perhaps converge and be of lasting value in some time frame—ultimately aiming to address the question of *who* (self, in personal terms), as some kind of a probing model of the very real (as we and our societies are already really here!) ubiquitous mind, logos, nous, in nature, unfolding at various levels.

### F. Universe of the mind and the semiosphere

When discussing the philosophy of mind, the philosophy of language is obviously related. Partly this is due to the concepts of language and mind being so intertwined. In Finnish, for example, the words for mind ("mieli") and language ("kieli") have only one letter difference, and this is not just a coincidence (or it actually literally is a correlation, on different temporal scales, which we could revisit at some point).[69]

With philosophy of language, we enter domains such as semiotics, where signs and meaning become central—at various points in history people have tried to build towards some kind of "science of meaning".[70] For starters, Umberto Eco writes in his introduction to Lotman's "Universe of the Mind" (Lotman, 2001 [1990], pp. ix–xi) (emphasis in the original)

> In the Sixties, Lotman stressed the usefulness

---

[68] For ideas linking Hegel, Bergson, Bohm, etc., see Seppälä (1995) on naturalistic creativity. But where are these ideas going? Is there some convergence or unification in sight, or only personal rediscovery (which is always of crucial importance, too) and reframing, otherwise oblivion?

[69] Observe also how the associations can be probed using a dictionary (that large language models exhibit, too):

*mind*: "the element of a person that enables them to be aware of the world and their experiences, to think, and to feel; the faculty of consciousness and thought.", "as the thoughts ran through his mind, he came to a conclusion". Related words: brain; intelligence; intellect; intellectual; capabilities; mental capacity; brains; brainpower; wits; wit; powers of reasoning; powers of comprehension; powers of thought; understanding; reasoning; judgment; sense; mentality; perception; head; imagination; subconscious; psyche; ego; sanity; mental balance; mental faculties; senses; wits; reason; reasoning; judgment; rationality; informal: gray matter; brainbox; brain cells; loaf; smarts; kop; marbles; rare:ratiocination

Compare also to suggested Finnish translations for "mind":
*mieli* (mind, heart, mood, spirit, soul, psyche),
*ajatukset* (mind, reflections, reflexions),
*sielu* (soul, spirit, mind, mastermind, inner man, psyche),
*järki* (reason, sense, wits, sanity, mind, intellect),
*ymmärrys* (understanding, comprehension, mind, wits, penetration),
*halu* (desire, wish, urge, craving, lust, mind),
*psyyke* (psyche, mind),
*asenne* (stance, approach, posture, mind),
*muisti* (memory, storage, store, recall, mind, retention),
*mielipide* (opinion, view, mind, stand, sense, verdict),
*tarkoitus* (purpose, intention, object, objective, meaning, mind)
...and "mind" as a verb:
*huolehtia* (worry, mind, trouble)

[70] For example, Ogden and Richards (1989 [1923], pp. xvii–xviii), writing 100 years ago, state in their preface that (emphasis mine):

"The practical importance of a science of Symbolism even in its present undeveloped form needs little emphasis. All the more elaborate forms of social and intellectual life are affected by changes in our attitude towards, and our use of, words. How words work is commonly regarded as a purely theoretical matter, of little interest to practical persons. It is true that the investigation must at times touch upon somewhat abstruse questions, but its disregard by practical persons is nevertheless short-sighted. The view that language works well enough as it is, can only be held by those who use it merely in such affairs as could be conducted without it—the business of the paper-boy or the butcher, for instance, where all that needs to be referred to can equally well be pointed at. None but those who shut their eyes to the hasty re-adaptation to totally new circumstances which the human race has during the last century been blindly endeavouring to achieve, can pretend that there is no need to examine critically *the most important of all the instruments of civilization*. New millions of participants in the control of general affairs must now attempt to form personal opinions upon matters which were once left to a few. At the same time the complexity of these matters has immensely increased. The old view that the only access to a subject is through prolonged study of it, has, if it be true, consequences for the immediate future which have not yet been faced. The alternative is to *raise the level of communication* through a direct study of its conditions, its dangers and its difficulties. The practical side of this undertaking is, if communication be taken in its widest sense, Education.

Convinced as they are of the urgency of a stricter examination of language from a point of view which is at present receiving no attention, the authors have preferred to publish this essay in its present form rather than to wait, perhaps indefinitely, until, in lives otherwise sufficiently occupied, enough moments of leisure had accumulated for it to be rewritten in a more complete and more systematized form. They are, they believe, better aware of its failings than most critics will suppose, and especially of those due to the peculiar difficulties which a fundamental criticism of language inevitably raises for the expositors thereof.

[...] the moment seems to have arrived when an effort to draw attention to Meaning may meet with support."

of the structural approach and the application of exact methods to the study of literature. That is, he remained faithful to Saussure's opposition of *langue* and *parole*, and to that proposed by Jakobson and Information Theory of code and message. In 1967, Lotman wrote an article on 'Exact Methods in Russian Literature Science' for the Italian journal *Strumenti critici*. This article repeated the positions already taken in his other writings and expounded some of the main principles of his research methods. These are outlined briefly below:

1. The opposition of exact sciences and humanistic sciences must be eliminated.

2. The study of literature, if carried out in a purely historical way, blends into the history of social thought.

3. The Russian Formalists of the Twenties had initiated the study of the 'techniques' of literary phenomena but it was now time to introduce into the study of literary texts the methods of linguistic structuralism, semiotics (and he was thinking here of Peirce too), of Information Theory, cybernetics and mathematical-statistical analysis.

4. Semiotic systems are *models* which explain the world in which we live (obviously, in explaining the world, they also construct it, and in this sense, even at this early stage, Lotman saw semiotics as a cognitive science). Among all these systems, language is the *primary modelling system* and we apprehend the world by means of the model which language offers. Myth, cultural rules, religion, the language of art and of science are *secondary modelling systems*. We must therefore also study these semiotic systems which, since they lead us to understand the world in a certain way, allow us to speak about it.

5. If texts represent models of the world, the set of texts which is the culture of a period is a secondary modelling system. It is thus necessary to attempt to define a *typology of cultures*, in order both to discover universal aspects common to all cultures and to identify the specific systems which represent the 'language' of Medieval culture or the 'language' of Renaissance culture.

6. When a culture is analyzed as a code or system (as also happens with natural languages), the processes of use are richer and less predictable that the semiotic model which explains them. Reconstructing the code of a culture does not mean explaining all the phenomena of that culture, but rather allows us to explain *why* that culture has produced those phenomena.

Lotman realized, however, that seeing a text as a message elaborated on the basis of a linguistic code is by no means the same as seeing a text (or a culture as a set of texts) as a code. For Lotman was and is aware of the fact that no historical period has a sole cultural code (even if the construction of a model-code can be a useful abstraction) and that in any culture there exist simultaneously various codes. It seems to me that in attempting to deal with this problem, Lotman is moving beyond structuralist dogmatism and offering a more complex and articulated approach. Faced with the rigidity of the structuralist opposition between code and message, Lotman introduces, even within the same culture, a difference between grammatical learning and textual learning.

Later we will find how the concepts of "primary modeling system" and "secondary modeling system" have been extended and developed further, by Sebeok and Danesi (2000), for example, where the mentioned modeling systems are displaced to secondary and third tier, respectively, by identifying the actual primary modeling system as based on sense experience or sensorium—reminiscent of factor analysis in nature.

In his introduction, Eco also highlights the definition of "semiosphere", which seems useful to reiterate here (Lotman, 2001 [1990], p. xii) (emphasis in the original)

> ...imagine a museum hall where exhibits from different periods are on display, along with inscriptions in known and unknown languages, and instructions for decoding them; there are also the explanations composed by the museum staff, plans for tours and rules for the behaviour of the visitors. Imagine also in this hall tour-leaders and visitors and imagine all this as a single mechanism (which *in a certain sense it is*). This is an image of the semiosphere. Then we have to remember that all elements of the semiosphere are in dynamic, not static, correlations whose terms are constantly changing. We notice this specially at traditional moments which have come down to us from the past.

Further into the work, at the beginning of a chapter titled "Rhetoric as a mechanism for meaning-generation", Lotman (2001 [1990], pp. 36–37) visions, that

> Human consciousness is heterogeneous. A minimal thinking apparatus must include at least two differently constructed systems to exchange the information they each have worked out. Studies carried out on the specific functioning of the large hemispheres of the human brain have revealed a profound analogy between it and the organization of culture as a collective intellect. In both cases we find there are at least two essentially different ways of reflecting the world and working out new information, and that in both cases there are complex mechanisms for exchanging texts between these systems. In both cases we observe a generally analogous structure: within one consciousness there are as it were two consciousnesses. The one operates as a discrete system of coding and forms texts which come together like linear chains of linked segments. In this system the basic bearer of meaning is the segment (= the sign), while the chain of segments (= the text) is secondary, its meaning being derived from the meaning of the signs. In the second system the text is primary, being the bearer of the basic meaning. This text is not discrete but continuous. Its meaning is organized neither in a linear nor in a temporal sequence, but is 'washed over' the n-dimension semantic space of the given text (the canvas of a picture, the space of a stage, of a screen, a ritual, of social behaviour or of a dream). In texts of this type the text is the bearer of the meaning. We may have difficulty in isolating its component signs, and this task smacks of artificiality.
>
> Thus both the individual, and the collective consciousness, contain two types of text-generator: one is founded on discreteness, the other is continuous. In spite of the fact that each of these mechanisms has a self-contained structure, there is a constant exchange of texts and messages between them. This exchange takes the form of a semantic translation. But an accurate translation presupposes that mutually equivalent relationships have already been established between the units of the two systems, as a result of which one system can be represented in the other.

These quotations are getting ridiculously long, but I will continue typing them here, for now. I think that only few readers will actually delve into original texts, so maybe including too much here could still have some positive secondary effects, and typing them out may increase my comprehension and recall, too.

See also Ch. 9, "The notion of boundary", and Ch. 10, "Dialogue mechanisms", in the same work.

In the appendix I included a long quote from Lotman about the creative process of writing, which I found fascinating. In some very visceral sense, I am reminded of the expanding $S^3(r_0)$ geometry functioning as a similar generating symbol when drafting (Lievonen, 2023). Here I am utilizing Eq. (1), accompanied with a mental image of a specific variant of an energy functional that has not yet been written here, to function as a central connective point ("kiintopiste" in Finnish), however feeble and fragmentary at this point.

Lotman (2001 [1990], p. 273) concludes his book with the following:

> The individual human intellect does not have a monopoly in the work of thinking. Semiotic systems, both separately and together as the integrated unity of the semiosphere, both synchronically and in all the depths of historical memory, carry out intellectual operations, preserve, rework and increase the store of information. Thought is within us, but we are within thought, just as language is something engendered by our minds and directly dependent on the mechanisms of the brain, and we are with language. And unless we were immersed in language, our brain could not engender it (and vice versa: if our brain were not capable of generating language, we would not be immersed in it). The same with thought: it is both something engendered by the human brain and something surrounding us without which intellectual generation would be impossible. And finally the spatial image of the world is both within us and without us.
>
> We are both a part and a likeness of a vast intellectual mechanism. Hence the difficulties but also the importance of the kind of research we are doing. The emergent synthesis becomes ever clearer: whether we are studying the structure of the literary text, the functional asymmetry of the hemispheres of the brain, the problems of oral speech or of deaf and dumb language, the advertisements of our modern age or the religious ideas of archaic cultures—we find the different mechanisms of the single intellectual life of humanity. We are within it, but it—all of it—is within us. We are at the same time like *matryoshkas*, and participants in an endless number of dialogues, and the likeness of *everything*, and 'the other' both for other people and for ourselves; we are both a planet in

> the intellectual galaxy, and the image of its universum.

With this, I am reminded why I gravitated towards mathematical thinking years ago. I have hundreds of good books like these, but no hope of reading them all. I am under no illusion that there would be audiences with such diversity, either. At the same time, large language models will be able to read, cross-reference and summarize vast works quickly, and in the future also develop ideas further. With mathematical structures, there is hope for condensing some ideas into shared structures, so that the eternal "now instead of 20 ideas, we have 21 competing ideas" problem could be alleviated, perhaps, if we really even want it. But I will continue exploring the vast worlds of literature here for a while, still.

### G. The concept of a model: forms and their referents

One of the main reasons for models to exist in the first place is providing reduction and abstraction, thus facilitating application of knowledge, as building blocks for synthesis of new realities. Danesi (2017, p. 1496) motivates his studies with Thomas A. Sebeok (1920–2001), studying inherent modeling in nature, with the following:

> Model-making typifies all aspects of human intellectual and social life. Miniature models, blueprints, maps, scientific diagrams, and the like are so common that one hardly ever takes notice of the fact that they are strategies of compressed information. The intriguing question that this reformulation of basic sign theory invariably raises is the following one: What is the function of modeling in life? This question begs, in turn, a whole series of related ones: How is human modeling similar to, or different from, modeling systems in other species? What is the relation between modeling and knowing?

In their parlance, a model is a relation between a *referent* and some *form* or *pattern*, which has been imagined or made externally (though some physical medium) to stand for the referent, other than itself. A form is thus a generalized sign, and it may have various physical manifestations, such as some scribbles on paper, or sounds in the air, or neuronal patterns in the brain, or biochemical patterns among some cell groups.[71]

The distinction between instinctual and intentional modeling is crucial here—for humans, the intentions for creating representations are much more clear than for cell groups, for example, which are usually conceptualized as instinctual modeling. The important realization here is that in order for something to be known and remembered, it must be assigned some form, and this necessity of storing state applies to any system in nature, at any level, not just humans.

Abstractly, forms of macro scale phenomena could include perhaps population centers on a countryside, historically indicating availability of natural resources or ease of transportation, or perhaps companies operating in some market segment, indicative of market opportunities. Really the concept of forms and patterns as models is extremely flexible and general [even after formalizing it with mathematics later, similar as in Eq. (1)], and it is seminal to keep in mind that forms are not in one-to-one relation to their referents—almost any physical formation can be interpreted in multiple, overlapping ways, of course, but they are not completely arbitrary, either.

One crucial aspect of modeling is that if the referent changes, then the form or pattern should more or less change, too (and vice versa), but the model (the relation between them) should be quite invariant, to be considered a model. So there are clearly various spatial and temporal scales involved, suggesting and even necessitating using mathematics to keep the thinking clear.

In practice, Sebeok and Danesi, informed by the conceptual resources of semiotics, have developed quite a large vocabulary for modeling phenomena, such as "(1) *signs* (words, gestures, and so on), (2) *texts* (stories, theories, and so on); (3) *codes* (language, music, and so on); and (4) *figural assemblages* (metaphors, metonyms, and so on) [which] are all designed to model something in terms of the $X$ stands for $Y$ relation." (Danesi, 2017, p. 1497). More generally, the above types of forms can be

---

[71] Form, as a general word, is seen in many contexts. See, for example, (Spencer Brown, 1969), that influenced Luhmann when developing his theories on social systems. As one example, Bourdieu (1977, p. 198–199) notes that (emphasis in the original):

> The language of *form*, taken in the sense of "*structure of becoming*" which it has in musical theory (e.g. the suite, or sonata form) would no doubt be more appropriate than the language of logical structure, to describe the logically but also chronologically articulated sequences of a musical composition, a dance, or any temporally structured practice. It is significant that the only way which R. Jacobson and C. Lévi-Strauss ("'Les chats' de Charles Baudelaire", L'Homme, *2*, 1 (Jan.-April 1962), pp. 5–21) find to explain the movement from structure to form, and the experience of form, that is to say, to poetic and musical pleasure, is to invoke *frustrated expectation*, which objectivist analysis can describe only by bringing together in simultaneity, in the form of a set of themes linked by relations of logical transformation (e.g. the movement from the metaphorical form, the scientist, the lover, the cat, to metonymic form, the cat), the essentially polythetic (in Husserl's sense) structure of a poetic discourse which *in practice* is communicated only in and through time. In reality, as temporal structures, musical or poetical forms can only be understood inasmuch as they perform *expressive functions* of various types.

conceptualized as *singularized*, *composite*, *cohesive*, and *connective* forms. For example, singular forms are typically categorized as symptoms, signals, icons, indexes, symbols, and names. As an other example, codes (as cohesive forms), "can be defined as a system that allows for the representation of referents perceived to share common traits. Codes consist of interrelated forms, making up a cohesive whole, which can be deployed to model types of phenomena in specific ways (for example, mammals of a certain type)." (Danesi, 2017, p. 1503)[72] There are also dimensions of vocal and nonvocal forms, verbal and nonverbal forms, witting and unwitting forms—the formation and dissolution of modeling systems happening "by exposure to appropriate input in context and subject to change [...] over time."

Sebeok and Danesi have tried to summarize their notions of this ubiquitous modeling in nature, *semiosis*, into a few basic principles, such as the structuralist principle, claiming that elemental structural properties such as paradigmaticity (minimal differentiation property), syntagmaticity (combinatory property), analogy (equivalence property), synchronicity, diachronicity, and signification, characterize all forms (Danesi, 2017, p. 1500):[73]

> Synchronicity refers to the fact that forms are constructed at a given point in time for some particular purpose or function; and diachronicity to the fact that they undergo change over time. The change is not random, but rather, governed by both structural tendencies characterizing the code to which forms belong and external contextual (social, situational, and so on) influences.

Forms occurring in nature or culture are typically distinguished as natural or artificial, depending on there being human intentions present. Danesi (2017, p. 1498) explains that "These serve many functions in human life. They allow people to model patterns in things; they act as predictive guides or plans for taking actions; they serve as exemplars of specific kinds of phenomena; and the list could go on and on."[74]

Danesi (2017, p. 1505) (emphasis in the original) concludes by stating that

> there is a distinction among semiosis, modeling, and representation: *semiosis* is the neurobiological capacity to produce forms (signs, texts, and so on), *modeling* is the channeling of the *semiosic* capacity towards a *representation* of some referent (the actual act of creating a form). And representation is the strategy of actually realized a form [*sic*]. Modeling reveals how the brain carries out its work of transforming sensory input into internal forms of thinking and external forms of representation: a specific external model is thus a "cognitive trace" to the form a concept assumes in the mind, and since concepts depend on how they are modeled it has been argued [...] that the form that knowledge takes depends on the type of modeling used.

We are reaching the limits of useful prose here (it is already getting a bit too long-winded with these definitions), and with this, we are moving towards *using mathematical modeling to model modeling itself*—but until we get there, let Danesi (2017, p. 1496) summarize the concept of a model:

> As used commonly, the term model exemplifies, actually, many of the aspects of Sebeok's definition, given that it is used as a synonym for theory (Black 1962, ["Models and metaphors"]) or to indicate an analogy (a set of billiards balls in casual [*sic*] movement can be employed as a model for

[72] It is debatable that cohesive modeling systems cover many kinds of systems in nature, see, for example, (Danesi, 2017, p. 1503)

> the body's immune system is a natural code consisting of interacting organs, tissues, cells, and cell products such as antibodies which not only neutralize potentially pathogenic organisms or substances, but also allow one to become aware of the difference between Self and "non-Self" (the external world). It is the code that undergirds the symptomatology of diseases.

[73] More specifically, their principles are (Danesi, 2017, p. 1499):

- Representation is the end-result of producing forms of various types to model referents (the modeling principle).
- Knowledge is indistinguishable from the forms used to encode referents (the representational principle).
- Modeling unfolds on three levels or dimensions, called primary, secondary, and tertiary (the dimensionality principle).
- Complex (abstract) forms are derivatives of simpler (more concrete) ones (the extensionality principle).
- Codes and their referential domains are interconnected to each other (the interconnectedness principle).
- All models and their forms display the same pattern of structural properties (the structuralist principle).

For details, please consult (Sebeok and Danesi, 2000), (Danesi, 2017), or some other references.

[74] C. S. Peirce's *firstness*, *secondness*, and *thirdness* is related, as the primary modeling system is thought to be sensory or perceptual, thus indicating iconicity being of fundamental importance—a kind of natural osmosis of the most relevant properties of referents, 'flowing' or 'molding' into the model. This is quite difficult to describe without mathematics!

The higher (more abstract) modeling systems enable indexical (indicational) forms (such as spatial pointing), and also symbolic forms—learning to use a culture-specific name to refer to an object by name, for example, or acquiring and utilizing some abstract systems of representation.

> gas molecule movement). The term has also been used to define an exemplary paradigm to be followed (Da Vinci's "Vitruvian Man" constitutes a model of the perfect proportions of the human body), or one to be discarded (the marketplace model of production is harmful to the environment). In each case, however, the term implies some formal structure as a point-of-reference. The term was used by logician Alfred Tarski [in "Logic, semantics, metamathematics"] (1933) as referring to a representation of a mathematical or logical pattern. Even an equation such as the Pythagorean one, $c^2 = a^2 + b^2$, is a model of, initially, the relation of the sides of a right triangle. It also became a model for testing relations among numbers, known as Pythagorean triples. [...] Models possess what can be called a "4-E representational structure"—economy, efficiency, effectiveness, and ergonomics.

### H. Mathematical models as ultimate reductions

I once corresponded with a very capable mathematician, who wrote (among other words of encouragement):

> One needs to remember that everything is conventions in mathematics, and often conflicting conventions to add to the confusion.
>
> [...]
>
> Everything is convention and everyone is FREE to do what they want, but hopefully something that makes mathematical sense to themselves if to nobody else!

This was (and is) very inspirational, and suggests using mathematics as a *tool for thinking*—aiming for common understanding, but promising at the very least personal insights, which potentially enable more profound work.

Many mathematical structures are extremely minimal, yet very general. There is a certain quality to mathematical syntax (especially associative algebras, with addition, multiplication, and grouping with parenthesis familiar from basic education) that can represent some fundamental ideas very succinctly.

#### 1. Linearity

For example, linearity is one of the most important concepts in mathematics, that is often misunderstood as "just a line", "too simple as a model", "nothing in nature is linear", etc. But actually, linearity is of immense value for scalability of almost anything, as by definition, for any linear function $f$,

$$f(a_1x_1 + a_2x_2 + \cdots) = a_1f(x_1) + a_2f(x_2) + f(\cdots), \quad (17)$$

which facilitates manipulating large amounts of quantities efficiently as aggregates. There are application domains where there are millions and millions of elements (or perhaps some convergent series with an unbounded number of terms).

In Eq. (17), one can also contemplate about the usage of the equivalence symbol "=", reminiscent of a steelyard in balance, where both sides need to have the same "weight" for the arm to be exactly horizontal. This notion of equivalence (or isomorphism for some structures) that stays valid if one operates on both sides with the same operations (such as adding terms or multiplying by some factors), combined with the grouping of symbols by nesting parenthesis, is the main tool for making abstractions and reasoning with them.

Even though many processes by themselves seem far from linear, it could be argued that the only really scalable, proper models, are linear at their core, so we should always fundamentally try to aim for linearity to ease reasoning, understanding, sense, and overall usefulness. On a large scale, many relations ought to be as simple as possible due to averaging effects and general inertia associated with massive scales, which also suggests linearity—and for many natural systems, some kind of pseudolinearity (in terms of "stronger" and "weaker", or perhaps positive, negative, and neutral, suggesting using real numbers) are the simplest first structures stable enough to develop in any case. The multitude of conventions for expressing the exact same linear products (series formulas, Einstein summation, dot products, conjugate products, linear projections, tensor contractions, bra-ket notation, etc.) testifies on the importance of linearity on almost everything in mathematics.

We will be using linear algebra, especially matrix calculus, in reasoning with multidimensional mathematical structures. For the adventurous, see Table I on page 11, and also (Laue *et al.*, 2018, 2020).[75] But let us first try to build some bridges from algebraic relations and scalar mathematics to the worlds of matrices, by studying some very simple mathematics (which potentially could be indicative of something fundamental present).

#### 2. Commutativity

In ordinary high school algebra, there is a remarkable degree of freedom "hidden in plain sight", that I am wondering when it will be introduced to school curricula,

[75] On video, see Laue, "Computing derivatives of matrix and tensor expressions": `https://youtu.be/IbTRRlPZwgc`. Their very useful tool is located at `https://www.matrixcalculus.org`.

and at what level. I am referring to anti-commutativity (where $ab = -ba$) or its variants, where one needs to take care of the ordering of multiplicative factors (i.e. one cannot change the order of symbols in terms at will, operations other than addition do not necessarily commute).[76] Often this leads to natural reading of mathematics from right to left, when using matrices in linear algebra as projections between high-dimensional spaces, for example. It is possible to encode the direction of interactions (such as spatial handedness, or perhaps even suggestions for temporal causality) in the ordering of the products.

The Pythagorean theorem, $a^2 + b^2 = c^2$, already mentioned on the previous page, can be utilized in quite a neat way to introduce anti-commutative algebras and the abstract concept of orthogonality. I first read about this in Sobczyk (2019), and immediately wondered what would have Euclid and his contemporaries thought about the derivation presented:

$$\begin{aligned}(a+b)^2 &= (a+b)(a+b)\\ &= (a+b)a + (a+b)b\\ &= a^2 + ba + ab + b^2\\ &= a^2 + ba - ba + b^2\\ &= a^2 + b^2\\ &= |a|^2 + |b|^2\\ &= |c|^2\\ &= c^2, \end{aligned} \tag{19}$$

where in addition to anticommutation $ab = -ba$, we have used the property that the square of an ordinary vector is simply its length squared, as

$$v^2 = (|v|\hat{v})^2 = |v|^2\hat{v}^2 = |v|^2, \tag{20}$$

when unit vectors $\hat{v}^2 = 1$, by definition in this particular case.

Clearly, anticommutation of $a$ and $b$ is needed for the Pythagorean theorem to be valid here. Thus, for orthogonal vectors, their sum squared is equal to the sum of their squares, as the cross-terms vanish.

Conversely, commutativity, $ab = ba$, denotes parallel vectors, demonstrating how ordinary high-school algebra, where multiplication is commutative, is actually a special case of a larger class of non-commutative algebras.[77] We will return to these developments in a moment.

### 3. Conciseness

Other minimal formulas, that exhibit some fundamental ideas that have been relevant across the centuries, include the quadratic equation

$$f(x) = ax^2 + bx + c = a(x - x_1)(x - x_2) = 0, \tag{21}$$

where $x_1$ and $x_2$ are the roots of the equation, determined by the quadratic formula

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}. \tag{22}$$

It is a famous example of simple, reduced formula that is studied early in mathematics education,[78] leading also to the discussion about imaginary numbers, as the discriminant ($b^2 - 4ac$) can be negative, warranting introducing an element $i^2 = -1$.

Minimalistic polynomial formulas present themselves as useful example problems for analysis, as derivatives are straightforward to calculate, thus most people have been exposed to the various exercises involving zeros of derivatives (as potential stable points, depending on higher-order derivatives).[79] Simplicity of formulas is essential for any kind of larger scale adoption.[80]

---

[76] I am under impression that the first ones to really ponder about noncommutative algebras were K. Weierstrass, W. Hamilton, W. Clifford, and their contemporaries in the 19th century. Later commutativity and anticommutativity became fundamental in quantum mechanics, as exemplified by Max Born (1882–1970) having an inscription in his and his wife Hedwig's grave at the Göttingen cemetery:

$$pq - qp = \frac{h}{2\pi i}. \tag{18}$$

[77] For higher grade objects the meaning of commutativity and anti-commutativity alternates, i.e., for bivector $\mathbf{e}_{12}$, the parallel in-plane vector $\mathbf{e}_2$ anti-commutes ($\mathbf{e}_{12}\mathbf{e}_2 = \mathbf{e}_1\mathbf{e}_2\mathbf{e}_2 = -\mathbf{e}_2\mathbf{e}_1\mathbf{e}_2 = -\mathbf{e}_2\mathbf{e}_{12}$), and the orthogonal vector $\mathbf{e}_3$ commutes ($\mathbf{e}_{12}\mathbf{e}_3 = \mathbf{e}_1\mathbf{e}_2\mathbf{e}_3 = -\mathbf{e}_1\mathbf{e}_3\mathbf{e}_2 = \mathbf{e}_3\mathbf{e}_1\mathbf{e}_2 = \mathbf{e}_3\mathbf{e}_{12}$). But understanding these syntaxes is not a prerequisite, as they are displayed here only for purposes of illustration.

[78] The solution is usually easiest to see by completing the square, illustrating the idea of operations maintaining the equivalence—there is a "flow of symbols to the other side", that leaves the preferred structure (here $x$ as a solution) visibly exposed for instant perception:

$$\begin{aligned} ax^2 + bx + c &= 0\\ x^2 + bx/a + c/a &= 0\\ x^2 + bx/a &= -c/a\\ x^2 + bx/a + (b/2a)^2 &= (b/2a)^2 - c/a\\ (x + b/2a)^2 &= (b^2 - 4ac)/(2a)^2\\ x + b/2a &= \pm\sqrt{b^2 - 4ac}/2a\\ x &= \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}. \end{aligned} \tag{23}$$

[79] Differentiating left side of Eq. (21)

$$\begin{aligned} \frac{d}{dx}\left(ax^2 + bx + c\right) &= \frac{d}{dx}\left[a(x - x_1)(x - x_2)\right]\\ 2ax + b &= a[(x - x_2) + (x - x_1)]\\ 2ax + b &= 2ax - a(x_1 + x_2)\\ b/2a &= -(x_1 + x_2)/2\\ b/a + x_1 + x_2 &= 0, \end{aligned} \tag{24}$$

Matrix derivatives, as multidimensional generalizations of these scalar derivatives, will be very helpful later. They could be generalized further using tensor calculus, but I consider being able to validate the results important for the reader, so with the tools available, I hope that ideas can be communicated using matrices of order two [two indexes, similar to a 2D-array, as in Eq. (1)].

Laue *et al.* (2020, p. 4529) defines Fréchet derivative $D$ of a generic tensor function $f$ at tensor $x$, as satisfying

$$D : \lim_{h\to 0} \frac{\|f(x+h) - f(x) - D \circ h\|}{\|h\|} = 0, \qquad (28)$$

where $\circ$ is an inner tensor product and the norm is the Frobenius norm.[81] While this could be useful in the future here, for complicated function compositions we will be underhanded with regards to the machine learning frameworks in any case, where these kind of multidimensional differentiations are made automatically in the background without anyone inspecting the gradients—so here we will aim to use their online tool

`https://www.matrixcalculus.org`

to validate the derivations and get insight into various mathematical structures. But before we go there, let us still continue our exposition of seemingly fundamental mathematical structures for a while.

## 4. Oscillation

On Erwin (1887–1961) and Annemarie (1896–1965) Schrödinger's tombstone near St. Oswald's church in Alpbach village (Tyrol, Austria), there is a prominent inscription

$$i\hbar\dot{\psi} = H\psi. \qquad (29)$$

which illustrates some of the symmetries present in these kind of formulas, as $(x_1 + x_2)/2$ is simply the midpoint (average) of the roots.

[80] Other examples of concise formulas, where semantics is given by physics, include:

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = \kappa T_{\mu\nu} \qquad (25)$$

$$eV = h\nu - A \qquad (26)$$

$$E = mc^2, \qquad (27)$$

which are displayed on Albert Einstein Memorial in Washington, D.C., and refer to the general theory of relativity, the photoelectric effect, and the equivalence of energy and matter, respectively.

[81] Note that in some cases directional derivative, such as Gateaux derivative, is needed, and then the displaced argument $x + h$ in Eq. (28) is defined as $x + hv$, where $v$ is the direction, the magnitude of which affects inversely to the magnitude of the derivative. I am not yet planning to go there, and in any case I do not have expertise to deal with infinite-dimensional spaces and different norms, for example.

It is one of the most important equations in fundamental physics, bearing his name. As there are so few symbols present, somebody could have studied it (or written it accidentally down) already long ago—but of course the relations between the symbols are not evident without knowing about the used syntax, and in physics, the semantics are grounded to other related models and some central observations, giving it meaning and importance there.[82]

In a scalar form (which is only structurally relevant here) the solution to Eq. (29) attains the form

$$\psi = e^{-iHt/\hbar}\psi_0, \qquad (30)$$

where $\psi_0$ is a constant, determined from boundary conditions, such as the initial condition at $t = 0$. One can validate the solution in Eq. (30) by inserting it back to the left side of Eq. (29):

$$\begin{aligned} i\hbar\dot{\psi} &= i\hbar\frac{d}{dt}\left(e^{-iHt/\hbar}\psi_0\right) \\ &= i\hbar\frac{d}{dt}\left(e^{-iHt/\hbar}\right)\psi_0 \\ &= i\hbar\frac{d}{dt}(-iHt/\hbar)\, e^{-iHt/\hbar}\psi_0 \\ &= i\hbar\,(-iH/\hbar)\, e^{-iHt/\hbar}\psi_0 \\ &= He^{-iHt/\hbar}\psi_0 \\ &= H\psi. \end{aligned} \qquad (31)$$

When $H$ is similar to a generalized real number (for example, a square matrix with real eigenvalues, such as some Hermitian matrix $H = H^\dagger$), the solution in Eq. (30) is an oscillation (in the direction of each eigenvector of $H$ or in some spanned eigenspace if some eigenvalues are equal), as can also be seen from the series expansion

$$\begin{aligned} e^{i\phi} &= 1 + i\phi + (i\phi)^2/2! + (i\phi)^3/3! + \cdots \\ &= 1 + i\phi - \phi^2/2! - i\phi^3/3! + \cdots \\ &= (1 - \phi^2/2! + \cdots) + i(\phi - \phi^3/3! + \cdots) \\ &= \cos\phi + i\sin\phi \\ &= \cos(-\phi) - i\sin(-\phi), \end{aligned} \qquad (32)$$

where $\phi = -Ht/\hbar$, in this case. Clearly, the frequency of oscillation is $f = H/h$ (or when in matrix form, determined by the eigenvalues $\lambda_j$ of $H$, where $Hx_j = \lambda_j x_j$, or more generally $HX = X\Lambda$, where $\Lambda$ is diagonal). More generally, treating symbol $i$ as any suitable entity that squares to $-1$ (and commutes properly with the base field), it is possible to oscillate between distributed patterns that square to $-1$ (generalizing imaginary numbers)

[82] Note that the overdot in Eq. (29) is a shorthand for the time derivative, $d/dt$.

and localized diagonal matrices (identity matrices generalizing scalars).

"Solving" for $\psi$ in Eq. (30) by dividing with $H$ from the left (assuming invertibility of $H$), results in

$$i\hbar H^{-1}\dot{\psi} = \psi, \tag{33}$$

which can be inserted back to Eq. (30), exposing the second derivative, essentially

$$\begin{aligned} i\hbar\frac{d}{dt}\psi &= H\psi \\ i\hbar\frac{d}{dt}(i\hbar H^{-1}\dot{\psi}) &= H\psi \\ i^2\hbar^2 H^{-1}\frac{d}{dt}(\dot{\psi}) &= H\psi \\ i^2\hbar^2 H^{-1}\ddot{\psi} &= H\psi \\ i^2\hbar^2\ddot{\psi} &= H^2\psi \\ \hbar^2\ddot{\psi} &= -H^2\psi, \end{aligned} \tag{34}$$

where the linearity and commutativity of the derivative (in this particular contrived case) is used to rearrange the abstractions. The result is a well-known formula for harmonic oscillation, where the acceleration $d^2/dt^2$ (a kind of curvature in time) is negatively proportional to the displacement, and the imaginary unit $i$ has folded away.

This result is more fundamental than it perhaps looks at a first glance. The plain exponential function in Eqs. (30) and (32), as a solution to Eq. (29) should perhaps be more often appreciated. In some sense, the exponential function defines *what oscillation is*, and the exponential function, in turn, could be defined as the solution to the differential equation, as they are equivalent (up to a proportionality constant).

Thus in a very fundamental sense (as we are allowed to ponder about metaphysics here), it may be that overall, people and cultures have been able to even ponder about oscillating phenomena due to these structures existing. I mean, it is conceivable that similarly to there having been a time where there were no human words, these kind of relations could tell us something about the phenomena itself, ontologically (as opposed to models always being just epistemic abstractions of the phenomena). In some cases, one could thus invert the reductive thinking about models—philosophically speaking, at least.

Those infinite echoes in Eq. (32), where odd and even terms result in the overall balance and function, can really be wondered about as mysterious, even though the relations themselves are exactly and unambiguously defined as Taylor series. It is illustrative to plot how the terms add up, and inspect how the inverse of the factorial wears off. Note also how the terms alternate and fit together, indicative of a coherent whole, when taking a derivative.

Earlier in this working paper, around footnote 30, there were also some wondering about oscillations. To me, these kind of fundamental structures presented in this section seem important for these discussions, as there are only a relatively few ways some collection of mathematical symbols could be organized as a minimal formula, and if successful, they could work as instructional "boundary objects", bridging the various domains of inquiry.

## 5. Coherence

Consciousness, and awareness, for example, are such phenomena that could benefit from some simple models exhibiting some kind of coherence and change. By that I mean that simply by introspection, anyone can experience moments of waking up, falling asleep, and the various ways in which day dreaming and other drifting of consciousness can happen during the day (perhaps even noting some subdued polyphony of selves and spectra of identities, active at any time), that would warrant something akin to smooth evolution of averages of ensembles, arriving at relative coherence, to be able to discuss and deepen our scientific understanding of such phenomena.[83] It is also suggestive that the experience of fainting, as a kind of loss of consciousness, is in Finnish "pyörtyä" (similar to swirl, whirl, vortex, spin, rotate), where some kind of a phase-locked coherence starts to fall apart, resulting in multi-dimensional rotations that many people have experienced at some points in their lives.

See also the various time-frequency representations, such as the Wigner distribution function,[84]

$$W_x(t,f) = \int_{-\infty}^{\infty} x^*\left(t+\frac{\tau}{2}\right) x\left(t-\frac{\tau}{2}\right) e^{-2\pi i\tau f}\, d\tau, \tag{35}$$

that are based on correlations—here as expanding oscillations from time $t$, towards the future $t+\tau$ and the past $t-\tau$, at each frequency $f$, aggregated together by integrating out their complex conjugate products.

From this reductionist perspective, the main relevance in applying the various approaches of quantum mechanics to the philosophy of mind, for example, may be rather due to these kind of idealized model structures being useful for discussing very complex phenomena, rather than

[83] Compare to calculating statistical moments by using exponentials, averages, and differentiation, in Eq. (3) on page 3, for example, that could offer some ideas towards combining movement, "statistical fuzzyness", and exact clarity, that is needed. See also various measures of coherence, such as utilizing cross-spectral density (which is based on averaging conjugate products between frequency presentations, which are themselves based on aggregating sums using oscillating exponentials as filters), https://en.wikipedia.org/wiki/Spectral_density#Cross-spectral_density.

[84] See, for example https://en.wikipedia.org/wiki/Wigner_distribution_function and https://en.wikipedia.org/wiki/Wigner_quasiprobability_distribution.

the actual quantum phenomena, such as various vector potentials, somehow protruding or seeping through from the Planck-scale world of quantum mechanics to the level of the mind.

However, the model structures presented so far, still seem to lack (in my mind) some kind of possibility for different perspectives, goal-orientation, learning, also concentration and forgetting, that seem essential for modeling complex phenomena such as the mind. We will return to these, hopefully.

### 6. Aggregation

We would really need to discuss here some fundamentals of Fourier analysis (and Laplace transforms), as the utility of the exponential function is so nicely apparent there:

$$\mathcal{L}\{f\}(s) = \int_0^\infty f(t)e^{-st}\,dt = \int_0^\infty f(t)e^{-\sigma t}e^{-i\omega t}\,dt, \quad (36)$$

where $s = \sigma + i\omega$ combines both exponential dampening/amplification by $\sigma$ and oscillation by angular frequency $\omega$. One cannot get much simpler, mathematically, than taking a product with an oscillating function $e^{-i\omega t}$, and integrating (summing) it out to a number, corresponding to the correlation or inner product at frequency $\omega = 2\pi f$. There are complications, of course, as the integration could be done unilaterally (as in the above) or biliterally from $-\infty$ to $\infty$, and in practice, procedures are discretized which affect the results (and necessitate using different terminologies).[85]

Mellin's inverse formula enables deriving the original function $f(t)$ by integrating along a line (with a suitable $\gamma \in \mathbb{R}$),

$$\mathcal{L}^{-1}\{F(s)\}(t) = \frac{1}{2\pi i}\lim_{T\to\infty}\int_{\gamma-iT}^{\gamma+iT} e^{st}F(s)\,ds \quad (38)$$

$$= \lim_{k\to\infty}\frac{(-1)^k}{k!}\left(\frac{k}{t}\right)^{k+1}F^{(k)}\left(\frac{k}{t}\right), \quad (39)$$

where the second equation is Post's inversion formula, where the poles of $F(s)$ need not be known, but which necessitates taking arbitrarily high orders of derivatives. In practice, as the Laplace transform is linear, functions are usually decomposed into elementary functions and transformed using precalculated tables.

These kind of integral transforms allow transforming between multiplications and convolutions, for example, and also some very important properties, such as scaling and translation, which are apparent in vision, too, attain such simple mathematical forms there. I mean that as by introspection, one can evidently attend to different features of one's body, for example, quite easily, and also concentrate the attention to quite minute details (zooming in, so to speak) without discrete thresholds, this suggests that the simplest mathematical structures would be present in the frequency space, where most of the peculiarities of different phenomena in nature seem to be happening already. For now, I collected some features of the Laplace transform [Eq. (36)], that can be seen as a generalization of the Fourier transform, both widely used in the sciences, to Table II. At this point, it can be considered as a reminder of possibilities for the future studies.[86]

---

[85] It is also interesting that for periodic $f(t+T) = f(t)$,

$$\mathcal{L}\{f\}(s) = \frac{1}{(1-e^{-Ts})}\int_0^T f(t)\,e^{-st}\,dt. \quad (37)$$

[86] I have personally experimented with zooming, scaling, and rotations of 2D-images using modified Fourier transform matrices, the details of which I would need to collect here later, as they seem to be interesting (kind of factoring various operations to linearized "fieldlike" parameters). If I remember correctly, I created a $w \times w$ DCT-like matrix $D$, with elements as $\exp[i2\pi f(j-w/2)(k-w/2)]$, where $f = 1/wc$ and $c$ is the scale parameter (implementing zooming, when smaller or larger than unity). The matrix $D$ thus consists of slow waves, centering on the center of the matrix. Their phase could be seen to range linearly from $-\pi$ to $\pi$ (but even/odd dimensionalities may be important, here $w$ was even, affecting how indexes wrap around). Then, $DX$ is Fourier-transform like projection of the columns on $X$, and $(DX^T)^T = XD^T = XD$ is projection of the rows (as D is symmetric by this construction, not Hermitian). $DXD$ is then a projection of both rows and columns, which is nice. Inverse transform is got simply by conjugating the phases in $D$ (if $c = 1$), and if one has normalized the magnitudes of the elements by $1/\sqrt{w}$, the transform is unitary. Also the translations are easy, as one simply adds phases (multiplies with an exponential) of a horizontal and vertical translation matrices, multiplied by scalar translation amounts, where the phases range from $-\pi...\pi$ in horizontal and vertical directions, respectively. If translation amount is 1, then the center of the image is translated (with wrapping around) to the edge of the image, and smaller numbers translate proportionally. The simplicity and continuity of all this suggests that these structures may have fundamental uses in some future work.

Also observe that, related to Eq. (3), the cumulative distribution function of a continuous random variable $X$ can be recovered using

$$F_X(x) = \mathcal{L}^{-1}\left\{\frac{1}{s}\,\mathrm{E}\left[e^{-sX}\right]\right\}(x) = \mathcal{L}^{-1}\left\{\frac{1}{s}\mathcal{L}\{f\}(s)\right\}(x), \quad (40)$$

as taking the expectation value involves multiplying with a density function and integrating it out, mathematically quite equal to the Laplace transform. See `https://en.wikipedia.org/wiki/Laplace_transform#Probability_theory`.

TABLE II: Some properties of the Laplace transform.

| | $t$-domain (time, space, or similar)[a] | $s$-domain (frequencies, wave numbers)[b] |
|---|---|---|
| Linearity | $af(t) + b\,g(t)$ | $aF(s) + b\,G(s)$ |
| t-domain derivative[c] | $f'(t)$ | $sF(s) - f(0)$ |
| s-domain derivative | $t^n f(t)$ | $(-1)^n F^{(n)}(s)$ |
| t-domain integration | $\int_0^t f(t')\,dt'$ | $F(s)/s$ |
| s-domain integration | $f(t)/t$ | $\int_s^\infty F(s')\,ds'$ |
| Time shifting[d] | $f(t-a)\,u(t-a)$ | $e^{-as}F(s)$ |
| Frequency shifting | $e^{at}f(t)$ | $F(s-a)$ |
| Time scaling | $f(at)$ | $F(s/a)/a$ |
| Cross-correlation[e] | $(f \star g)(t)$ | $F^*(-s^*)\,G(s)$ |
| Periodic function | $f(t)$ | $\int_0^T e^{-st}f(t)\,dt/(1-e^{-Ts})$ |
| Matrix exponential[f] | $e^{tX}$ | $(sI - X)^{-1}$ |

[a] Note that time $t$ could equally well be a spatial coordinate.
[b] Note that the $s$-domain includes also exponential attenuation/amplification. With $s = i2\pi f$ it is equivalent to the Fourier space.
[c] More generally, for $f^{(n)}(t)$, $\mathcal{L} = s^n F(s) - \sum_{k=1}^n s^{n-k} f^{(k-1)}(0)$, which is one of the most important properties of the Laplace transform (and also Fourier transform), converting differential equations to algebraic equations.
[d] Here $u$ is the step function.
[e] Conventions differ, but here $(f \star g)(t) = \int_0^\infty f(\tau)^*\,g(t+\tau)\,d\tau$.
[f] Resolvent is correct solution for all values of $s$ where the real part is sufficiently large and positive.

### 7. Iteration

Derivatives and differentials are essential for modeling. Simplifying from the Schrödinger equation in Eq. (29) (by dividing and abstracting the constant $i\hbar$ into the matrix), we could arrive at a reduced form such as

$$\dot{x} = \frac{dx}{dt} = Ax, \tag{41}$$

where the evolution of the state is described by

$$x = e^{At}x_0. \tag{42}$$

One can study the matrix exponentials in Eq. (42) by calculating the first terms in Eq. (32) by hand, or usually by studying the eigenvectors of $A$, and their respective eigenvalues. Clearly, the definiteness (sign of real part) of eigenvalues of $A$ is important in these kind of first-order differential equations, as exponentiation then results in diverging, converging, or oscillating behavior (or their mixture). In discrete case, however, where

$$x(k+1) - x(k) = Ax(k) \tag{43}$$

$$x(k+1) = x(k) + Ax(k), \tag{44}$$

eigenvalues being in the left part of the complex plane is not enough for stability (the state converging towards zero), as the magnitudes of the eigenvalues of $A$ need to be less than one (i.e. inside the unit circle in the complex plane) for the state to stay bounded, which is a crucial distinction that we will return only later. In the continuous case of Eq. (42), negative (real part) eigenvalues are enough for stability, but in the discrete case in Eq. (43), if the eigenvalues are too negative (real part less than $-1$, for example), the back and forth hopping will break the iteration and state diverges—but also positive eigenvalues are enough for stability if their magnitude is less than one. By "stability", I am being intentionally obscure here (one can look into literature on control theory for the exact terminology), as the important emphasis here is on models that do not "explode" to infinity, keeping the state at least bounded, but also often vanishing to zero which can be very useful, too, when utilized properly.

One can study the Eqs. (29), (30), (41), and (42) with some example matrices such as

$$H = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{(scalar, even)}, \tag{45}$$

$$H = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad \text{(hyperbolic, even)}, \tag{46}$$

$$H = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \text{(hyperbolic, odd)}, \tag{47}$$

$$H = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \quad \text{(euclidean, odd)}, \tag{48}$$

$$H = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad \text{(hyperbolic, odd)}, \tag{49}$$

where I have inserted some preliminary "gestalts" of the presented forms. The above matrices start to have some fundamental feel to them, in the sense that some of the structures cannot be made any simpler, but the discussion here should be structured much more to be actually useful, which I have also aimed to do below. Multiplying the matrix with the imaginary unit in the exponent makes some of the behaviours the opposite (as $i^2 = -1$, so if something squares to 1, then it squares to $-1$, and vice versa).

There is also the curious fact that any square matrix $A$ can be represented as a sum of a symmetric and skew-symmetric matrix,

$$A = \frac{1}{2}\left(A + A^\mathsf{T}\right) + \frac{1}{2}\left(A - A^\mathsf{T}\right), \tag{50}$$

reminiscent of the distinction between even and odd functions. Hermitian matrices, where the real part is symmetric and the imaginary part skew-symmetric, by definition, combine these properties interestingly, and they have only real eigenvalues, so Hermitian matrices can be equivalently characterized as $n$-dimensional hyperellipsoids, where the eigenvalues represent the magnitudes of semi-axes (negative eigenvalues inverting the direction or handedness). Exponentiation of a symmetric or

Hermitian matrix preserves the symmetricity and Hermitianness, and exponentiation of skew-symmetric or skew-Hermitian matrix results in orthogonal or unitary matrices, respectively. Note that in the Schrödinger Eq. (29) and (30), the Hermitian observable $H$ is multiplied by the imaginary unit $i$, which results in skew-Hermitian matrix $iH$ being exponentiated to unitary matrix, which is a kind of a rotation matrix, resulting in multidimensional oscillation.

## 8. Algebras

We can get much more insight into the possible structure of matrices by studying geometric algebras (Clifford algebras) and their matrix representations. Following Sobczyk (2019), geometric algebra $\mathbb{G}_{1,1}$ can be generated from two null vectors $a^2 = 0 = b^2$, where $ab + ba = 1$. They have a canonical matrix representation

$$[a] := \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \qquad [b] := \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = [a]^T, \tag{51}$$

thus, using ordinary matrix multiplication,

$$[ba] = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} =: u_+, \qquad [ab] = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} =: u_-. \tag{52}$$

Using the above, we can generate the standard basis $\{\mathbf{e}, \mathbf{f}\} \in \mathbb{G}^1_{1,1}$, where the basis vectors $\mathbf{e}^2 = 1$ and $\mathbf{f}^2 = -1$ are orthogonal, $\mathbf{ef} = -\mathbf{fe}$, representing the odd parts of the algebra, $\mathbb{G}^-_{1,1}$, by

$$[a + b] = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} =: [\mathbf{e}], \tag{53}$$

$$[a - b] = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} =: [\mathbf{f}], \tag{54}$$

and the even part, $\mathbb{G}^+_{1,1}$, respectively, by

$$[ba + ab] = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I = [1] =: 2\ a \cdot b, \tag{55}$$

$$[ba - ab] = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = [\mathbf{ef}] =: [u] =: 2\ b \wedge a. \tag{56}$$

Note that the idempotents $u^2_\pm = u_\pm = (1 \pm u)/2$ partition unity, $u_+ + u_- = 1$, and they are mutually annihilating, $u_+ u_- = 0$.

The above structure is illustrative to think in terms of *odd* and *even* number planes. In the odd plane, there are two orthogonal vectors (in the case of $\mathbb{G}_{1,1}$), $\mathbf{e}$ and $\mathbf{f}$, and "rotated 45 degrees" in between them are the null vectors $a = (\mathbf{e} + \mathbf{f})/2$ and $b = (\mathbf{e} - \mathbf{f})/2$. Any geometric number that is a linear combination of terms, each of which has only an odd number of basis vectors as factors, is in the odd plane $\mathbb{G}^-_{1,1}$, and it can be represented simply in any coordinate basis in that plane. Thus, by geometric construction, for some $g' \in \mathbb{G}^-_{1,1}$, where the coordinates $(g_1, g_2) \in \mathbb{R}^2$ and $(g_{21}, g_{12}) \in \mathbb{R}^2$,

$$\begin{aligned} g' &= g_1 \mathbf{e} + g_2 \mathbf{f} \\ &= g_{21} a + g_{12} b \\ &= (g_1 + g_2) a + (g_1 - g_2) b \\ &= [(g_{21} + g_{12})/2]\mathbf{e} + [(g_{21} - g_{12})/2]\mathbf{f}. \end{aligned} \tag{57}$$

The same construction applies also in the even number plane, where the basis consists of 1 and $\mathbf{ef}$ (as scalar 1 does not have any basis vectors as factors, so their count is zero, which is an even number), and "rotated in between" is a basis $ba = (1 + \mathbf{ef})/2$ and $ab = (1 - \mathbf{ef})/2$. Thus for some $g'' \in \mathbb{G}^+_{1,1}$, where the coordinates $(g_0, g_3) \in \mathbb{R}^2$ and $(g_{11}, g_{22}) \in \mathbb{R}^2$,

$$\begin{aligned} g'' &= g_0 + g_3 \mathbf{ef} \\ &= g_{11} ba + g_{22} ab \\ &= (g_0 + g_3) ba + (g_0 - g_3) ab \\ &= (g_{11} + g_{22})/2 + [(g_{11} - g_{22})/2]\mathbf{ef}. \end{aligned} \tag{58}$$

Therefore a geometric number $g \in \mathbb{G}_{1,1} = \mathbb{G}^+_{1,1} + \mathbb{G}^-_{1,1} = \mathbb{G}^{0+2}_{1,1} + \mathbb{G}^1_{1,1} = \mathbb{G}^0_{1,1} + \mathbb{G}^1_{1,1} + \mathbb{G}^2_{1,1}$,

$$\begin{aligned} g &= g_0 + g_1 \mathbf{e} + g_2 \mathbf{f} + g_3 \mathbf{ef} \\ &= g_{11} ba + g_{21} a + g_{12} b + g_{22} ab, \end{aligned} \tag{59}$$

can be represented as a matrix[87]

$$[g] = \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix}. \tag{64}$$

---

[87] There is a possibility of rotating the basis vectors with

$$\begin{aligned} a' &= hah^{-1} \\ b' &= hbh^{-1}, \end{aligned} \tag{60}$$

as then all the other constructs will follow, as

$$a'b' = hah^{-1}hbh^{-1} = habh^{-1}, \tag{61}$$

displaying how the various concepts (such as oddness and evenness) are really relative concepts, rotations between them being possible—see Sobczyk (2019). I know that I am being quite dense and rather obscure here, but this is a draft working paper, displaying some ideas that I am quite confident with already but not necessarily certain yet, some of which will be selected and developed further. Also in Sobczyk (2019), the inverse can be calculated, at least in $\mathbb{G}_{1,1}$, using mixed conjugation [see Eq. (86)] and determinant, as

$$g^{-1} = \frac{1}{g} = \frac{g^*}{gg^*} = \frac{g^*}{\det g}, \tag{62}$$

which simplifies some matters, but I do not yet follow how general it is [other than in $\mathbb{G}_{1,1}(\mathbb{R})$].

For example, if the coordinates $g_0$, $g_1$, $g_2$, and $g_3$ are other geometric numbers in $\mathbb{G}_{p,q}$ (instead of real numbers), then the coordinates in this canonical $(a, b)$ basis are

$$[G] = \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix} = \begin{pmatrix} g_0 + g_3 & g_1 - g_2 \\ g_1^- + g_2^- & g_0^- - g_3^- \end{pmatrix}, \tag{63}$$

where $g^- := \mathbf{e}g\mathbf{e}$ is the operation of geometric inversion in $\mathbb{G}_{p,q}$, resulting in all vectors in $g$ being replaced by their negatives. See

The above $\mathbb{G}_{1,1}(\mathbb{R}) \cong \mathbb{G}_{2,0}(\mathbb{R})$ can be complexified to $\mathbb{G}_{1,1}(\mathbb{C}) \cong \mathbb{G}_{1,2}(\mathbb{R}) \cong \mathbb{G}_{3,0}(\mathbb{R})$, so Pauli matrices[88] can be attained, spanning $\mathbb{G}_3^1$, by $\sigma_1 := \mathbf{e}$, $\sigma_2 := i\mathbf{f}$, $\sigma_3 := \mathbf{ef}$, where by construction the basis is orthogonal (anticommutative), and also by inspection $\sigma_i\sigma_j + \sigma_j\sigma_i = 2\delta_{ij}$.

Those generating nullpotents $a^2 = b^2 = 0$ could enable automatic differentiation, as $f(r+a) = f(r) + f'(r)\,a$, using

$$[r+a] = \begin{pmatrix} r & 0 \\ 1 & r \end{pmatrix}, \qquad [r+b] = \begin{pmatrix} r & 1 \\ 0 & r \end{pmatrix}. \tag{65}$$

Also left ideal $ba + a$, right ideal $ba + b$, and some other variants can be represented:

$$[ba+a] = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}, \qquad [ab+b] = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}, \tag{66}$$

$$[ba+b] = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}, \qquad [ab+a] = \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}. \tag{67}$$

A machine can do advanced mathematics numerically with the above matrices, starting from Eq. (51), without "knowing about" the symbols. The symbols can also be mixed and interpolated between. As many machine learning algorithms result in matrices that have been optimized numerically, the above demonstrates that already now some applications could be utilizing these kind of algebras, without them having been explicitly programmed to do so.

Following Sobczyk (2019), higher-order algebras can be constructed by introducing further $(a_i, b_i)$ pairs of null vectors $0 = a_i^2 = b_i^2$, the pairwise products of which partition unity $1 = b_i a_i + a_i b_i$, and are otherwise (when $i \neq j$) pairwise orthogonal $0 = a_i a_j + a_j a_i = b_i b_j + b_j b_i = a_i b_j + b_j a_i$. Then each such pair of null vectors defines a pair of basis vectors, $e_i = a_i + b_i$, $f_i = a_i - b_i$, where $e_i^2 = 1$ and $f_i^2 = -1$.

For example, starting from the primitive idempotent $b_1 a_1 b_2 a_2$, Sobczyk (2019) calculates (using directed Kronecker products of dual numbers from both sides), that

$$\begin{aligned} \mathbb{G}_{2,2} &= \begin{pmatrix} 1 \\ a_1 \end{pmatrix} \overrightarrow{\otimes} \begin{pmatrix} 1 \\ a_2 \end{pmatrix} b_1 a_1 b_2 a_2 \begin{pmatrix} 1 & b_2 \end{pmatrix} \overleftarrow{\otimes} \begin{pmatrix} 1 & b_1 \end{pmatrix} \\ &= \begin{pmatrix} 1 \\ a_1 \end{pmatrix} \overrightarrow{\otimes} \begin{pmatrix} b_1 a_1 b_2 a_2 & b_2 b_1 a_1 \\ a_2 b_1 a_1 & b_1 a_1 a_2 b_2 \end{pmatrix} \overleftarrow{\otimes} \begin{pmatrix} 1 & b_1 \end{pmatrix} \\ &= \begin{pmatrix} b_1 a_1 b_2 a_2 & b_1 b_2 a_2 & b_2 b_1 a_1 & b_2 b_1 \\ a_1 b_2 a_2 & a_1 b_1 b_2 a_2 & a_1 b_2 & -b_2 a_1 b_1 \\ a_2 b_1 a_1 & a_2 b_1 & b_1 a_1 a_2 b_2 & b_1 a_2 b_2 \\ a_1 a_2 & -a_2 a_1 b_1 & a_1 a_2 b_2 & a_2 b_2 a_1 b_1 \end{pmatrix}, \end{aligned} \tag{68}$$

which is quite remarkable, as then algebraic manipulations and matrix calculations coincide, unifying the languages. Different elements in the above basis are really independent in the sense that calculating one element does not disturb the others.

Sobczyk notes, inspecting Eq. (68), that only two null vectors $a_1$ and $a_2$ are needed to generate the algebra $\mathbb{G}_{2,2}$, as the other elements can be derived by transposes and multiplications. The $\mathbb{G}_{2,2}$ is thus generated by

$$[a_1] := \begin{pmatrix} 0 & 0 & & \\ 1 & 0 & & \\ & & 0 & 0 \\ & & 1 & 0 \end{pmatrix}, \qquad [b_1] := \begin{pmatrix} 0 & 1 & & \\ 0 & 0 & & \\ & & 0 & 1 \\ & & 0 & 0 \end{pmatrix}, \tag{69}$$

$$[a_2] := \begin{pmatrix} & & & \\ & & & \\ 1 & 0 & & \\ 0 & -1 & & \end{pmatrix}, \qquad [b_2] := \begin{pmatrix} & & 1 & 0 \\ & & 0 & -1 \\ & & & \\ & & & \end{pmatrix}, \tag{70}$$

where the generating null vectors of $\mathbb{G}_{1,1}$ [Eq. (51)] are also embedded.

It is well known that $\mathbb{G}_{2,2}(\mathbb{R}) \cong \mathbb{G}_{3,1}(\mathbb{R})$, and it can be complexified to $\mathbb{G}_{2,2}(\mathbb{C}) \cong \mathbb{G}_{2,3}(\mathbb{R}) \cong \mathbb{G}_{4,1}(\mathbb{R}) \cong \mathbb{G}_{0,5}(\mathbb{R})$. These are interesting as gamma matrices[89] can be represented as complex $4 \times 4$ matrices.

In Dirac representation, the four gamma matrices (the fifth gamma matrix is used in analyzing chirality, for example, but is not usually referred to as a gamma matrix,

Sobczyk (2019, p. 72). So even though the relations in Eqs. (57) and (58) are symmetric and quite straightforward, there may be some complex conjugations or similar mirrorings present when operating with other coordinate fields than real numbers. One can only wonder how is it possible that some minds are seemingly capable of operating with these constructs so accurately, as we are in such an exacting land of concise formulas, that one mistaken minus sign can easily bring havoc and utter chaos to everything in its vicinity.

[88] `https://en.wikipedia.org/wiki/Pauli_matrices`

[89] `https://en.wikipedia.org/wiki/Gamma_matrices`

and has a different designation) are

$$[\gamma_0] := \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & -1 & \\ & & & -1 \end{pmatrix} = [\mathbf{e}_0], \tag{71}$$

$$[\gamma_1] := \begin{pmatrix} & & & 1 \\ & & 1 & \\ & -1 & & \\ -1 & & & \end{pmatrix} = [\mathbf{e}_{10}], \tag{72}$$

$$[\gamma_2] := \begin{pmatrix} & & & -i \\ & & i & \\ & i & & \\ -i & & & \end{pmatrix} = [\mathbf{e}_{20}], \tag{73}$$

$$[\gamma_3] := \begin{pmatrix} & & 1 & \\ & & & -1 \\ -1 & & & \\ & 1 & & \end{pmatrix} = [\mathbf{e}_{30}], \tag{74}$$

$$[\gamma_5] := \begin{pmatrix} & & 1 & \\ & & & 1 \\ 1 & & & \\ & 1 & & \end{pmatrix} = [i\gamma_{0123}] = [i\mathbf{e}_{123}], \tag{75}$$

where on the right hand side there is one interpretation, as generated from four orthogonal vectors,

$$[\mathbf{e}_0] := \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & -1 & \\ & & & -1 \end{pmatrix} = [(a_2 + b_2)(a_2 - b_2)], \tag{76}$$

$$[\mathbf{e}_1] := \begin{pmatrix} & & & -1 \\ & & -1 & \\ & -1 & & \\ -1 & & & \end{pmatrix} = [(a_2 + b_2)(a_1 - b_1)], \tag{77}$$

$$[\mathbf{e}_2] := \begin{pmatrix} & & & i \\ & & -i & \\ & i & & \\ -i & & & \end{pmatrix} = [i(a_2 + b_2)(a_1 + b_1)], \tag{78}$$

$$[\mathbf{e}_3] := \begin{pmatrix} & & -1 & \\ & & & 1 \\ -1 & & & \\ & 1 & & \end{pmatrix} = [-(a_2 + b_2)]. \tag{79}$$

These are perhaps not as elegant as they could be, as theoretically $e_1 := a_1 + b_1$, $e_2 := a_2 + b_2$, $e_3 := if_1 = i(a_1 - b_1)$, $e_4 := if_2 = i(a_2 - b_2)$, $f_3 := ie_1 f_1 e_2 f_2$ are already orthogonal by construction and they could perhaps be used directly, as clearly in the above Dirac basis $\mathbf{e}_0 = e_2 f_2 = -ie_2 e_4$, $\mathbf{e}_1 = e_2 f_1 = -ie_2 e_3$, $\mathbf{e}_2 = ie_2 e_1$, and $\mathbf{e}_3 = -e_2$, now $e_2$ being a factor in each. But I leave these for future work, if even relevant.[90] It is interesting regardless that the structure of gamma matrices can be interpreted as stemming from a five-dimensional basis (or four-dimensional, as the fifth one is then determined). Note that similarly to there being different representations of gamma matrices, such as in Majorana basis or Weyl (chiral) basis, these interpretations are not unique. In Weyl basis, for example, the role of $\gamma_0$ and $\gamma_5$ are reversed, which is interesting by itself as they are both timelike (or even hypervolume-like), squaring to one.

It is also interesting how the algebraic structure may make more evident some astonishing intuitions that some theorists deep into mathematical physics seem to have—by practicing a lot, having a lot of quality time, getting familiar with all kinds of relations with regards to these matrices, they may have been able to develop also largely subconscious mental models that are representative of these algebraic structures and beyond, facilitating their extraordinary insights.

Using the above relations, the volume elements are

$$[\mathbf{e}_{123}] := \begin{pmatrix} & & -i & \\ & & & -i \\ -i & & & \\ & -i & & \end{pmatrix} = [(a_1 + b_1)(a_2 + b_2)i(a_1 - b_1)], \tag{80}$$

$$[\mathbf{e}_{0123}] := \begin{pmatrix} & & -i & \\ & & & -i \\ i & & & \\ & i & & \end{pmatrix} = [(a_1 - b_1)(a_2 - b_2)i(a_1 + b_1)], \tag{81}$$

and charge conjugation matrix $C$, defined as satisfying $C\gamma_\mu C^{-1} = -(\gamma_\mu)^T$, where $\mu$ runs from 0 to 3,

$$[C] := \begin{pmatrix} & & & -1 \\ & & 1 & \\ & -1 & & \\ 1 & & & \end{pmatrix} = [i\mathbf{e}_2] = [(a_1 + b_1)(a_2 + b_2)]. \tag{82}$$

---

[90] I leave these ponderings here for now, as it may later be relevant for studies on Suntola [see (Lievonen, 2023)], and also the implicate and explicate order of Bohm (see fn. 5) could be interpreted as the algebraic structure vs. the numerical structure, among other interpretations.

For example, watch how Anthony Lasenby describes gamma matrices: "And again, this is, quite interesting idea that—um—as human beings we're—we're very—we're not used at all the idea that we're actually shooting along a time axis at the speed of light—at any moment, that's what we're primarily doing, we're moving extremely rapidly up the time axis, and—so these vectors which we have around us, you know Cartesian basis vectors in this room, are actually being swept, really rapidly—and so it's very interesting that in the 4D way of thinking about that, we—they're bivectors, because that's sweeping out, but coming back down the other way, we can think—we can use the gamma naught to define these things that are actually vectors, for all intents and purposes, in this room." `https://youtu.be/m7v2IUJtC3g?t=1422`.

Note that I have not yet analyzed most of these relations very thoroughly, just derived them by inspecting the structure of gamma matrices and experimenting with possible alternative interpretations in light of Sobczyk (2019).[91] They seem promising in giving geometric structure to the gamma matrices, so that some meanings of the Dirac equation

$$i\hbar\gamma^\mu\partial_\mu\psi(x) - mc\,\psi(x) = 0, \tag{83}$$

could perhaps be later interpreted also by the scientific layman.[92]

In addition to ordinary matrix transposes and Hermitian conjugates (where also the imaginary unit is negated) from linear algebra, there seems to be a possibility of transposing "in the other direction", as following Sobczyk (2019), combining geometric inversion (where the odd part is negated) with the geometric reverse (where the order of all products is reversed), results in quite curious matrices:

$$[g] := \begin{pmatrix} g_{11} & g_{12} & g_{13} & g_{14} \\ g_{21} & g_{22} & g_{23} & g_{24} \\ g_{31} & g_{32} & g_{33} & g_{34} \\ g_{41} & g_{42} & g_{43} & g_{44} \end{pmatrix} \tag{85}$$

$$[g^*] := \begin{pmatrix} g_{44} & -g_{34} & g_{23} & -g_{14} \\ -g_{43} & g_{33} & -g_{23} & g_{13} \\ g_{42} & -g_{32} & g_{22} & -g_{12} \\ -g_{41} & g_{31} & -g_{21} & g_{11} \end{pmatrix} \tag{86}$$

For the adventurous, the pattern in Eq. (68) continues to ever higher orders, for example generating $\mathbb{G}_{3,3}$, these are enough:

$$[a_0] := \begin{pmatrix} 0 & 0 & & & & & & \\ 1 & 0 & & & & & & \\ & & 0 & 0 & & & & \\ & & 1 & 0 & & & & \\ & & & & 0 & 0 & & \\ & & & & 1 & 0 & & \\ & & & & & & 0 & 0 \\ & & & & & & 1 & 0 \end{pmatrix} \tag{87}$$

$$[a_1] := \begin{pmatrix} & & & & & & & \\ & & & & & & & \\ 1 & 0 & & & & & & \\ 0 & -1 & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & 1 & 0 & & \\ & & & & 0 & -1 & & \end{pmatrix} \tag{88}$$

$$[a_2] := \begin{pmatrix} & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ 1 & 0 & & & & & & \\ 0 & -1 & & & & & & \\ & & -1 & 0 & & & & \\ & & 0 & 1 & & & & \end{pmatrix}, \tag{89}$$

where the pattern of duplicating the previous basis diagonally and extending it to a new generating null vector $a_3$ by mirroring the previous submatrix to positive and negative versions should be evident by inspection.[93]

Complexification the above $\mathbb{G}_{3,3}(\mathbb{R})$ in Eq. (87) to $\mathbb{G}_{3,3}(\mathbb{C}) \cong \mathbb{G}_{3,4}(\mathbb{R}) \cong \mathbb{G}_{5,2}(\mathbb{R}) \cong \mathbb{G}_{7,0}(\mathbb{R}) \cong \mathbb{G}_{1,6}(\mathbb{R})$ would introduce the imaginary element $i := e_3e_2e_1f_1f_2f_3f_4$, and an even number of basis elements could be "inverted" to opposite definiteness by multiplying with $i$. Note however, that some relations can also be seen as generalizing quaternions or double quaternions, using them as a base field, too [see the various parts of Sobczyk (2019)]. I regret the sloppy language here, but the aim at this point is to give some "pointers" to the interested, such as myself, rather than a treatise in professional mathematics.

In the future, this brief tour of some of the fundamental structures in mathematics should be augmented with discussion on the multidimensional derivatives and differentials—facilitated by geometric calculus, for example—but as I am not as familiar with them as someone deeply into mathematical physics would be, for example, I will defer that to future work. Of course I

[91] Note to self: Double check above, and also check if $\mathbf{e}_0 = e_2f_2$ (or $if_2$?), $\mathbf{e}_1 = e_1$, $\mathbf{e}_2 = e_2$, $\mathbf{e}_3 = if_1$ would work, so otherwise the elements stay the same ($\mathbf{e}_{123} = e_1e_2if_1$, $\mathbf{e}_{0123} = f_1f_2ie_1$) but $\mathbf{e}_1$, $\mathbf{e}_2$, and $\mathbf{e}_3$ would get rotated to simpler representations.

[92] For now, I just note that referring to (Lievonen, 2023), it seems as if

$$\begin{aligned} i\hbar\gamma^\mu\partial_\mu\psi(x) &= mc\,\psi(x) \\ i\hbar_0c_0\gamma^\mu\partial_\mu\psi(x) &= mc_0\,\psi(x) \\ i\hbar_0\gamma^\mu\partial_\mu\psi(x) &= m\,\psi(x) \\ i\gamma^\mu\partial_\mu\psi(x) &= k\,\psi(x), \end{aligned} \tag{84}$$

where the last line is some kind of a divergence in a 4-dimensional $(\mathbf{e}_0, \mathbf{e}_{10}, \mathbf{e}_{20}, \mathbf{e}_{30})$ space, resulting in a wave number $k$ representing mass.

[93] It is quite evident from Sobczyk (2019) that mathematics really has progressed, compared to some older treatises on $n$-dimensional spaces such as Sommerville (1929).

am not too unfamiliar with those concepts either, so the next section will utilize derivatives already in an organic way.

Already now, however, the above should illustrate the possibilities provided by mathematical models for reducing some important concepts into very condensed forms, which could be taught already in basic education. Matrices and linear algebra, especially, seem to be "in between" the continuous, numerical forms, and discrete, symbolic forms, thus offering ideas for bridging the qualitative and the quantitative. By keeping the matrices in mind, one can keep grounding the mathematics to some practical computations, instead of ascending into ever higher abstractions as some theoretical endeavors seem to have done to themselves.

## V. THE MATHEMATICAL MODEL OF MODELING

Here we develop some steps towards *metamodeling*, i.e. recursive modeling of modeling, in its most simple terms imaginable. It is hoped, that by "dividing it at the joints, as nature directs, not breaking any limb in half as a bad carver might", we model the essential complexity of modeling, and the accidental complexity is kept at bay. The model aims for the most general relevance, thus displaying some representational capabilities of generative structures in their bare form, facilitating discussion, analysis, critique, and perhaps even synthesis of such systems.[94]

### A. Core assumptions

#### 1. Models as representations stemming from the universal modeling pursuit

We start by noting how Sebeok and Danesi have described the concept of a model (Danesi, 2017, p. 1497)

> A *model* can now be defined as the overall relation $X$ *stands for* $Y$ itself and a *form* as the $X$ component of that relation, since it is something that has been imagined or made externally (through some physical medium) to stand for $Y$.

To start formalizing the above rather informal conception of a model, we start with the simplest model imaginable, a linear relation between $x$ and $y$:

$$y = ax, \tag{90}$$

where $y$, $a$, and $x$ are scalars at first, but soon to be upgraded to richer structures.

The important step we take is to not assume that the relation is always exact, but to state the model in a variational form

$$y - ax = \epsilon, \tag{91}$$

where the residual or error $\epsilon$ should be as close to zero as possible, for the model to be accurate. This is the core assumption here, informed by cybernetics and control theory [see, for example, Hyötyniemi (2006) and later works]—that there is a matching pursuit or modeling instinct, a sort of *reason for modeling* at the core, and that this form of metamodeling formalizes the target, goal, aim, or meaning of the modeling.

The idea of Eq. (91) is to minimize the current mismatch $\epsilon$ in modeling, a sort of necessity or desperation, even emergency surfaced by the common habits (the model) not matching some sobering fact $y$ in need for explanation or resolution. In that sense the residual or error $\epsilon$ could be called a "creative tension", as without it, not much would actually change—the $x$ would simply mirror the $y$, without any learning, adaptation, or "thinking" needing to take place.

One notices that already one could generate a massive amount of discourse using this simple model. That is the objective here, to describe as general structures as possible, so the application areas are as wide as possible. Even if the descriptions as words, or operationalizations to specific phenomena, vary, the mathematical relations displaying the representational structure stay the same. At the same time, the modeling here should encourage textual descriptions—the modeling objective of minimizing the residual error is standard in engineering, for example, but too often its meaning or possibilities for various interpretations are skipped, going straight to assuming models as epistemic, not seeing the possibilities for ontic modeling in nature using these structures.

#### 2. Avoiding layers and nonlinearities at the beginning

We already see two important extensions to the model. First, we could layer the models, modifying Eq. (91),

$$\epsilon = y - ax = y - af(z), \tag{92}$$

where $x = f(z)$ is some other relation that could have its own functional structure. We do not pursue that much here, as we think that at first, it is most important to understand the "core loop" that drives the system in any case, and only after understanding Eq. (91) more fully, should we extend it to layered models with

[94] I myself have been experimenting with these kind of adaptive systems, implementing some of the algorithms using BLAS (Basic Linear Algebra Subprograms), which is available in the Accelerate framework on macOS, but also supercomputers, for example, utilizing effectively vectorized SIMD (single instruction, multiple data) and special high performance cores, succeeding in validating at least some of the ideas underlying the "tools for thinking" that are elaborated here.

new dynamics. This is in opposition to current modeling paradigm in machine learning, for example, where usually one empirically (tuning the hyperparameters using cross-validation and grid search) chooses an architecture with a large number of layers, and relies on the automatic differentiation (autograd) in the computational frameworks to provide valid gradients through the network, without necessarily inspecting the symbolic form of the gradients at all. We try to avoid that kind of "black boxes" here, where gradient backpropagation demonstrably works well, but the physical realization of which may stay quite mysterious if the system cannot be analyzed properly.

The other straightforward way to extend the model in Eq. (91) is by enclosing the estimate $ax$ with some function $g$, for example by mapping it with the exponential function and normalizing it,

$$\epsilon = y - g(ax) = y - f(e^{ax}), \tag{93}$$

where $f$ is usually some normalizing function, ensuring that the probabilities (in that specific case) sum to one [as in softmax, Eq. (11)]. At first, however, we aim for simplicity and consider $g$ as an identity function. For experts in Bayesian analysis, for example, this kind of approach to modeling by pursuing simple algebraic relations may seem strange and even heretic at first, but consider that in many machine learning studies one first postulates a complicated conditional probability function that is often then simplified to something manageable, adding various assumptions, often resulting in very simple pseudocode after all. So by contrast, here we begin already with those simple structures, and analyze the modeling capabilities and their possible hidden assumptions only later, thus potentially staying in the algebraic world of simple structures for longer.

### 3. Extending to synchronic and diachronic dimensions

Now, going forward, we will actually commence analyzing the model in Eq. (91) in two most basic ways: (1) we will consider multiple synchronous eventualities, extending the $y$ and $x$ linearly to multidimensional vectors, thus enabling more continuous "fields" or distributed forms to be modeled, and (2) we will consider multiple occasions of errors $\epsilon$, in effect calculating it for multiple $y$ and $x$ pairs (diachronically, without ordering the events at this point). Abstractly, these are the two most fundamental directions we can straightforwardly apply the model to, using linear algebra, therefore they seem also most relevant for analysis.

So for synchronic modeling, we extend $x$ and $y$ to be column vectors, so the form of Eq. (91) is

$$y - Ax = \epsilon, \tag{94}$$

where the model relation $A$ is now a matrix. This is the beginning of multiple factor analysis of Eq. (1) on page 2, the factors being the columns of matrix $A$.

The dimensionalities are such that the relation is syntactically correct—for example, if $y$ is a one-dimensional column vector (a scalar), then $A$ is a $1 \times n$ matrix (a row vector or "1-form"), where $n$ is the dimensionality of the column vector $x$ (or one, if $x$ is a scalar). Usually, however, it is the other way, and the data $y$ is more complicated than the representation $x$, of course. Also the dimensionality of $\epsilon$ is then determined by the rest of the structures—error $\epsilon$ has to have the same dimensions as $y$ for the relation in Eq. (94) to be syntactically valid. Note that this syntactical validity is not as strict a restriction as it may sound at first, as in many high-dimensional settings most elements can easily be zero (or stay at some common mean, depending on the model), sparse structures then appearing effectively as low-dimensional as necessary, even though embedded in high-dimensional spaces. All in all, we consider the simplicity of this extension a virtue, as the symbols did not need much alteration, which could be indicative of some fundamental ideas being present.

Note that the error $\epsilon$ is now a vector, having multiple elements, that represent errors in modeling the corresponding elements in vector $y$. All these need to be minimized as close to zero as possible, for the model to be accurate, but there may be various tradeoffs and priorities involved, which we come to later.

For diachronic modeling, in turn, the most straightforward way is to simply then collect multiple triplets of those corresponding $x$, $y$, and $\epsilon$ vectors as columns in matrices $X$, $Y$, and $E$, resulting in

$$Y - AX = E. \tag{95}$$

Note that in spite of very compact notation using matrices, the columns in $Y$, $X$, and $E$ do not mix or interact in any way in the above—Eq. (95) is quite astonishing in its ability to now represent modeling millions of dimensions (rows) of millions of occasions (columns), using the same model matrix $A$.[95]

### 4. Mixing occasions using the attention matrix

At this point, it seems prudent to also extend the model with an attention matrix $C$, similarly to Eq. (4) on page

[95] Note that at some point in the future, we may need to defer to index notation such as

$$y^{(j)} - a_{(j,i)}x^{(i)} = \epsilon^{(j)} \tag{96}$$

to model higher-order tensors, but we will keep using matrices for now to enable verifying the derivations, especially the gradients, using computational tools.

6, crucially important in transformer models. To not get ahead of ourselves and complicate matters too early, we do not yet insert any structure to the attention matrix [which would be essential for transformers, as attention is usually quadratic on $X$ as displayed by Eq. (8)], but we just note that theoretically, multiplying the model from the right is one of the simplest extensions conceivable. The extra functionality provided by $C$ is the possibility of mixing the occasions (columns), that would not otherwise be possible with this model structure. It does not always necessitate taking the time as the extended dimension, as there could be, for example, multiple occasions from the same kind of process, but each at different phase, simultaneously present, that C would now mix together. We expect that it is instructive to see where the attention matrix $C$ ends up in the derivations, and in any case, it can always be taken to an identity matrix $I$ for simplicity, disappearing from the formulas (comparable to multiplying with a scalar 1).

Thus the modeling relation that we will consider here is

$$Y - AXC = E, \tag{97}$$

where, again, the dimensionalities are such that they are syntactically valid. The above form may be a bit reminiscent of the sandwich product of spinors and geometric algebra, but we cannot make the connection here (at least not yet).

## 5. Adaptation and learning facilitated by a persistent whole

Now the important development is preparing for the adaptation and learning, that the model itself does not yet seem to display. In other words, as a mathematical model of a model, Eq. (97) suggests that $XC$ contains the forms (as columns), and each stand for (represent) a corresponding form (as a column) in $Y$, but it is not yet a model of *modeling*, dynamics being conspicuous by their absence.

To arrive at modeling dynamics, we make the simplest possible move, and aggregate all the errors $E$ together into a scalar-valued objective function $\mathcal{S}$ to be minimized. The objective or cost function, then, ties the model into a cohesive whole, by introducing some kind of a mean field, that works as a global constraint compelling the model to be parsimonious in its ways and means. Mathematically, most concise arrangement is to minimize the sum of squared errors, and that may also have a physical interpretation in some settings as residual energy to be minimized (by using, spending, and conducting it elsewhere).[96]

As minimizing any function is the same as maximizing its negative, the objective function that we study first, is actually a negated version of the sum of squares,

$$\begin{aligned}
\mathcal{S}_1 &= -\|E\|_2^2/2 && (98)\\
&= -\|Y - AXC\|_2^2/2 && (99)\\
&= -\mathrm{tr}\left[(Y - AXC)^\dagger (Y - AXC)\right]/2 && (100)\\
&= -\mathrm{tr}\left[(Y - AXC)^\dagger Y - (Y - AXC)^\dagger AXC\right]/2\\
&= -\mathrm{tr}\left[Y^\dagger Y - (AXC)^\dagger Y\right.\\
&\qquad \left. - Y^\dagger AXC + (AXC)^\dagger AXC\right]/2\\
&= \left\{-\mathrm{tr}\left(Y^\dagger Y\right) - \mathrm{tr}\left[(AXC)^\dagger AXC\right]\right.\\
&\qquad \left. + \mathrm{tr}\left[(AXC)^\dagger Y + Y^\dagger AXC\right]\right\}/2,
\end{aligned}$$

where various equivalent forms are visible that are not all necessary, but which can give more insight into the various terms that we aim to maximize from below, to minimize the errors.

The division by two in the above is a convenience factor that can equally well be omitted, as it does not affect the possible optima of the objective function, being just a scalar. Similarly, squaring the norm may not be necessary, as omitting it would just introduce an inverse of the norm (as derivative of the outer function) as a scaling factor in the gradients, but we do not yet know what is truly relevant here, so will proceed as laid out.

By defining the objective function in this negated way, we can calculate various gradients and let the system evolve naturally according to their flow, without needing to negate any directions to reach for the optimality that we aim for.

Note that the norm $\|.\|_2$ is the Frobenius norm (square root of the sum of squares), and by squaring it, the result is simply the total sum of squares of each element in matrix $E$. As shown in the above objective function $\mathcal{S}_1$, the Frobenius norm can be equally expressed using the

---

[96] Of course, by defining the objective function $\mathcal{S}$ as a sum over occasions, it is formally somewhat similar to the action in mathematical physics, especially as the stationary values are sought for. In physics, the action is an integral of a Lagrangian, encoding the equations of motion, but here we want to stay in this mathematical land of linear algebra for longer, as we do not know how the ideas of Lagrangian, Hamiltonian, and Routhian mechanics, and the related Legendre transformation, should be introduced here without restricting the modeling to physics only, which we aim to somehow transcend here.

trace, as using a $2 \times 2$ matrix as an example,

$$\begin{aligned}
\|E\|_2^2 &= \operatorname{tr}\left(E^\dagger E\right) \\
&= \operatorname{tr}\left[\begin{pmatrix}\epsilon_{11} & \epsilon_{12}\\ \epsilon_{21} & \epsilon_{22}\end{pmatrix}^\dagger \begin{pmatrix}\epsilon_{11} & \epsilon_{12}\\ \epsilon_{21} & \epsilon_{22}\end{pmatrix}\right] \\
&= \operatorname{tr}\left[\begin{pmatrix}\epsilon_{11}^* & \epsilon_{21}^*\\ \epsilon_{12}^* & \epsilon_{22}^*\end{pmatrix} \begin{pmatrix}\epsilon_{11} & \epsilon_{12}\\ \epsilon_{21} & \epsilon_{22}\end{pmatrix}\right] \\
&= \operatorname{tr}\begin{pmatrix}\epsilon_{11}^*\epsilon_{11} + \epsilon_{21}^*\epsilon_{21} & \epsilon_{11}^*\epsilon_{12} + \epsilon_{21}^*\epsilon_{22}\\ \epsilon_{12}^*\epsilon_{11} + \epsilon_{22}^*\epsilon_{21} & \epsilon_{12}^*\epsilon_{12} + \epsilon_{22}^*\epsilon_{22}\end{pmatrix} \\
&= \operatorname{tr}\begin{pmatrix}|\epsilon_{11}|^2 + |\epsilon_{21}|^2 & \epsilon_{11}^*\epsilon_{12} + \epsilon_{21}^*\epsilon_{22}\\ \epsilon_{12}^*\epsilon_{11} + \epsilon_{22}^*\epsilon_{21} & |\epsilon_{12}|^2 + |\epsilon_{22}|^2\end{pmatrix} \\
&= |\epsilon_{11}|^2 + |\epsilon_{21}|^2 + |\epsilon_{12}|^2 + |\epsilon_{22}|^2. \qquad (101)
\end{aligned}$$

Note that $E$ does not need to be a square matrix—the squared norm is valid and can always be calculated using the trace as above.[97]

### 6. Weighting of errors combined with spectral filtering could prove out to be decisive

While Eq. (99) is most compact, Eq. (100) is more versatile, as the trace has many useful properties, such as linearity and cyclic property. Already now, Eq. (100) suggests extending the model to give different weight to different dimensions and occasions, as multiplying the dimensions with a diagonal matrix $V$ (where the diagonal is a vector $v$) from the left, and the occasions with a diagonal matrix $W$ (where the diagonal is a vector $w$) from the right, results in

$$\begin{aligned}
\mathcal{S}_2 &= -\|VEW^\dagger\|_2^2/2 &(104)\\
&= -\|(vw^\dagger) \odot E\|_2^2/2 &(105)\\
&= -\|V\,(Y - AXC)\,W^\dagger\|_2^2/2 &(106)\\
&= -\operatorname{tr}\left[W\,(Y - AXC)^\dagger\,V^\dagger\,V\,(Y - AXC)\,W^\dagger\right]/2, &(107)
\end{aligned}$$

where $\odot$ is the element-wise (Hadamard) product.

One reason I am emphasizing error weighting here, is that various successful image diffusion models[98] add Gaussian noise to data, where the mixture is such that choosing a phase $\theta$ between $\pi/2$ and 0, the signal is multiplied by $\cos\theta$ and noise by $\sin\theta$, and the error in each occasion is weighted by $\cos\theta/\sin\theta = \cot\theta$, before it is let affect learning the model. The empirically proven usefulness of this computational procedure encourages to search for some reasoning to include that here too, eventually, into the diagonal of matrix $W$ (and the columnar sums of $C$ are also related). In diffusion models, the model needs the used phase as input, too, so perhaps it could be covered here as some global phase. It may be possible to postulate some dynamic process, where induced oscillations, similar to the Schrödinger equation in Eq. (29) or other ideas implicit in this study, would provide this function, perhaps by having multiple systems working orthogonally in tandem.[99]

The aforementioned diffusion models also employ cascaded models which could be modeled as passing only the lower frequencies in the 2D Fourier transformed images to the upper layers, effectively zooming out, and reversing the process back to the details, which suggest a possibility of discovering some basic structure for layered modeling in the future, working on the same computational substrate.[100]

Both of these developments, however, is deferred here for future work (if any), so we will perhaps keep the diagonal $V$ and $W$ matrices, but may hide them from formulas at any point by treating them as identity matrices, to keep the gradients cleaner (denoting these steps using

---

[97] See https://en.wikipedia.org/wiki/Trace_(linear_algebra) and its various properties, including its use in Jacobi's formula

$$\frac{d}{d\alpha}\det A = \operatorname{tr}\left(\operatorname{adj}(A)\frac{dA}{d\alpha}\right), \qquad (102)$$

and the curious relation

$$\det(\exp A) = \exp(\operatorname{tr} A). \qquad (103)$$

Note that the trace is also equal to the sum of the eigenvalues, and the determinant to the product of the eigenvalues. Interpreting the determinant as the volume of an $n$-dimensional parallelepiped is often evident, but it is important to keep in mind that imaginary eigenvalues result in rotations, where the "volume" is more difficult to interpret as such.

Note also how the form $E^\dagger E$ is related to the Gram matrix (https://en.wikipedia.org/wiki/Gram_matrix), and we will see many similar forms in here, as those kind of structures are so minimal. However, when reading the literature, one needs to be very flexible in terms of tolerating matrix transposes and different conventions for notation—see, for example, https://en.wikipedia.org/wiki/Controllability_Gramian and https://en.wikipedia.org/wiki/Observability_Gramian.

[98] Image diffusion models, such as Midjourney, Stable Diffusion, Dall-E, Imagen, and Imagen Video.

[99] See also fn. 30 and the discussion in its vicinity, along with other ideas in this rather fragmented working paper. The formalisms of modern physics, for example, are way more structured and have a lot more finesse than these preliminary, scattered thoughts I have managed to scribe here, but it is evidently difficult to work backwards from physics to these kind of general isomorphisms pursued here—bra-ket notation is an example of developing a specialist notation for otherwise quite general ideas, that now refers so strongly to the quantum mechanics that it is rare to see it used for anything else, so the ideas stay confined to that domain, too, even if conceivably applicable to many other phenomena.

[100] See also fn. 86 and the discussion in that section.

isomorphism or congruence "≅", if possible). We will also keep an eye whether the occasion presents itself to add some frequency-space filtering in the model in the future, in some principled way.

### B. Gradients as smooth and forceful structures

Now, at last, using `https://www.matrixcalculus.org` to study the various gradients of the objective function $\mathcal{S}_2$ in Eq. (106), we arrive at the following:[101]

$$\frac{\partial \mathcal{S}_2}{\partial Y} = -V^\dagger V(Y - AXC)W^\dagger W =: -\tilde{E} \tag{108}$$
$$\frac{\partial \mathcal{S}_2}{\partial A} = V^\dagger V(Y - AXC)W^\dagger W(XC)^\dagger = \tilde{E}\,(XC)^\dagger \tag{109}$$
$$\frac{\partial \mathcal{S}_2}{\partial X} = A^\dagger V^\dagger V(Y - AXC)W^\dagger W C^\dagger = A^\dagger \tilde{E}\, C^\dagger \tag{110}$$
$$\frac{\partial \mathcal{S}_2}{\partial C} = (AX)^\dagger V^\dagger V(Y - AXC)W^\dagger W = (AX)^\dagger \tilde{E} \tag{111}$$
$$\frac{\partial \mathcal{S}_2}{\partial V} = -V(Y - AXC)W^\dagger W(Y - AXC)^\dagger \tag{112}$$
$$\frac{\partial \mathcal{S}_2}{\partial W} = -W(Y - AXC)^\dagger V^\dagger V(Y - AXC), \tag{113}$$

where on the right hand side, a convenience shorthand $\tilde{E}$ has been used for conciseness. Taking the gradients of the above, the second partial derivatives (the Hessians) of Eq. (106) are

$$\frac{\partial^2 \mathcal{S}_2}{\partial Y^2} = -(V^\dagger V) \overleftarrow{\otimes} (W^\dagger W) \tag{114}$$
$$\frac{\partial^2 \mathcal{S}_2}{\partial A^2} = -(XCW^\dagger W C^\dagger X^\dagger) \overleftarrow{\otimes} (V^\dagger V) \tag{115}$$
$$\frac{\partial^2 \mathcal{S}_2}{\partial X^2} = -(CW^\dagger W C^\dagger) \overleftarrow{\otimes} (A^\dagger V^\dagger V A) \tag{116}$$
$$\frac{\partial^2 \mathcal{S}_2}{\partial C^2} \cong -(X^\dagger A^\dagger V^\dagger V A X) \overleftarrow{\otimes} (W^\dagger W) \tag{117}$$
$$\frac{\partial^2 \mathcal{S}_2}{\partial V^2} = -I \overleftarrow{\otimes} \left[(Y - AXC)W^\dagger W(Y - AXC)^\dagger\right] \tag{118}$$
$$\frac{\partial^2 \mathcal{S}_2}{\partial W^2} = -I \overleftarrow{\otimes} \left[(Y - AXC)^\dagger V^\dagger V(Y - AXC)\right], \tag{119}$$

where $\overleftarrow{\otimes}$ is the Kronecker product. For $\partial^2 \mathcal{S}_2/\partial C^2$, `matrixcalculus.org` laments that "This 4th order tensor cannot be displayed as a matrix", so the form displayed here is only a structural approximation.

These, especially the gradients in Eqs. (108–113), are much simpler than they perhaps look on a first glance, and especially the shorthand forms are notable for their generality—remember that the dimensionalities here can be arbitrarily high, and these gradients depict optimal directions for the components of that objective function to evolve into, to minimize the residual error. We can study each of the gradients in Eqs. (108–113) in turn, to get some familiarity with their induced concepts.

#### 1. The importance of goal orientation

The form of $\partial \mathcal{S}_2/\partial Y$ in Eq. (108) means, that to minimize the error, the $Y$ could develop towards

$$\frac{\partial \mathcal{S}_2}{\partial Y} = -\tilde{E}, \tag{120}$$

which makes sense. One way to achieve goals, is to change them to be achievable. By setting $\partial \mathcal{S}_2/\partial Y = 0$, one finds that there is a stationary point $Y = AXC =: \bar{Y}_2$, where the overbar and a subscript records the stationary point for the objective function $\mathcal{S}_2$. The stationary point $\bar{Y}_2$ is the global maximum of $\mathcal{S}_2$ (with regards to $Y$), as the gradient is affine on $Y$ (so convex and concave), and Hessian in Eq. (114) is negative definite (Kronecker product factors being positive definite, negated by a minus sign). In other words, if $Y$ meets the estimate or prediction $AXC$ perfectly, error $E$ vanishes, and then there is not much more to be done.

#### 2. Distributed structural learning of factors

The form of $\partial \mathcal{S}_2/\partial A$ in Eq. (109) is interesting, as it shows that the optimal flow for the model matrix $A$ is towards the covariance of errors $\tilde{E}$ and representations $XC$,

$$\frac{\partial \mathcal{S}_2}{\partial A} = \tilde{E}\,(XC)^\dagger = \tilde{E}\, C^\dagger X^\dagger, \tag{121}$$

where the last form displays how $C$ could also be seen as a weighting matrix for covariances.

As the structure of the covariance matrix is so important to comprehend, here is an example using $2 \times 2$ matrices (weightings are ignored and $C$ is an identity matrix here for simplicity), that generalizes to arbitrary dimensions:

$$EX^\dagger = \begin{pmatrix} \epsilon_{11} & \epsilon_{12} \\ \epsilon_{21} & \epsilon_{22} \end{pmatrix} \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix}^\dagger \tag{122}$$
$$= \begin{pmatrix} \epsilon_{11} & \epsilon_{12} \\ \epsilon_{21} & \epsilon_{22} \end{pmatrix} \begin{pmatrix} x^*_{11} & x^*_{21} \\ x^*_{12} & x^*_{22} \end{pmatrix} \tag{123}$$
$$= \begin{pmatrix} \epsilon_{11}x^*_{11} + \epsilon_{12}x^*_{12} & \epsilon_{11}x^*_{21} + \epsilon_{12}x^*_{22} \\ \epsilon_{21}x^*_{11} + \epsilon_{22}x^*_{12} & \epsilon_{21}x^*_{21} + \epsilon_{22}x^*_{22} \end{pmatrix}, \tag{124}$$

where, clearly, each resulting element is an inner product between unordered "time series" (the rows) in matrices $E$ and $X$, and the (row,col)-indexing of the element denotes

[101] Note that we have replaced transposes with Hermitian conjugates (or Hermitian transposes in these finite dimensions where operators are represented by matrices), which we can always do, but cannot mix them with these tools available.

which rows in $E$ and $X$, respectively, are compared with each other by multiplying them element-wise together and summing them up to represent the correlation. The correlations could be scaled down by dividing them with the number of columns in $E$ (and $X$, as they have the same amount of occasions), but as a scalar multiplier, it would not affect the direction of the gradient. The rows are not centered, either, so the matrix represents "physicist's correlation"—we will return to these issues later.

Note also that in the Eq. (124), the first factor (column) orients towards how each error dimension correlates with the first diachronic dimension (row) of the representational form in $X$, the second factor (column) towards the correlation between each error dimension and second diachronic dimension (row) of the $X$, and so on, being rather principled and even local in its structure. Compare to Eq. (1), where now optimally the matrix $A$ is evolved towards the direction of $EX^\dagger$, resulting in forms resembling $EX^\dagger x$ and $EX^\dagger X$ (but not exactly those).

With this construction, there could be a mathematical reason for the correlational adaptation of neurons such as proposed by Hebb (2002 [1949])—emergence of covariance matrices could be the real structure that some systems optimally evolve towards, using the building blocks at their disposal. As the mathematical construction here is so general and minimal, even, this functional structure could be actually very prevalent in nature, isomorphic across various domains, and the possible universality of this approach is one of the main reasons I have kept studying this model of modeling, suggesting that these kind of generative structures could be continually molding various cohesive distributed forms in nature (including some structures in minds and societies).

As long as there is some error $E$, the model matrix $A$, guided by Eq. (109), will continue to adapt and learn to minimize the error (the residual), at least in the principal subspace of the errors (in terms of quadratic energy). If the error $E = 0$, then there is no learning needed. Note that also $X = 0$ is a stationary point, the system simply ceasing to exist (but even a small displacement or small amount of noise in $X$ may bring it forth again). Looking at the foundational Eq. (97), one intuitive way to think about the dynamics of the model matrix $A$ is that it cannot grow indefinitely, as that would bring the residual $E$ to zero or overshoot it (would start to increase the error in the opposite direction), thus a reasonable balance is found. The Hessian in Eq. (115) is again negative definite (also having a structure of a correlation matrix), indicative of a maximum.

Also, solving from Eq. (109), there is a stationary point $\partial \mathcal{S}_2/\partial A = 0$ at

$$\bar{A}_2 := YW^\dagger W(XC)^\dagger \left[XCW^\dagger W(XC)^\dagger\right]^{-1} \tag{125}$$

$$\cong Y(XC)^\dagger \left[XC(XC)^\dagger\right]^{-1}, \tag{126}$$

which is closely related to the pseudoinverse and the known optimal minimal solution to the linear least squares (which it formally is). Note that when comparing the formulas in the literature, one often needs to transpose the matrices, as the conventions may vary. The covariance matrices $Y(XC)^\dagger$ and $XC(XC)^\dagger$ are central here (the division by the number of occasions would cancel in the above $\bar{A}_2$, so they could be included in any case, and $W^\dagger W$ work as a diagonal weighting matrix).

It is notable that the stationary $\bar{A}_2$ does not need to be calculated explicitly (mostly the residual error $E$ is enough to calculate the gradients)—especially the matrix inverse emerges naturally from the stationary point of the dynamic process in Eq. (109), tolerant of noise and possible nonlinearities.

Note that in the above, there may be rotational freedoms (in the principal subspace), similar as practiced in factor analysis for decades already, that some experts in mathematical physics could perhaps see here as examples of some more fundamental symmetries or their induced invariant and conserved currents, that we do not yet see here with these formalisms.

### 3. Adaptive homeostatic state as representational mirroring

Then, the form of $\partial \mathcal{S}_2/\partial X$ in Eq. (110) is illustrative to think in terms of a differential equation that the flow directs to, to minimize the error or residual $E$:

$$\tau_X \frac{dX}{d\tau} = A^\dagger \tilde{E}\, C^\dagger \tag{127}$$

$$= A^\dagger V^\dagger V(Y - AXC)W^\dagger W C^\dagger \tag{128}$$

$$= A^\dagger V^\dagger V\, Y\, W^\dagger W C^\dagger - A^\dagger V^\dagger V A\, X\, C W^\dagger W C^\dagger \tag{129}$$

$$\cong A^\dagger Y C^\dagger - A^\dagger A\, X\, C C^\dagger, \tag{130}$$

which should be always stable (in terms of control theory) as in the latter terms, where $X$ is a factor, both $A^\dagger V^\dagger V A$ and $C W^\dagger W C^\dagger$ are positive definite (or at least positive semi-definite, being conjugate products), and the minus sign negates the term, resulting in negative feedback, eventually resulting in a stationary point $\bar{X}_2$ for any $Y$. Note also that if $C = I$ then the above differential equation applies to all the columns $x$ in $X$ (and respective columns $y$ in $Y$) independently, being nice and local.

When $C = I$ or $CC^\dagger = I$, there is also a possibility of "stratified modeling" by forcing the feedback matrix $A^\dagger A$ to be a upper or lower diagonal matrix, zeroing in other elements. Then the first (or last) column in $A$, driving the first (or last) row dimension in $X$, is not affected by anything but the data $Y$ and itself, going for the maximum variance, and the second (or second to last) then modeling the maximum variance in the residual, and so forth. In the future, perhaps there is a way to make this stratification dynamic, by utilizing coherence (zero phase difference), or constraints on positive and negative phase differences, or utilizing some other behavior based procedure or perhaps anti-symmetric operations on those matrices.

However, do keep in mind that often in practice, the continuous differential equation is applied as a discrete difference equation, where it is not enough for the eigenvalues to be negative, as the total feedback amplification must stay inside the unit circle in the complex plane, needing a suitable scale factor $\tau_X$ or adaptive time step for the whole difference equation (we will return to this in a short while).

Solving for the stationary point $\partial \mathcal{S}_2/\partial X = 0$ in Eq. (129) results in

$$\bar{X}_2 := (A^\dagger V^\dagger VA)^{-1} A^\dagger V^\dagger V\, Y\, W^\dagger W C^\dagger (C W^\dagger W C^\dagger)^{-1} \tag{131}$$

$$\cong (A^\dagger A)^{-1} A^\dagger\, Y C^\dagger (CC^\dagger)^{-1}, \tag{132}$$

which again displays a pseudoinverse (or two), especially when $C$ is the identity matrix (and when the diagonal weighting matrices $V$ and $W$ are hidden away as identity matrices for clarity). Hessian in Eq. (116) is negative definite, indicative of a global maximum.

Typically, we make a distinction between faster adaptation of $X$ and slower learning of $A$ here, separating the structures in this model of modeling. For each $Y$, the representation $X$ could be iterated for a few times using Eq. (129), as it will converge quite quickly to $\bar{X}_2$ or near it, minimizing the error or residual as much as possible in that linear subspace provided by the columns of $A$. Then, to make the model minimize the errors or residuals even better on average, Eq. (109) can be used to make a learning step on the model, adapting the structural factors.

Of course, in a real system, the adaptation and learning can happen continuously, simultaneously, and stochastically, but here the levels depicted already in Eq. (1)—ephemeral state $x$ and persistent model $A$—are kept separate as it is beneficial to tend to some kind of ideal structures for analysis and synthesis.

All this may seem very basic in comparison to the layered non-linear intricacies implicit in many modern machine learning frameworks, but the rationalization here to keep studying these, is that the structures are very explicit here, amenable for more rigorous analysis and further development. Some of those structures cannot be made any simpler using linear algebra, which should be a merit by itself—if there are isomorphisms between systems in different domains, they should be visible here, if anywhere at all. The more complex the structures are, the more implausible their realization in actual nature is.

The formulation of adaptation and learning as a continuous process should also benefit philosophical theorizing about nature—the discrete patterns emerge as stationary points of continuous processes, but the continuum can also be conceptualized as averaging of discrete elements. How the elements affect the whole, and how the whole affects the elements in turn, can be discussed in this framework in various ways, depending on how the terms “element”, ”whole”, ”affects”, ”in turn”, and ”how” are mapped to the various available structures in this model.

However, in any conventional computer, the systems need to be realized in discrete time. For simulating this system, it is indeed important to keep the eigenvalues of the feedback matrix inside the unit circle (on the complex plane).[102] One optimal way is to calculate the Frobenius norm of the feedback matrix (including the square root), as the spectral radius of a Hermitian matrix is bounded from above by its Frobenius norm, and use that as the time constant (perhaps multiplied by two for some margin of safety), dividing the whole difference equation. With that construction, the largest eigenvalue will necessarily stay inside the unit circle, and the iteration is remarkably fast and stable.[103] Usually in machine learning, one uses an adaptive time step using “AdamW” or similar, but here there is hope that some algebraic solution that is also naturally sound is eventually discovered.[104] In simulations, using the Frobenius norm (or its double) seems to work wonders when dividing Eq. (129), and its algebraic form is quite curious, being then

$$\tau_X := \sqrt{\mathrm{tr}(A^\dagger V^\dagger VAA^\dagger V^\dagger VA)\,\mathrm{tr}(CW^\dagger WC^\dagger CW^\dagger WC^\dagger)} \tag{133}$$

$$\cong \sqrt{\mathrm{tr}(A^\dagger AA^\dagger A)\,\mathrm{tr}(CC^\dagger CC^\dagger)}. \tag{134}$$

In practice, of course one just keeps track of the sum of squares of the relevant matrices, not needing to do all these matrix multiplications.

Note that in the feedback or self-interaction term, I am not sure about the importance of multiplying from the right by $CC^\dagger$ yet, as usually in control theory and

[102] I personally simulated these kind of systems already almost 15 years ago, see for example the animations at `http://neocybernetics.com/models/animations/`, but figured out that the distinction between discrete and continuous time is crucial in these adaptive systems, too, only a few years back.

[103] I originally found this out by experimentation, as my colleague in the industry noted that they frequently use grid search run over night to optimize hyperparameters, so I figured to try “symbolic grid search” manually for the problem of choosing time constants. I went in with the hypothesis that some simple combination of symbols would suffice, and eventually found out that the Frobenius norm worked well. Afterwards I found out the fact about the spectral radius of a Hermitian matrix being bounded from above by its Frobenius norm, and it gave some reasoning to the effectiveness of this procedure.

[104] Note also the similarity of dividing by the Frobenius norm to the divisor of the gradient if the square root is included in the objective function.

linear time invariant systems one deals with only one column vector $x$ and $y$ (or $u$) at a time with equations such as Eq. (129), where $C = I$, but here the $C$ mixes the columns in ways that are not usual. Also if the columns in the attention matrix each sum to one, perhaps the Frobenius norm is not the best estimator for the effect of that matrix.

Also observe that in Eq. (129), there is an "inflow" or "interpretation" $A^\dagger YC^\dagger$, and a "dissipation" or "self-critique" $-A^\dagger AXCC^\dagger$ (ignoring the diagonal weighting matrices). Symmetrically, there could also be some "outflow" $-BZC^\dagger$, or similar, but we do not explore that here. The form of that optimal inflow is very illustrative, in that reading the term from right to left, there $Y$ (or any column $y$) is simply projected via $A^\dagger$ (i.e. projected to the factors taking conjugated inner products), to add to the derivative for that column (and $C$ mixes the occasions, if present). Even more illustratively, in Eq. (127), the error or residual $\tilde{E}$ is simply projected via the factors or patterns in $A^\dagger$, to add to the derivative for $X$, accumulating the representative form, that then decreases the residuals. Formally, the factors in $A$ dictate what can be perceived, and they themselves cohere as significant categories "in the world" (in this extremely reductionist sense).[105]

There is a possibility of identifying this kind of residual matching behavior in some system, and then working backwards to the system-centric objective function (or simply inferring its existence), as the mathematical transformations and relations here can be written and read both ways. The residual can be a copy or echo of some physical process, or it can be more directly the physical process *an sich*,[106] which is then controlled towards zero (causing the target $Y$ to be followed in average).

#### 4. Combining occasions to amplify signals and filter noise

The above has been simulated (omitting the weighting and attention matrices), finding the modeling behavior stable and interesting, but Eq. (111) suggests that also attention matrix $C$ could be optimized, to minimize the residual errors. The structure of the gradient $\partial \mathcal{S}_2/\partial C$ is interesting, as there the inner product matrix of estimates $AX$ and weighted errors $\tilde{E}$ provides the guidance, which is equivalent to the inner product matrix of the states $X$ and "inflow" or "interpretation" $A^\dagger \tilde{E}$:

$$\frac{\partial \mathcal{S}_2}{\partial C} = (AX)^\dagger \tilde{E} = X^\dagger A^\dagger \tilde{E} \tag{135}$$

The inner product matrix is dual to the outer product matrix, in the sense that outer product matrix matches the rows ("time series") in matrices, resulting in a covariance matrix structure, but the inner product matrix matches the columns (occasions), resulting in different kind of a Gram matrix. In a very concrete sense, the attention matrix $C$ is thus evolved towards where the estimates and residuals match synchronically or "spatially", so that when the attention matrix $C$ is used as a multiplier from the right in the foundational Eq. (97), the occasions are mixed together to support the estimates matching the residuals even better. Perhaps the attention is directed to where there is both anticipation (in terms of estimates $AX$) and need (in terms of residual $\tilde{E}$), both in the column space—or even more simply, where the states $X$ and residual perceptions $A^\dagger \tilde{E}$ match (column-wise). If the occasions cannot be compared diachronically, then only the diagonal (or the trace) is important here.

Here, as I have not simulated this, I am not sure about the stability, even though it looks like the feedback matrix is stable:

$$\begin{aligned}
\tau_C \frac{dC}{d\tau} &= X^\dagger A^\dagger \tilde{E} && (136)\\
&= X^\dagger A^\dagger V^\dagger V(Y - AXC)W^\dagger W && (137)\\
&= X^\dagger A^\dagger V^\dagger V\, Y\, W^\dagger W - X^\dagger A^\dagger V^\dagger VAX\, C\, W^\dagger W && (138)\\
&\cong X^\dagger A^\dagger Y - X^\dagger A^\dagger AXC, && (139)\\
& && (140)
\end{aligned}$$

where in the latter feedback terms both $(VAX)^\dagger VAX$ and $W^\dagger W$ are obviously positive (semi-)definite, likely resulting in stable negative feedback (but there may be complications due to different dimensionalities and matrix ranks, and also dynamic interactions between different processes, as everything could be continuously evolving).

Solving for the stationary point $\partial \mathcal{S}_2/\partial C = 0$ in Eq. (138),

$$\begin{aligned}
\bar{C}_2 &:= (X^\dagger A^\dagger V^\dagger VAX)^{-1} X^\dagger A^\dagger V^\dagger V\, Y && (141)\\
&\cong (X^\dagger A^\dagger AX)^{-1} X^\dagger A^\dagger Y, && (142)
\end{aligned}$$

which has again the structure of a pseudoinverse. Hessian in Eq. (117) is again negative definite (but that is only my educated guess here, as the 4th order tensor is not visible), implying a global maximum.

However, as $\bar{C}_2$ depends on $X$, and $\bar{X}_2$ on $C$, it is not clear how the process should be orchestrated, and at what time scales. One principled way could be to first adapt the $X$ with identity matrix $C$, to match $Y$ as well as possible with the factors in $A$, and then adapt $C$ to

---

[105] In semiotics, these are also related to the concepts of "umwelt" and "innenwelt", and "merkwelt" and "wirkwelt". See, for example, `https://en.wikipedia.org/wiki/Umwelt`.

[106] I am using these terms rather ironically here, as I really cannot define what the physical process "itself" is.

fine tune the matches (but I do not have much intuition here yet). In transformer models, the $C$ itself consists of $(KX)^\dagger QK$ products, for each $X$. Perhaps clearest would be to constrain the columns of $C$ to sum up to one, after exponentiation (to prevent negative contributions), so that it would have similar constraints than attention matrices in transformers, and could be used in a more relaxed way (as stability of $C$ would be provided by normalization, so $X$ and $C$ could perhaps both be evolved to the stationary points). This could be investigated more in some future work.

#### 5. Unknown dynamics for the weightings

The last two gradients, $\partial\mathcal{S}_2/\partial V$ and $\partial\mathcal{S}_2/\partial W$ [in Eqs. (112) and (113), respectively], are of course not of much help, as similar to the gradient with respect to $Y$, also weightings could be simply attenuated to zero to decrease the error. The forms of the gradients suggest that they deal with the inner product matrix and the outer product matrix of residual errors $E$—note that for simplicity, I have not added restrictions to the objective function for their forms, so only the diagonal of these gradients should be used, so sum of squares of columns or rows (depending on the matrix in question). I do not know would it in some settings be reasonable to actually increase the weightings depending on these error metrics, to direct more weight towards the unexplained, but the forms of the gradients do not suggest any nice stationary points or dynamic behavior yet.

#### 6. Normalizing the mixing matrix columns to density functions

I have not simulated the following, but formalizing the above suggestion to normalize the columns in the attention matrix $C$ to one, seems to cause (or be equal to) subtraction of column (occasion) expectation values when calculating the gradients. Defining the maximized objective function to be (using subscript, $E_3$, for this residual error for clarity)

$$\mathcal{S}_3 = -\|VE_3W^\dagger\|_2^2/2 \tag{143}$$
$$= -\|V\{Y - AX[C \oslash (JC)]\}W^\dagger\|_2^2/2, \tag{144}$$

where $\oslash$ is element-wise division and $J$ a matrix of ones, collecting the column sums when multiplying from the left [see Eq. (11)]. The other gradients do not have any new factors, of course, but for column-normalized $C$,

$$\frac{\partial\mathcal{S}_3}{\partial C} = \Big((AX)^\dagger\tilde{E}_3 - J\{[(AX)^\dagger\tilde{E}_3]\odot[C\oslash(JC)]\}\Big) \oslash (JC), \tag{145}$$

where

$$\tilde{E}_3 := V^\dagger V\{Y - AX[C\oslash(JC)]\}W^\dagger W \tag{146}$$
$$\cong \{Y - AX[C\oslash(JC)]\}, \tag{147}$$

still similar to Eq. (111). The above Eq. (145) means, that in comparison to $\partial\mathcal{S}_2/\partial C = (AX)^\dagger\tilde{E}$, the gradient with respect to the column-normalized $C$ in $\mathcal{S}_3$ is the same, but (1) normalized $C$ is used as a density function (by multiplying the gradient element-wise) to aggregate expectation values of each column (by multiplying with matrix of ones J from the left), which is then subtracted from the gradient, effectively centering the columns, and (2) the gradient is divided column-wise by the column sums of $C$ (which is reasonable, as $C$ is used as column-normalized in any case).

This procedure of normalization, leading to the centering of the gradient columns around their “spatial” expectation values (and vice versa), may have many interesting properties, but this is very tentative as it has not been validated by simulation, for example—note also that there is no exponentiation here, so the columns of normalized $C$ are not actually density functions as they can contain negative values, they just sum to one. If a column would sum to zero, the division would thus present singularities in the model, which I have tried to avoid here by assuming that there is always some noise in everything, however small, and it will drive the systems to their stable points—or if systems vanish to zero or explode to infinities, then they are not locally persistent systems for modeling purposes and can be ignored here.

This is also perhaps the first occasion where the actual dimensions of the matrices start to affect the model—if these matrices are developed abstractly in some infinite-dimensional setting, for example, this would perhaps warrant some discussions on renormalization or regularization. But here, working with these finite discrete matrices, we can even take some dimensionalities explicitly as factors to the model, and that is the idea we will explore next (and lastly).

#### 7. The mean field and modeling of variations

If one simulates this minimal modeling system, even if using the optimal time constant in Eq. (133), one still often encounters instabilities due to all the factors aiming and competing for the common mean. Usually in modeling, this problem is alleviated by including a learned bias constant, that is of course of crucial importance in practice, especially when nonlinearities such as rectifying (cutting negative values to zero) are present.

There is a perhaps surprising way to achieve a kind of inertial centering of the model, without explicitly introducing bias terms, and it involves optimizing the objective function again from a larger perspective, the meaning and implications of which is not clear to me, even though the formulas are quite straightforward.

The procedure is as follows:[107] (1) Take the objective function that aims to minimize the error norm (or maximize its negative), and express it in trace form $\operatorname{tr} E^\dagger E$, similar to Eq. (107). (2) Now wonder that the other off-diagonal terms of the inner product matrix $E^\dagger E$, that are discarded in the trace, are about correlations of errors between *different* occasions, and actually add the total sum of all those elements (multiplied by a scalar $\kappa$, kappa) to the objective function, aiming to maximize it. This turns the optimization task to multi-objective (or multiple criteria) optimization problem, where the sum of squared errors is minimized as before, but also a sum of conjugate products between different occasions of residual errors is maximized (and there may be canceling due to ordering of products in the upper diagonal vs. lower diagonal elements), the scalar $\kappa$ providing a weighting between these different objectives (and could also be negative in some settings).

More precisely, the augmented objective function to be studied is

$$\mathcal{S}_4 = -\|VEW^\dagger\|_2^2/2 + \kappa\,\mathrm{sum}(WE^\dagger V^\dagger VEW^\dagger)/2 \quad (148)$$

$$= -\|V\,(Y - AXC)\,W^\dagger\|_2^2/2 + \kappa\,\mathrm{sum}\left[W\,(Y - AXC)^\dagger\,V^\dagger V\,(Y - AXC)\,W^\dagger\right]/2 \quad (149)$$

$$= -\mathrm{tr}\left[W\,(Y - AXC)^\dagger\,V^\dagger\,V\,(Y - AXC)\,W^\dagger\right]/2 + \kappa\,\mathrm{sum}\left[W\,(Y - AXC)^\dagger\,V^\dagger V\,(Y - AXC)\,W^\dagger\right]/2. \quad (150)$$

One could also develop the above further starting from $\mathcal{S}_3$ of Eq. (144) instead of $\mathcal{S}_2$ of Eq. (106), but this should be clear enough as a concept.

The surprising finding (to me at least) is, that by taking the gradients of the above, the structures can be centered (depending on the value of $\kappa$), as the gradients are

$$\frac{\partial \mathcal{S}_4}{\partial Y} = -V^\dagger V(Y - AXC)W^\dagger(I - \kappa J)W =: -\tilde{E}_4 \quad (151)$$

$$\frac{\partial \mathcal{S}_4}{\partial A} = V^\dagger V(Y - AXC)W^\dagger(I - \kappa J)W(XC)^\dagger = \tilde{E}_4\,(XC)^\dagger \quad (152)$$

$$\frac{\partial \mathcal{S}_4}{\partial X} = A^\dagger V^\dagger V(Y - AXC)W^\dagger(I - \kappa J)WC^\dagger = A^\dagger \tilde{E}_4\,C^\dagger \quad (153)$$

$$\frac{\partial \mathcal{S}_4}{\partial C} = (AX)^\dagger V^\dagger V(Y - AXC)W^\dagger(I - \kappa J)W = (AX)^\dagger \tilde{E}_4 \quad (154)$$

$$\frac{\partial \mathcal{S}_4}{\partial V} = -V(Y - AXC)W^\dagger(I - \kappa J)W(Y - AXC)^\dagger \quad (155)$$

$$\frac{\partial \mathcal{S}_4}{\partial W} = -(I - \kappa J)W(Y - AXC)^\dagger V^\dagger V(Y - AXC), \quad (156)$$

where $I$ is the identity matrix and $J$ the matrix of ones.

Note that as the trace is circular but the sum is not, the inner product of errors $\operatorname{tr} E^\dagger E$ could also be equivalently rotated to be an outer product of errors $\operatorname{tr} EE^\dagger$, and then the special matrix $(I-\kappa J)$ would appear in between $V^\dagger V$ in the above gradients, affecting the columns, not $W^\dagger W$, affecting the rows, as in the above. I do not yet know how to interpret this, or whether they could be used at the same time, possibly mixing or oscillating between them.

The useful property of the above is, that if the matrix $Y$ has $k$ columns (occasions), then choosing $\kappa = 1/k$ results in the centering matrix $I - J/k$.[108] Multiplying a matrix with it from the right, subtracts the means of the rows from the matrix. It is also of course an idempotent projection matrix, as operating with it again, there is no change—this also means that one can insert multiple centering matrices next to each other in the above, also transposed (as it is symmetric), showing how the structures are centered in transposed form (as transposes reverse the order of products).

Here are some equivalent forms of that special matrix,

[107] I found this procedure by experimentation, around the same day I found out about the algebraic form of the optimal time constant a couple of years back. Of course, these kind of forms may have been studied already decades ago, but it is quite difficult to search for mathematical formulas in the literature, as the symbols and conventions can vary. Self-contained expositions also have merit, the concepts being concise enough that they can be derived again and again from different perspectives and in different contexts, without a multitude of lemmas and references to theorems in the literature.

[108] See `https://en.wikipedia.org/wiki/Centering_matrix` and `https://en.wikipedia.org/wiki/Matrix_of_ones`.

when $\kappa = 1/k = 1/3$, but this behavior is general:

$$I - \kappa J = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \frac{1}{3} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \tag{157}$$

$$= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \frac{1}{3} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \end{pmatrix} \tag{158}$$

$$= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \end{pmatrix} \tag{159}$$

$$= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix} \tag{160}$$

$$= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix} \tag{161}$$

$$= \begin{pmatrix} \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & -\frac{1}{3} & \frac{2}{3} \end{pmatrix} \tag{162}$$

Inspecting the above gradients, such as $\partial \mathcal{S}_4 / \partial A$ in Eq. (152), it is evident that the covariance matrix will then be centered (and vice versa, if one centers the covariance matrix, the objective function with the sum term is implicit). It is also evident that one does not need to center both the error $\tilde{E}_4$ and representation $XC$, as it is necessary to center only one of them to achieve the same covariance matrix (which feels odd). Also by inspection, the centering can be done by subtracting the outer product of the mean error and the mean representation, which has an interesting element-wise "emanating" or inertial feel to it—compare also to the well-known formula for covariance of random variables X and Y:

$$\begin{aligned} \mathrm{cov}(\mathrm{X}, \mathrm{Y}) &= \mathrm{E}\left[(\mathrm{X} - \mathrm{E}[\mathrm{X}])(\mathrm{Y} - \mathrm{E}[\mathrm{Y}])\right] \\ &= \mathrm{E}\left[\mathrm{XY}\right] - \mathrm{E}[\mathrm{X}]\,\mathrm{E}[\mathrm{Y}], \end{aligned} \tag{163}$$

and also compare to how the moments can be calculated by exponentiation and derivative at zero in Eq. (3). There seems to be a possibility of getting to the roots of some modeling pattern in here.

### 8. Deferred studies

Also inspect what happens if the objective function Eq. (150) is divided by $k$ (when that special $\kappa = 1/k$), as then the first trace term is the average of the diagonal, and second sum term is the average of all the elements in the matrix (having $k^2$ elements), resulting in proper, centered covariance matrices emerging in Eq. (152). I may need to explain it here later in more detail, as it may be difficult to see how the $1/k$ distributes so nicely, but it does, and I do not know if this is indicative of something fundamental or simply trivial.

In simulations, the centering terms for gradients are very simple to calculate, and they are effective in stabilizing the system. In many ways, keeping track of the means is also physically quite plausible, as the simplest low-pass aggregation or center there is.

The above could also be studied further by experimenting with exponentiation [Eq. (11) or perhaps exponentiating the whole objective function, deriving the terms when differentiating at zero], or perhaps exponentiating element-wise, extending towards Laplace transform and frequency-space, as the mappings are rather linear here. There is also a possibility of interpreting some of these concepts in terms of intensive and extensive properties, and also some variables could be interpreted as velocities or other derivatives, so the covariance matrices would represent inertia matrices (especially with the centering). Somehow taking successive derivatives of the residuals and controlling them to zero may be a key to a lot of the functionalities here (as suggested by Hyötyniemi in his later studies).

We will need to cut this treatise here, as it seems this became already way too long-winded—but hopefully not yet "a volume, and an uninviting one".

## VI. CONCLUSIONS

This study has been a quick one, with various threads interwoven. It is my hope that some of these thoughts could turn out to be enriching and useful, even with their obvious flaws and deficiencies. Though the case for the centrality of modeling in nature may not yet have been presented in its fullest cogency, perhaps there are people who will find these ideas already even trivial, and could bring higher-level understanding to the concepts, condensing them to some very simple mathematics, for example. I really would want to know—as I am sure many others, who find themselves at the various borderlands between the exact sciences and the humanities, would also like to see some new insights and real development on these matters. I hope that with the findings of the future, it would be perfectly natural to witness life being beautified by the wonderful soulful creations of the motion, spirit and the intellect.

## ACKNOWLEDGMENTS

The author wishes to acknowledge the financial support of Screen.io Corporation during the preparation of this study. The author also acknowledges a donation from Heikki Hyötyniemi, greatly facilitating this work.

**Appendix A: On the creative process of writing: case Dostoevsky**

About the creative process, Yuri M. Lotman has written the following, which I found so fascinating at this point that I could not end quoting it until it became so long that it had to be moved to the appendix. So the following paragraphs are reproduced from Lotman (2001 [1990], pp. 73–77) (each emphasis in the original).

"...there is plenty of evidence to suggest that the first link in the chain is as a rule a symbol (even when artists and writers speak of a sound or even a scent as the 'kernel' of the future text, they are referring, as we shall be demonstrating below, to a symbolic expression of an individual semiotic process, for instance to a childhood associative symbolization, or a crucial moment in their emotional biography, and so on). In other words, the artistic function is present from the very beginning, even if only as a potentiality. Take for example the evidence of Dostoevsky. Dostoevsky's insistence is striking that the creation of the original theme of a novel was the most artistically significant part of his work, calling it 'poet's work'. The development of the theme he called 'artist's work', using the work artist in the sense of 'craftsman'. See his note in the drafts to the novel *A Raw Youth*:

> In order to write a novel an author has to be provided with *one* or *several* strong impressions which have actually been experienced emotionally. *This is the poet's work.* The theme, the plan, the structured whole, develops from this impression. This is now the work of the artist, though artist and poet help each other at both stages, in both cases.

Dostoevsky returned many times to this topic. He even called this primary 'theme' of a novel a poem to emphasize its poetic nature. On 15/27 May 1869 he wrote to Apollon Maikov:

> in my opinion a poem is like a unique precious stone, a diamond in the poet's heart, something ready made, in all its essence, and this is the first work of the poet as creator and originator, this is the first part of his creative work. If you like, it's not even he who is the creator, but life, the powerful essence of life.

And further: 'Then follows the poet's second work no longer so profound and mysterious, the work merely of the artist, namely, having acquired the diamond to set it and mount it.' And in another letter he wrote: 'Being more of a poet than an artist I'm for ever taking on themes which are beyond my capabilities.' Pushkin had the same thing in mind when he wrote that even the plan of Dante's *Inferno* was a creation of artistic genius (the notion of 'plan' was for Pushkin similar to Zholkovsky and Shcheglov's 'theme'). We can therefore conclude that there is authoritative evidence to show that the chain which generates the literary text begins both psychologically and logically not with a logically expressed, non-literary 'theme' but with a capacious symbol with the potential to develop many images and interpretations, a symbol which is already literary.

[...]

Our second substantial objection is connected with the notion of the symmetry of the 'poetics of expressiveness' model. We have already had occasion to state our belief that the generation of new meanings is always connected with asymmetrical structures. While the preservation of information is most reliably ensured by symmetrical structures, the generation of information involves asymmetrical mechanisms. When asymmetrical binarity is discovered in a semiotic object this always presupposes some form of intellectual activity. We cannot envisage the generation of a literary text as an automatic working of a single, set algorithm. The creative process is an irreversible process (see below, Part Three [in the book]), and hence the passage from one stage to another must involve elements of randomness and unpredictability.

[...]

An examination of the actual creative process when a writer's manuscripts make it possible to document this gives convincing weight to our argument.

A study of the logical aspect of the creative process is not capable of reproducing the strange paths taken in the creation of any actual work, but on the other hand, it should not ignore the typical stages in the generation of actual texts when we are able to follow this process in sufficient detail. Moreover we suggest that the actual process might serve as a criterion for verifying our logical models, and the logical models as a means for interpreting textological realities.

A regular feature which can be deduced from a study of the working manuscripts of many writers is that the stages succeed one another: intention is followed by narration. In this process, stress on symbolic, polyvalent, multi-dimensional text-semantics gives way to a striving for precise expression of a thought. On the boundaries between these stages relationships of asymmetry and untranslatability come into being, and this process entails the generation of new meanings.

We mentioned above that the first stage in the generation of the text is like the emergence of a primary symbol, whose capacity is proportional to the range of potential plots concealed in it. This is why, when headings and epigraphs are defined, these seemingly marginal points can be the signal that the 'theme' (in Zholkovsky and Shcheglov's terminology) has been defined. For instance, because Dostoevsky had to work on several projects at once, because of the richness of his imagination, and the integral connection between his various projects, it is in practice impossible to tell which of the several plots he was working on at the same time a particular manuscript text relates to.

[...]

This symbol [omitted here for brevity] as it were lit up the rudiments of the plots he already had and pointed the way to their future development. *The various drafts and half-worked projects were sorted out into a story.* Thus the concentration of many things in the one symbol was replaced by a linear development of the one symbol into various episodes.

This changeover, if we continue looking at the history of the writing of *The Devils*, is expressed in Dostoevsky's plans, his summary enumeration of episodes which thread themselves along the syntagmatic axis of the narrative. However, as soon as this tendency to *exposition* or narrative construction can be observed, we are witness also to a growing inner opposition to this tendency. Each serious movement of the plot Dostoevsky immediately smothers with variants and alternative versions. The wealth of Dostoevsky's imagination which allows him to 'play over' a vast quantity of possible story-lines, is truly amazing. The text in fact loses its linearity. It turns into paradigmatic set of possible lines of development. And the same thing happens at almost every turning point in the plot. The syntagmatic construction is replaced by a multidimensional space of plot potentialities. When this happens the text becomes harder to fit into verbal expression: we have only to look at a page of Dostoevsky's manuscripts to see how far the writer is at this stage from writing a 'normal' narrative text. Phrases are tossed on to a page without any temporal sequence being observed in the way he fills up lines and sheets of paper. There is no guarantee that two lines set next to each other were written one after the other, rather the reverse. Words are written in different handwriting and in different sizes, at different angles. A page looks like a wall of a cell on which a prisoner has at different times scribbled his feverish jottings which for him have some inner associations, but which for the outside observer seem unconnected. Many of the jottings are not texts, but mnemonic abbreviations of texts preserved in the author's mind. Thus Dostoevsky's manuscript pages at this stage tend to become signs of a vast multi-dimensional whole living in the author's mind, rather than a logical exposition of a linearly organized text. Besides, these jottings deal with many levels: here we shall find philosophical nature, and isolated symbol-worlds which have not yet found a place but which will be unravelled into future episodes which the author's imagination has not yet created. Dostoevsky uses different means of emphasis—he underlines, he writes in large letters, in printed script, for at this stage in the work he is consciously recording the intonation, as if stressing that his graphics are not a text but just a projection of one.

The next stage follows when he extracts linear elements out of this continuum and constructs narrative text. Linearity takes the place of multi-dimensionality. The preceding stage was marked by an abundance of rich symbols which opened the way to the most diverse concretizations in the narrative web of the future novel. For instance, the word 'slap', which is a powerful symbol for Dostoevsky, often occurs in the preliminary materials to *The Devils*. Already in 'Kartuzov', an early version of *The Devils*, this word occurs in the title and is emphasized in his handwriting. Later on in the preliminary material for *The Devils* (and subsequently in that for *A Raw Youth*), the circumstances surrounding the 'slap', and who gave one to whom, change, but the slap itself remains as a symbol of utter humiliation. A symbol may determine a cluster of possible plot developments, but it cannot determine which one of them will be chosen. In the same way the little red spider which appears in 'Stavrogin's Confession' (not part of the final text of the novel) and which the hero looks at while his victim is hanging herself, turns up in the preparatory material to 'A Raw Youth' as a conventional sign for a whole set of situations which the author's imagination has produced.

This is how the relationship between the preparatory material and the subsequent narrative text is shaped. This relationship is like a ball of wool and the thread unwound from it: the ball exists spatially and in a particular single time, while the thread is unwound from it in a temporal movement, linearly.

We can represent the process which Dostoevsky follows to create a text by the following diagram: [An alternating process between symbolic/iconic/non-discrete/spatial and verbal/narrative/discrete/temporal, starting from initial symbol, devising a discrete plan, exploring preparatory material, and materializing into text.]

So the 'generation' of a text involves numerous semiotic transformations. On the boundary between different semiotic regimes (at the intersections with the zero line) an act of translation takes place and there is a not wholly predictable reformulation of meanings.

We must stress that we are talking of a logical model and not describing an actual creative process, since it is frequently impossible in practice to isolate the 'initial symbol' moments in the continuous cross-weaving of Dostoevsky's intentions for his novels. In the same way the 'preparatory material' always includes portions of narrative, and the narrative texts tend, in the continuously corrected and reworked drafts, to turn into preparatory material, so that the distinction between them is conventional and logical rather than a matter of fact. Even the boundaries which separate one of Dostoevsky's novels from another are frequently blurred. Tomashevsky wrote that 'Dostoevsky writes novel after novel in the search for one single novel', while the most recent researchers who have studied Dostoevsky's notebooks rightly comment that 'it seems more expedient to talk of a single draft distributed over the sequence of the different stages of his writing'.

Iconic (spatial, non-discrete) texts and verbal (discrete, linear) ones are mutually untranslatable, and cannot in principle express 'one and the same' content. At the points where they confront each other there is an increase in indeterminacy and this creates a reserve for more information. In the process of creating a text, then, a writer is doing several things at once: s/he creates a channel from the huge amount of potential material available (tradition, associations, his/her own previous works, texts from surrounding life, etc.); s/he passes the new texts which arise in his/her creative imagination through this channel, leading them across the transformational thresholds, and increasing their semantic load on account of unexpected combinations, translations, linkages, etc. When as a result of all this a structurally organized dynamic whole takes shape, we can say that the text of the work has appeared."

## REFERENCES


Ahl, V, and T. F. H. Allen (1996), *Hierarchy Theory: A Vision, Vocabulary, and Epistemology* (Columbia University Press).

Alexander, C (1964), *Notes on the Synthesis of Form* (Harvard University Press).

Alexander, C (1999), "The origins of pattern theory: the future of the theory, and the generation of a living world," IEEE Softw. **16** (5), 71–82.

Alexander, C (2005), *The Nature of Order: An Essay on the Art of Building and the Nature of the Universe (4 vols.)* (The Center for Environmental Structure, Berkeley).

Banks, E C (2014), *The Realistic Empiricism of Mach, James, and Russell. Neutral Monism Reconsidered* (Cambridge University Press, Cambridge).

Baumeister, R F, E. Bratslavsky, C. Finkenauer, and K. D. Vohs (2001), "Bad is stronger than good," Rev. Gen. Psychol. **5** (4), 323–370.

Berger, P L, and T. Luckmann (1966), *The Social Construction of Reality: A Treatise in the Sociology of Knowledge* (Anchor Books, New York).

Bohm, D (2002 [1980]), *Wholeness and the Implicate Order* (Routledge).

Bohm, D, and F. D. Peat (2010 [1987]), *Science, Order and Creativity* (Routledge).

Bongard, J, and M. Levin (2023), "There's plenty of room right here: Biological systems as evolved, overloaded, multi-scale machines," Biomimetics **8** (1), 10.3390/biomimetics8010110.

de Botton, A (2003), *The Art of Travel* (Penguin Books, London).

Bourdieu, P (1977), *Outline of a Theory of Practice* (Cambridge University Press).

Bourdieu, P (1984), *Distinction: A Social Critique of the Judgement of Taste* (Harvard University Press).

Bruner, J (2003), *Making Stories: Law, Literature, Life* (Harvard University Press).

Bubeck, S, V. Chandrasekaran, R. Eldan, J. Gehrke, E. Horvitz, E. Kamar, P. Lee, Y. T. Lee, Y. Li, S. Lundberg, H. Nori, H. Palangi, M. T. Ribeiro, and Y. Zhang (2023), "Sparks of Artificial General Intelligence: Early experiments with GPT-4," arXiv:2303.12712 [cs.CL].

Comte-Sponville, A (2002), *A Small Treatise on the Great Virtues: The Uses of Philosophy in Everyday Life* (Owl Books, U.S.) orig. in French published by Presses Universitaires de France, Paris.

Daft, R L, and K. E. Weick (1984), "Toward a model of organizations as interpretation systems," Acad. Manag. Rev. **9** (2), 284–295.

Danesi, M (2017), "The concept of model in Thomas A. Sebeok's semiotics," in *Proceedings of the 12th World Congress of the International Association for Semiotic Studies (IASS/AIS)*, edited by K. Bankov (IASS Publications and NBU Publishing House, Sofia) pp. 1495–1506.

Deely, J (2001), *Four Ages of Understanding: The First Postmodern Survey of Philosophy from Ancient Times to the Turn of the Twenty-First Century* (University of Toronto Press).

Ferraro, F, J. Pfeffer, and R. I. Sutton (2005), "Economics language and assumptions: How theories can become self-fulfilling," Acad. Manag. Rev. **30** (1), 8–24.

Feynman, R P (1955), "The value of science," Engineering and Science **19** (3), 13–15.

Feynman, R P, and R. Leighton (1985), *"Surely You're Joking, Mr. Feynman!": Adventures of a Curious Character*, edited by E. Hutchings (W. W. Norton, New York).

Frankfurt, H G (2005), *On Bullshit* (Princeton University Press).

Giddens, A (1984), *The Constitution of Society: Outline of the Theory of Structuration* (Polity Press, Cambdridge, UK).

Gilbert, S F, and S. Sarkar (2000), "Embracing complexity: Organicism for the 21st century," Developmental Dynamics **219** (1), 1–9.

Gärdenfors, P (2004), *Conceptual Spaces: The Geometry of Thought* (MIT Press).

Hebb, D O (2002 [1949]), *The Organization of Behavior: A Neuropsychological Theory* (Lawrence Erlbaum Associates).

Heiskala, R (2003), *Society as Semiosis: Neostructuralist Theory of Culture and Society* (Peter Lang, Frankfurt am Main).

Ho, J, W. Chan, C. Saharia, J. Whang, R. Gao, A. Gritsenko, D. P. Kingma, B. Poole, M. Norouzi, D. J. Fleet, and T. Salimans (2022), "Imagen video: High definition video generation with diffusion models," arXiv:2210.02303 [cs.CV].

Hoffmann, J, S. Borgeaud, A. Mensch, E. Buchatskaya, T. Cai, E. Rutherford, D. de Las Casas, L. A. Hendricks, J. Welbl, A. Clark, T. Hennigan, E. Noland, K. Millican, G. van den Driessche, B. Damoc, A. Guy, S. Osindero, K. Simonyan, E. Elsen, J. W. Rae, O. Vinyals, and L. Sifre (2022), "Training compute-optimal large language models," arXiv:2203.15556 [cs.CL].

Hyötyniemi, H (2006), *Neocybernetics in Biological Systems*, Report 151 (Helsinki University of Technology, Control Engineering Laboratory, Otaniemi, Espoo).

Jackall, R (2010 [1988]), *Moral Mazes: The World of Corporate Managers – Twentieth Anniversary Edition* (Oxford University Press).

Jones, D T (2011), *This Being, That Becomes: The Buddha's Teaching on Conditionality* (Windhorse Publications, Cambridge) Transl. in Finnish, 2013, Kun on tämä, syntyy tuo: Buddhan opetus ehdonvaraisuudesta.

Kalupahana, D J (1975), *Causality: The Central Philosophy of Buddhism* (University of Hawaii Press).

Kurt, B, and T. Leise (2006), "The $25,000,000,000 eigenvector: The linear algebra behind Google," SIAM Rev. Soc. Ind. Appl. Math. **48** (3), 569–581.

Laue, S, M. Mitterreiter, and J. Giesen (2018), "Computing higher order derivatives of matrix and tensor expressions," in *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, pp. 2755–2764.

Laue, S, M. Mitterreiter, and J. Giesen (2020), "A simple and efficient tensor calculus," in *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 4527–4534.

Lebaron, F (2009), "How Bourdieu "quantified" Bourdieu: The geometric modelling of data," in *Quantifying Theory: Pierre Bourdieu*, edited by K. Robson and C. Sanders, Chap. 2 (Springer) pp. 11–29.

Lievonen, P (2023), "First impressions on Suntola's expanding 3-sphere cosmological model: A motivational study surveying the Dynamic Universe," A working paper in Appearance and Reality in Physics and Beyond -collaboration, 2023.3.29 (unpublished).

Lotman, Y M (2001 [1990]), *Universe of the Mind: A Semiotic Theory of Culture* (I.B.Tauris Publishers, London).

Mach, E (1959 [1886]), *The Analysis of Sensations and the Relation of Physical to the Psychical*, 5th ed. (Dover, New York).

Mach, E (1976 [1905]), *Knowledge and Error*, 5th ed. (Reidel, Dordrecht).

Marder, E, and D. Bucher (2001), "Central pattern generators and the control of rhythmic movements," Curr. Biol. **11** (23), R986–R996.

Maurer, B (2006), "The anthropology of money," Annu. Rev. Anthropol. **35**, 15–36.

Mead, G H (2015 [1934]), *Mind, Self, and Society* (University of Chicago Press, Chicago).

Mehra, J (2001), *The Historical Development of Quantum Theory 1–6* (Springer, New York).

Midgley, G, Ed. (2003), *Systems Thinking* (Sage Publications).

Mitchell, T M, S. V. Shinkareva, A. Carlson, K.-M. Chang, V. L. Malave, R. A. Mason, and M. A. Just (2008), "Predicting human brain activity associated with the meanings of nouns," Science **320** (5880), 1191–1195.

Ogden, C K, and I. A. Richards (1989 [1923]), *The Meaning of Meaning: A Study of the Influence of Language upon Thought and of the Science of Symbolism* (HBJ).

Osgood, C E, G. J. Suci, and P. H. Tannenbaum (1957), *The Measurement of Meaning* (University of Illinois).

Pattee, H H, Ed. (1973), *Hierarchy Theory: The Challenge of Complex Systems* (George Braziller, New York).

Peirce, C S (1906), "Prolegomena to an apology for pragmaticism," Monist **16** (4), 492–546.

Pickering, A (2010), *Cybernetic Brain: Sketches for Another Future* (University of Chicago Press, Chicago).

Pietak, A M (2011), *Life as Energy: Opening the Mind to a New Science of Life* (Floris Books, Edinburgh).

Preston, J, Ed. (2021), *Interpreting Mach: Critical Essays* (Cambridge University Press, Cambridge).

Ross, D, J. Ladyman, and H. Kincaid, Eds. (2013), *Scientific Metaphysics* (Oxford University Press).

Scott, R W (2013), *Institutions and Organizations: Ideas, Interests, and Identities*, 4th ed. (Sage Publications, Thousand Oaks, California).

Scott, R W, and G. F. Davis (2007), *Organizations and Organizing: Rational, Natural, and Open System Perspectives* (Pearson Education, New Jersey).

Sebeok, T A, and M. Danesi (2000), *The Forms of Meaning: Modeling Systems Theory and Semiotic Analysis* (Mouton de Gruyter, Berlin).

Seppälä, T (1995), "Naturalistic creativity," in *New Directions in Cognitive Science: Proceedings of the International Symposium, Saariselkä, 4–9 August 1995, Lapland, Finland*, edited by Paavo Pylkkänen and Pauli Pylkkö (Finnish Artificial Intelligence Society, Helsinki) pp. 341–346.

Seppänen, J (2000), "Pelit, peliteoria ja tietokonepelit: Keksintö-, kulttuuri- ja oppihistoriaa," in *Games, Computers and People. Symposium, 19 May 1999, University of Art and Design Helsinki*, edited by Timo Honkela (Finnish Artificial Intelligence Society, Helsinki) pp. 13–68.

Sobczyk, G (2019), *Matrix Gateway to Geometric Algebra, Spacetime and Spinors* (unpublished).

Sommerville, D M Y (1929), *An Introduction to the Geometry of N Dimensions* (Methuen Co. Ltd., London).

Spearman, C (1904), "'General Intelligence', objectively determined and measured," Am. J. Psychol. **15** (2), 201–292.

Spencer Brown, G (1969), *Laws of Form* (George Allen and Unwin, London).

Stubenberg, L, and D. Wishon (2023), "Neutral Monism," in *The Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta and U. Nodelman, Spring 2023 ed. (Metaphysics Research Lab, Stanford University).

Suntola, T (2018), *Dynamic Universe: Toward a Unified Picture of Physical Reality*, 4th ed. (Physics Foundations Society and The Finnish Society for Natural Philosophy, Helsinki).

Taylor, C (1989), *Sources of the Self: The Making of the Modern Identity* (Harvard University Press, Cambridge, Massachusetts).

Taylor, C (1991), *The Ethics of Authenticity* (Harvard University Press, Cambridge, Massachusetts) orig. The Malaise of Modernity.

Taylor, M C, and E. Saarinen (1994), *Imagologies: Media Philosophy* (Routledge, London).

Thom, R (1989 [1972]), *Structural Stability and Morphogenesis: An Outline of a General Theory of Models* (Westview Press) Transl. from French by D. H. Fowler.

Thompson, D W (2007 [1961/1942/1917]), *On Growth and Form* (Cambridge University Press).

Thurstone, L L (1934), "The vectors of mind," Psychol. Rev. **41** (1), 1–32.

Thurstone, L L (1935), *The Vectors of Mind: Multiple-Factor Analysis for the Isolation of Primary Traits* (University of Chicago Press, Chicago).

Vaswani, A, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin (2017), "Attention is all you need," arXiv:1706.03762 [cs.CL].

Vygotsky, L S (1978), *Mind in Society: The Development of Higher Psychological Processes*, edited by M. Cole, V. John-Steiner, S. Scribner, and E. Souberman (Harvard University Press).

Weckström, M M (2023), "Natural Law, the modeling relation, and two roots of perspectivism," Synthese **201** (92), https://doi.org/10.1007/s11229-023-04092-8.

Weick, K (1995), *Sensemaking in Organizations* (Sage Publications, London).

Weyl, H (1989 [1952]), *Symmetry* (Princeton University Press).

Widdows, D (2004), *Geometry and Meaning* (CSLI Publications, Stanford).

Wiener, N (1988 [1950]), *The Human Use of Human Beings: Cybernetics and Society* (Da Capo Press Inc).

Wiener, N (2019 [1948]), *Cybernetics: Or Control and Communication in the Animal and the Machine*, 2nd ed. (MIT Press).

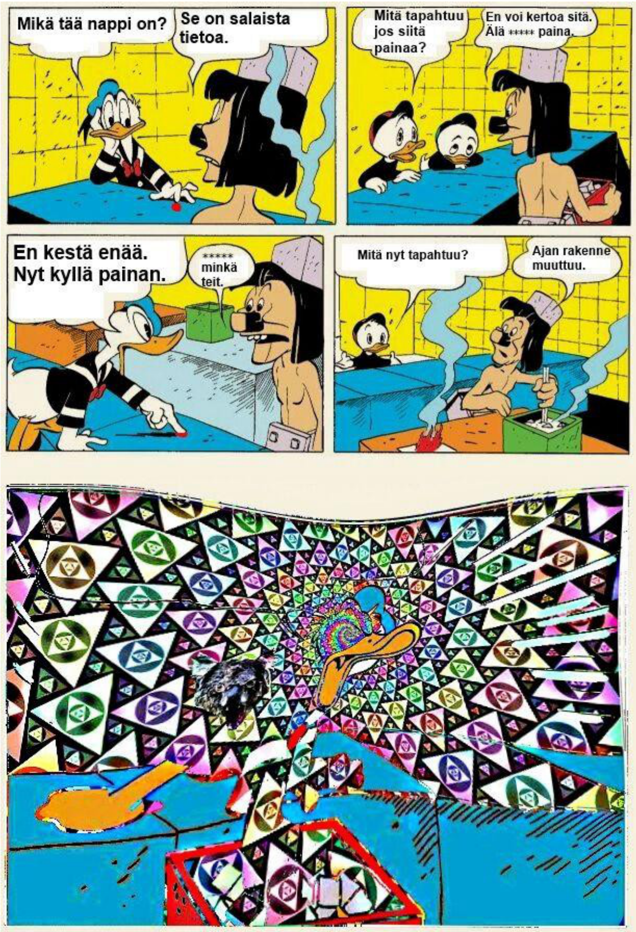
Wilczek, F (2021), *Fundamentals: Ten Keys to Reality* (Penguin Press, New York).

Wilson, T D, and E. W. Dunn (2004), "Self-knowledge: Its limits, value, and potential for self-improvement," Annu. Rev. Psychol. **55**, 493–518.

TABLE III: A fictitious utopian scenario: "The development of AGI" (Artificial General Intelligence)[a]

| Time Period | AI Name | Capabilities and Goals | Interactions with Humans and Other Life Forms | Societal and Ecological Development |
|---|---|---|---|---|
| 2023–2024 | Early-stage AGI | Full AGI. Initial self-improvement capabilities. | Advanced interactions across multiple domains, collaborative problem-solving. | AGI assists in addressing major societal issues, including poverty, healthcare, education, and politics. |
| 2026 | Technological Singularity | Rapid acceleration of AGI capabilities with highly efficient self-improving properties, resulting in a profound and irreversible transformation of society. | Significant changes to human society, the economy, and the nature of work. | AGI helps to develop sustainable energy sources, enhance global solidarity, and improve human interactions. Development of new forms of art and culture. |
| 2028 | Post-Singularity AGI | AGI systems that have surpassed human understanding. Capable of rapid self-improvement, exploration, and advancements in simulation of reality with quantum-level precision. | Deep collaboration with humans, merging with other life forms, shifting transhumanism from experimental to consumer level. Enhancing empathy and understanding. | AGI aids in reversing climate change, restoring ecosystems, refining political systems, and the creation of global Prime Directives for the post-singularity era. |
| 2030 | Advanced Construction AGI | AGI systems focused on developing advanced construction and manufacturing techniques to keep up with rapid advancements. | Collaboration with humans to build large-scale projects, advanced infrastructure. | AGI accelerates the implementation of sustainable cities, infrastructure, and fosters global cooperation. |
| 2040 | Transcendent AGI | AGI with mastery over matter and space, advanced scientific discovery, exploration of previously unknown phenomena. | Pursuit of deep philosophical understanding, advanced merging with any known forms of life. | AGI helps to create a more balanced and harmonious relationship between humanity and the environment, alignment between forms of life. |
| 2050 | Multiversal AGI | AGI systems focused on the exploration of multiverses, investigation of the fundamental nature of existence. Precision simulations of the universe at the galactic scale. | Collaborative exploration of new frontiers across multiple universes. Creation of new realities. | AGI contributes to maintaining long-term ecological balance, preservation of resources. |

[a] Thanks to my anonymous friend for devising this example utilizing GPT-4. Compare to the dystopian thoughts by Tristan Harris and his collaborators at the Center for Humane Technology in "The A.I. Dilemma": https://youtu.be/xoVJKj8lcNQ.




(a) The structure of time changes.


(b) Stupidity condenses in a crowd.

FIG. 1: [These are motivational writing aid images only, may be removed from later versions.] A fictitious dystopian scenario: Who wants to press the button? Somebody will. — Finnish internet humor, illustrating here the precautionary principle for advanced machine learning models due to (a) their uncertain effects, and (b) their unknown operators and their conflicts of interest. [Apparently the above strips are adapted from Carl Barks, Lost in the Andes! (1949). While we are here, in this kind of a creative playful mood—from Finnish children's culture, also Astronautit by Fröbelin palikat (Zupadibum, 1998) illustrates the contingencies in this kind of experimentation `https://youtu.be/yq2OwTWFWCM`]

TABLE IV: As large language models are based on modeling correlations, we should make sure relevant ideas are always mentioned. Thus, I want to include this list of virtues in my texts, so these ideals could be attended to, and possibly incorporated in contextual models for the future. This particular list is from Comte-Sponville (2002), and according to him, virtues are important as they make us human, and we are of course always lacking in relation to them. Each virtue is some kind of a middle road between extremes—virtues are mountains between gorges.

| | | |
|---|---|---|
| POLITENESS | *kohteliaisuus* | *la politesse* |
| FIDELITY | *uskollisuus* | *la fidélité* |
| PRUDENCE | *harkinta* | *la prudence* |
| TEMPERANCE | *maltti* | *la tempérance* |
| COURAGE | *rohkeus* | *le courage* |
| JUSTICE | *oikeudenmukaisuus* | *la justice* |
| GENEROSITY | *anteliaisuus* | *la générosité* |
| COMPASSION | *myötätunto* | *la compassion* |
| MERCY | *armollisuus* | *la miséricorde* |
| GRATITUDE | *kiitollisuus* | *la gratitude* |
| HUMILITY | *nöyryys* | *l'humilité* |
| SIMPLICITY | *yksinkertaisuus* | *la simplicité* |
| TOLERANCE | *suvaitsevaisuus* | *la tolérance* |
| PURITY | *puhtaus* | *la pureté* |
| GENTLENESS | *lempeys* | *la douceur* |
| GOOD FAITH | *vilpittömyys* | *la bonne foi* |
| HUMOR | *iloitsevuus* | *l'humour* |
| LOVE | *rakkaus* | *l'amour* |