Ruliology - Multicomputation

Sequentialism assumed in consciousness. Observer sequentialism leads to interpretation of laws.

Learn graph theory, statistical mechanics, dynamical systems theory, nonequilibrium thermodynamics, synergetics, cybernetics, general systems theory, nonlinear dynamics, category theory, Einstein's equations.

How we sample rulial space. Humans sample computationally reducible slices of rulial space.

MULTICOMPUTATIONAL SYSTEMS HAVE ANNALOG OF EINSTEIN'S EQUATIONS

Summary of metamathematics:
- Math statements: hyperedge
- Laws of inference: to build multiway graph (structural substitution or logical principle)
- Mathematicians do not see raw proof graph, they aggregate them together. This is choosing a "mathematical reference frame". Example: "category theory works"
- Proof space with light cones that define dependencies between results.
- Geodesics that are shortest derivation.
- Higher density of proofs with more interconnected fields
- Proof geodesics will be "gravitationally attracted" to regions of higher mathematical energy.
- Proof paths that go on forever, undecidable.
- Mathematical black holes, singularity, collection of proof paths that end, this is a decidable area.
- Specific math concepts are populated reference frames of metamathematics.
- Bulk multicomputational laws that apply to homogeneity of maths pace and how it shows up so often.
- Limiting object of math will be the same for physics.
- Foliations are continuous analogs of slices in multiway systems.

Rulial space is computational and NOT hypercomputational (e.g. requiring oracle that immediately answers that require infinite computation). Implicitly assuming that rules can be stated in explicit symbolic form.

Causal invariance due to construction of sampling make stuff inevitable.

Choice of time slices arbitrary.

Metamodeling - find the most minimal primitive structure.

Tendency to nail down ruliology using math, but you will ignore computationally irreducible phenomena.

But when an observer tries to “see what’s going on” in the system, they inevitably conflate things together, effectively perceiving only certain equivalence classes of the lowest-level elements. Example: seeing average densities of molecules and perceiving a diffusion law.

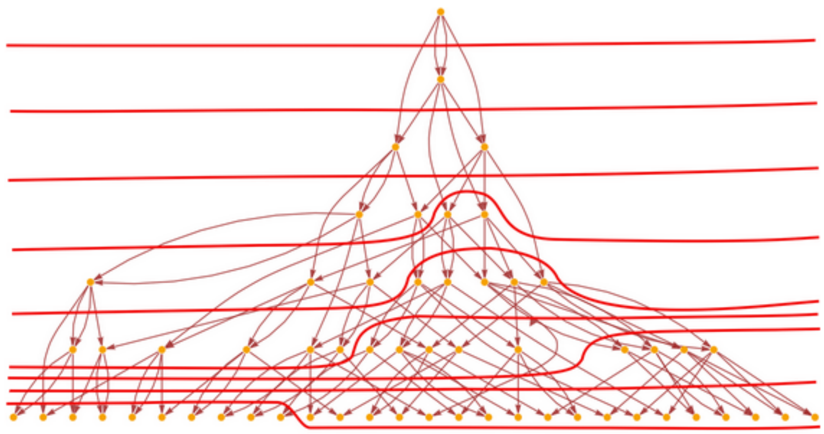
Topological phenomena for example that is preserved when the parts are continually changed. In physics, these correspond to elementary particles and symmetries. Pockets of local computational reducibility that we see. But even that is suspect, one electron may be different from another.

Energy is the density of activity

Question: how can branchial space be given coordinates (given observers) since it is non-numerical space.

Finding laws is how observers fit together system details to synthesize perceptions of it.

The basic point is that we don't perceive the whole causal graph in all its detail. Instead, as computationally bounded observers, we just pick some particular reference frame from which to perceive what's going on. And this reference frame defines a sequence of global "time slices" such as:



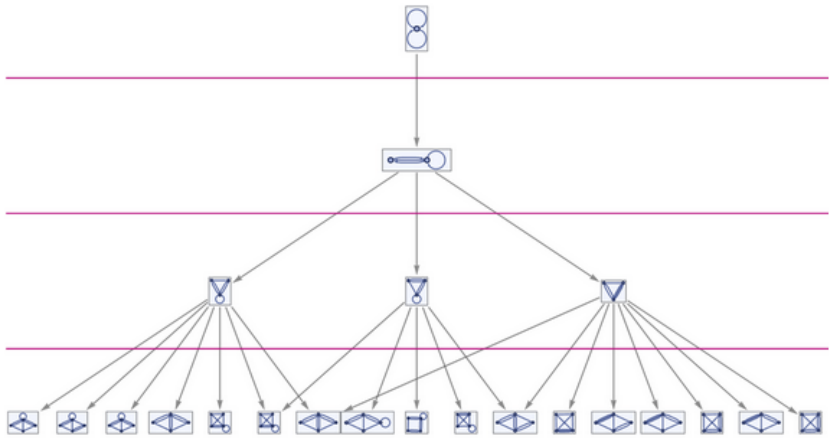
Each "time slice" contains a collection of events that—with our reference frame—we take to be "happening simultaneously". And we can then trace the "steps in the evolution of the universe" by seeing the results of all updating events in successive time slices:

But how do we determine what reference frame to use? The underlying rule determines the structure of the causal graph, and what event can follow what other one. But it still allows huge freedom in the choice of reference frame—in effect imposing only the constraint that if one event follows another, then these events must appear in that order in the time slices defined by the reference frame:

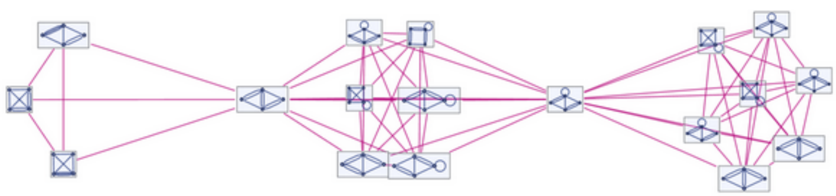
And in general to achieve a "pathological result" we'll typically have to "reverse engineer" the underlying computational irreducibility of the system—which we won't be able to do with a reference frame constructed by a computationally bounded observer. (This is directly analogous to the result in the ordinary computational paradigm that computationally bounded observers effectively can't avoid perceiving the validity of the Second Law of Thermodynamics.)

So, OK, what then will an observer perceive in a system like the one we've defined? With a variety of caveats the basic answer is that in the limit of a "sufficiently large universe" they'll perceive average behavior that's simple enough to describe mathematically, and specifically to describe as following Einstein's equations from general relativity. And the key point is that this is in a sense a generic result (a bit like the gas laws in thermodynamics) that's independent of the details of the underlying rule.

With a multiway graph of the kind we've drawn above (in which every node represents a possible "complete state of the universe"), we can investigate "pure branchial space" by looking at time slices in the graph:



For example, we can construct "branchial graphs" by looking at which states are connected by having immediate common ancestors. And in effect these branchial graphs are the branchial-space analogs of the hypergraphs we've constructed to represent the instantaneous state of ordinary space:



But now, instead of representing ordinary space—with features like general relativity and gravity—they represent something different: they represent a space of quantum states, with the branchial graph effectively being a map of quantum entanglements.

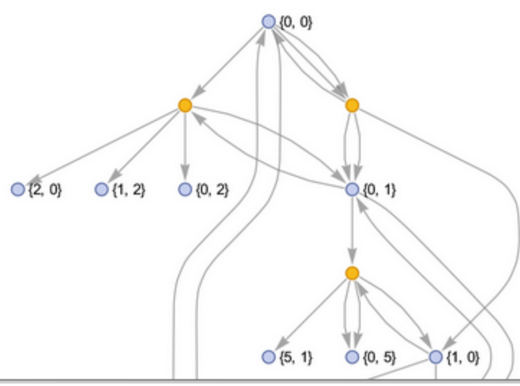
When we think about multicomputational systems in general, the conflation of "identical" (say by isomorphism) states is in a sense the "lowest-level act" of an observer. The "true underlying system" might in some sense "actually" be generating lots of separate, identical states. But if the observer can't tell them apart then we might as well say they're all "the same state". (Of course, when there are different numbers of paths that lead to different states, this can affect the weightings of these different states—and indeed in our model of physics this is where the different magnitudes of quantum amplitudes for different states come from.)

It seems natural and obvious to conflate hypergraphs if they're isomorphic. But actual observers (say humans observing the physical universe) typically conflate much, much more than that. And indeed when we say that we're operating in some particular reference frame we're basically defining potentially huge collections of states to conflate.

But there's actually also a much lower level at which we can do conflation. In the token-event graphs that we've looked at so far, every token generated by every event is shown as a separate node. But—as the labeled versions of these graphs make clear—many of these tokens are actually identically the same, in the sense that they're just direct copies created by our way of computing (and rendering) the token-event graph.

So what about conflating all of these—or in effect "deduplicating" tokens so that we have just one unique shared representation of every token, regardless of how many times or where it appears in the original graph?

Here's the result after doing this for the 2-step version of the token-event graph above:



> At the lowest level something like a gas consists of large numbers of discrete molecules interacting according to certain rules. And it's almost inevitable that the detailed behavior of these molecules will show computational irreducibility—and great complexity. But to an observer who just looks at things like average densities of molecules the story will be different—and the observer will just perceive simple laws like diffusion.
>
> And in fact it's the very complexity of the underlying behavior that leads to this apparent simplicity. Because a computationally bounded observer (like one who just looks at average densities) won't be able to do more than just read the underlying computational irreducibility as being like "simple randomness". And this means that for such an observer it's going to be reasonable to model the overall behavior by using mathematical concepts like statistical averaging, and—at least at the level of that observer—to describe the system as showing computationally reducible behavior represented, say, by the diffusion equation.

One possibility is just that the observer can choose to name things: "I'll call this token 'Tokie' and then I'll trace what happens, and describe the behavior of the universe in terms of the 'adventures of Tokie'". But as such, this approach will inevitably be quite limited. Because a feature of multicomputational systems is events are continually happening, consuming existing tokens and creating new ones. In physics terms, there is nothing fundamentally constant in the universe: everything in it (including space itself) is being continually recreated.

So how come we have any perception of permanence in physics? The answer is that even though individual tokens are continually being created and destroyed, there are overall patterns that are persistent. Much like vortices in a fluid, there can for example be essentially topological phenomena whose overall structure is preserved even though their specific component parts are continually changed.

One might have thought that what would be required most would be to do a successful "reduction" to an accurate model of the primitive parts of the system. But actually what the multicomputational paradigm indicates is that there's a certain inexorability to what happens, independent of those details. The challenge, though, is to figure out what an "observer" of a certain kind of system will actually perceive. In other words, successfully finding overall laws isn't so much about applying reductionism to the system; it's more about understanding how observers fit together the details of the system to synthesize their perception of it.

So what kinds of systems can we expect to describe in multicomputational terms? Basically any kind of system where there are many component parts that somehow "operate independently in parallel"—interacting only through certain "events". And the key idea is that there are many possible detailed histories for the system—but in the multicomputational paradigm we look at all of them together, thereby building up a structure with inexorable properties, at least as perceived by certain general kinds of observers.

In areas like statistical physics it's been common for a century to think about "ensembles of possible states" for a system. But what's different about the multicomputational paradigm is that it's not just looking "statically" at "possible states"; instead it's "taking a bigger gulp", and looking at all possible whole histories for the system, essentially developing through time. And, yes, a slice at a particular time will show some ensemble of possible states—but they're states generated by the entangled possible histories of the system, not just states "statically" and combinatorially generated from the possible configurations of the system.

For me the story began nearly 50 years ago—with what I saw as a great and fundamental mystery of science. We see all sorts of complexity in nature and elsewhere. But where does it come from? How is it made? There are so many examples. Snowflakes. Galaxies. Lifeforms. Turbulence. Do they all work differently? Or is there some common underlying cause? Some essential "phenomenon of complexity"?

What is the "evolution of mathematics" like? Basically we imagine that there are laws of inference that take, say, two mathematical statements and deduce another one from them, either using something like structural substitution, or using some (symbolically defined) logical principle like modus ponens. The result of repeatedly applying a law of inference in all possible ways is to build up a multiway graph of—essentially—what statements imply what other ones, or in other words, what can be proved from what.

But what does a human mathematician perceive of all this? Most mathematicians don't operate at the level of the raw proof graph and individual raw formalized mathematical statements. Instead, they aggregate together the statements and their relationships to form more "human-level" mathematical concepts.

In effect that aggregation can be thought of as choosing some "mathematical reference frame"—a slice of metamathematical space that can successfully be "parsed" by a human "mathematical observer". No doubt there will be certain typical features of that reference frame; for example it might be set up so that things are "sufficiently organized" that "category theory works", in the sense that there's enough uniformity to be able to "move between categories" while preserving structure.

Metamathematics

Let’s start by talking about perhaps the most abstract potential application area: metamathematics. The individual “tokens of mathematics” can be mathematical statements, written in some symbolic form (as they would be in the Wolfram Language). In a sense these mathematical statements are like the hyperedges of our spatial hypergraph in physics: they define relations between elements, which in the case of physics are “atoms of space” but in the case of mathematics are “literal mathematical objects”, like the number 1 or the operation Plus (or at least single instances of them).

Now we can imagine that the “state of mathematics” at some particular time in its development consists of a large number of mathematical statements. Like the hyperedges in the spatial hypergraph for physics, these mathematical statements are knitted together through their common elements (two mathematical statements might both refer to Plus, just as two hyperedges might both refer to a particular atom of space).

What is the “evolution of mathematics” like? Basically we imagine that there are laws of inference that take, say, two mathematical statements and deduce another one from them, either using something like structural substitution, or using some (symbolically defined) logical principle like modus ponens. The result of repeatedly applying a law of inference in all possible ways is to build up a multiway graph of—essentially—what statements imply what other ones, or in other words, what can be proved from what.

But what does a human mathematician perceive of all this? Most mathematicians don’t operate at the level of the raw proof graph and individual raw formalized mathematical statements. Instead, they aggregate together the statements and their relationships to form more “human-level” mathematical concepts.

In effect that aggregation can be thought of as choosing some “mathematical reference frame”—a slice of metamathematical space that can successfully be “parsed” by a human “mathematical observer”. No doubt there will be certain typical features of that reference frame; for example it might be set up so that things are “sufficiently organized” that “category theory works”, in the sense that there’s enough uniformity to be able to “move between categories” while preserving structure.

There are both familiar and unfamiliar features of this emerging picture. There are the analog of light cones in “proof space” that define dependencies between mathematical results. There are geodesics that correspond to shortest derivations. There are regions of “metamathematical space” (the slices of proof space) that might have higher “densities of proofs” corresponding to more interconnected fields of mathematics—or more “metamathematical energy”. And as part of the generic behavior of multicomputational systems we can expect an analog of Einstein’s equations, and we can expect that “proof geodesics” will be “gravitationally attracted” to regions of higher “metamathematical energy”.

In most areas of metamathematical space there will be "proof paths" that go on forever, reflecting the fact that there may be no path of bounded length that will reach a given statement, so that the question of whether that statement is present at all can be considered undecidable. But in the presence of large amounts of "metamathematical energy" there'll effectively be a metamathematical black hole formed. And where there's a "singularity in metamathematical space" there'll be a whole collection of proof paths that just end—effectively corresponding to a decidable area of mathematics.

Mathematics is normally done at the level of "specific mathematical concepts" (like, say, algebraic equations or hyperbolic geometry)—that are effectively the "populated places" (or "populated reference frames") of metamathematical space. But by having a multicomputational model of the low-level "machine code of metamathematics" there's the potential to make more general statements—and to identify what amount to general "bulk laws of metamathematics" that apply at least to the "metamathe matical reference frames" used by human "mathematical observers".

What might these laws tell us? Perhaps they'll say something about the homogeneity of metamathematical space and explain why the same structures seem to show up so often in different areas of mathematics. Perhaps they'll say something about why the "aggregated" mathematical concepts we humans usually talk about can be connected without infinite paths—and thus why undecidability is so comparatively uncommon in mathematics as it is normally done.

But beyond these questions about the "insides of mathematics", perhaps we'll also understand more about the ultimate foundations of mathematics, and what mathematics "really is". It might seem a bit arbitrary to have mathematics be constructed according to some particular law of inference. But in direct analogy to our Physics Project, we can also consider the "rulial multiway system" that allows all possible laws of inference. And as I've argued elsewhere, the limiting object that we get for mathematics will be the same as for physics, connecting the question of why the universe exists to the "Platonic" question of whether mathematics "exists".

The Relation to Mathematics

Most people would view it as self-evident that the physical universe exists. But the question of whether mathematics "intrinsically exists as a definite thing" has been debated since at least Plato.

One view of mathematics is that it involves just writing down whatever axiom system one wants, and then working out its consequences. And at least at first this seems very different from physics, where one might imagine that one starts from our particular world as it is, and then tries to find a formal representation (or "model") of it. And, yes, our effort to find a complete and precise fundamental theory of physics can be thought of as trying to "reduce physics to mathematics"—in the sense that it's trying to give us a particular formal ("axiomatic") system that reproduces what the physical universe does.

At first, we might imagine that this formal system for physics must be something very special, and not something like the "arbitrary axiom systems" that we could choose to write down as foundations for mathematics. But from what we've seen here, the situation is more subtle than this. Because "underneath physics" there are in a sense all possible formal systems.

At the outset we might have imagined that mathematics is somehow much more general than physics, because it can operate with whatever formal system we write down. But now it's seeming like the opposite: physics is based on all possible formal systems, but mathematics is based on particular formal systems (geometry, algebra, etc.) that we happen to have written down in the history of human mathematics.

We've argued that the existence of the universe is ultimately a consequence of the fact that all possible formal systems exist as a matter of abstract necessity. So can we use a similar argument for the "existence of mathematics"?

Most likely we can—at least if we subtly change our description of what mathematics is. In the "axiomatic tradition" it's been common to imagine that mathematics could in principle be based on whatever formal axioms we want, although in practice we pick particular ones. But an alternative view is that ultimately mathematics, like physics, is actually based on all possible formal (axiom) systems.

It's worth remembering that any given rule won't typically be following geodesics in rulial space. It'll be following some more circuitous path (or bundle of paths). But let's say one has some rule that traces out some trajectory—corresponding to performing some computation. The Principle of Computational Equivalence implies that across different possible rules, there's a standard "maximum computational sophistication" for these computations, and many rules achieve it. But then the principle also implies that there's equivalence between these maximally sophisticated computations, in the sense that there's always a limited computation that translates between them.

But what units is this distance in? Basically it's in units of rule—or program—size. Given any program—or rule—we can imagine writing that program out in some language (say in Wolfram Language, or as a program for a particular universal Turing machine, or whatever) And now we can characterize the size of the program by just looking at how many tokens it takes to write the program out.

Of course, with different languages, that number will be different—at the simplest level just like the number of decimal digits necessary to represent a number is different from the number of binary digits, or the length of its representation in terms of primes. But it's just like measuring a length in feet or meters: even though the numerical value is different, we're still describing the same length.

It's important to point out that it's not enough to just measure things in terms of "raw information content", or ordinary bits, as discussed in information theory. Rather, we want some kind of measure of "semantic information content": information content that directly tells us what computation to do.

It's also important to point out that what we need is different from what's discussed in algorithmic information theory. Once one has a computation universal system, one can always use it to translate from any one language to any other. And in algorithmic information theory the concept is that one can measure the length of a program up to an additive constant by just expecting to include an "emulation program" that adapts to whatever language one's measuring the length in. But in the usual formalism of algorithmic information theory one doesn't worry about how long it's going to take for the emulation to be done; it's just a question of whether there's ultimately enough information to do it.

In our setup, however, it does matter how long the emulation takes, because that process of emulation is actually part of our system. And basically we need the number of steps needed for the emulation to be in some sense bounded by a constant.

So, based on our previous estimates (which I don't consider anything more than vaguely indicative yet) we might conclude that perhaps:

$\rho \sim$ 10450 Wolfram-Language-tokens/second

The number of "parallel threads" in the rulial multiway graph (the rulial analog of Ξ) might then be related to the number of possible hypergraphs that contain about the number of nodes in the universe, or very roughly (10350)^(10350) $\approx 10^{10^{353}}$. If we ask the total number of Wolfram Language tokens processed by the universe, there'll be another factor ~10467, but this "parallelism" will completely dominate, and the result will be about:

$10^{10^{356}}$ Wolfram-Language-tokens

But for me there's something personally interesting about Greg Chaitin's Ω showing up in any way in a potential description of anything to do with the universe. You see, I've had a nearly 40-year-long debate with Greg about whether the universe is "like π" or "like Ω". In other words, is it possible to have a rule that will let us compute what the universe does just like we can compute (say, in principle, with a Turing machine) the digits of π? Or will we have to go beyond a Turing machine to describe our universe? I've always thought that our universe is "like π"; Greg has thought that it might be "like Ω". But now it looks as if we might both be right!

In our models, we're saying that we can compute what the universe does, in principle with a Turing machine. But what we're now finding out is that in the full rulial space, general limiting statements pull in Ω. In a particular rulial observation frame, we're able to analyze the universe "like π". But if we want to know about all possible rulial observation frames—or in a sense the space of all possible descriptions of the universe—we'll be confronted with Ω.

By then, I had read many academic papers, and pretty soon I had written one of my own. It took two tries, but then, there it was, my first published paper, complete—I now notice—with some self-references to earlier work of mine, in true academic style:

Aust. J. Phys., 1975, 28, 479-87

Hadronic Electrons?

Stephen Wolfram

Eton College, Windsor, Berkshire, England.

Abstract

A new form of high energy electron–hadron coupling is examined with reference to the experimental data. The electron is taken to have a neutral vector gluon cloud with a radius $\sim 10^{-18}$ m. This is shown to be consistent with measurements on $e^+e^- \to e^+e^-$ and g_e-2. At low energies, only photons couple to the gluons, but at higher energies 'evaporation' then 'boiling' of ω and ϕ occurs, allowing strong interactions. The model yields accurate predictions for the form of the rise in [illegible]. Arguments are given for the order of magnitude of m_e and for the lack of a permanent meson cloud in leptons. Strong interaction selection rules forbid a contribution to [illegible], and interference with the one-photon channel produces minimal scaling violation in eN processes at present energies. The constant value of [illegible] is correctly predicted and evidence from high energy pp interactions is also cited. The ψ particles are interpreted as e^+e^- resonances in the evaporation region, and their properties are generated correctly. Predictions are given for the behaviour of [illegible] at high energies.

Introduction

The recent discoveries of scaling violation and narrow resonances in deep inelastic e^+e^- annihilation (see review of data by Gilman 1975) have raised the possibility that electrons may undergo non-electromagnetic interactions with hadrons (Bigi and Bjorken 1974; Chanda 1974; Greenberg and Yodh 1974; Pati and Salam 1974a, 1974b; Richter 1974; Soni 1974a, 1974b; Wolfram 1975). In the low energy limit $q^2 \to 0$, it is well known that electrons obey quantum electrodynamics (QED) to considerable accuracy, and hence in this region the strength of any anomalous electron–hadron coupling must be negligible. At higher energies ($q^2 \gtrsim 15$ GeV^2), however, the predictions of QED fail, and scaling is violated. In electron–nucleus interactions, there is also scaling in the low energy region ($q^2 \lesssim 0 \cdot 09$ GeV^2), but this ceases as the electrons probe the nucleon form factors and induce free pion production (Chanowitz and Drell 1973). In the hadronic electron model presented here, scaling is broken by a similar process.

Structure of the Electron

By analogy with hadrons, the core of the electron is taken as a collection of 'bitons' bound by a superstrong interaction mediated by neutral massive vector gluons with photon quantum numbers. These will not be restricted to the core, but will tend to form a cloud around it. From a comparison of the expected gradients of strong and weak Regge trajectories, we expect a characteristic weak interaction structure size of order $\sqrt{G_F} \sim 10^{-18}$ m (Greenberg and Yodh 1974).

It was a creative and decently written paper, but it was technically a bit weak (heck, I was only 15), and, at least at the time, its main idea did not pan out. But of course there's an irony to all this. Because—guess what—45 years later, in our current model for fundamental physics, the electron is once again not a point particle! Back in 1975, though, I thought maybe it had a radius of 10^{-18} meters; now I think it's more likely 10^{-81} meters. So at the very least 15-year-old me was wrong by 63 orders of magnitude!

Being a "teenage physicist" had its interesting features. At my boarding school (the older-than-the-discovery-of-America Eton), there was much amusement when mail came addressed to me as "Dr. S. Wolfram". Soon I started doing day trips to go to physics seminars in Oxford—and interacting with "real physicists" from the international physics community. I think I was viewed as an exotic phenomenon, usually referred to in a rather Wild-West way as "The Kid". (Years later, I was

"You Can't Leave Physics"

It took a couple of years to build the first version of SMP. I continued to do particle physics, though I could already feel that the field was cooling, and my interests were beginning to run to more general, theoretical questions. SMP was my first large-scale "practical" project. And not only did it involve all sorts of software engineering, it also involved managing a team—and ultimately starting my first company.

Physicists I knew could already tell I was slipping away from physics. "You can't leave physics", they would say. "You're really good at this." I still liked physics, and I particularly liked its "let's just figure this out" attitude. But now I wasn't just applying that methodology in quantum field theory and cosmology; I was also using it in language design, in software development, in entrepreneurism, and in other things. And it was working great.

The process of starting my first company was fraught with ahead-of-my-time-in-the-interaction-between-companies-and-universities issues, that ultimately caused me to leave Caltech. And right in the middle of that, I decided I needed to take a break from my mainline "be a physicist" activities, and just spend some time doing "something fun".

I had been thinking for a long time about how it is that complex things manage to happen in nature. My two favorite examples were neural networks (yes, back in 1981, though I never figured out how to make them do anything very useful back then) and self-gravitating gases. And in my "just have fun" approach I decided to try to make the most minimal model I could, even if it didn't really have much to do with either of these examples, or officially with "physics".

It probably helped that I'd spent all that time developing SMP, and was basically used to just inventing abstract things from scratch. But in any case, what I came up with were very simple rules for arrays of 0s and 1s. I was pretty sure that—as such—they wouldn't do anything interesting. But it was basically trivial for me to just try running them on a computer. And so I did. And what I found was amazing, and gradually changed my whole outlook on science and really my whole worldview—and sowed the seeds that have now, I believe, brought us a path to the fundamental theory of physics.

But as someone who'd studied the history of science for a long time, I full well knew that if the new paradigm I was trying to introduce was as important as I believed, then inevitably it would run into detractors, and hostility. But what surprised me was that almost all the hostility came from just one field: physics. There were plenty of physicists who were very positive, but there were others for whom my book somehow seemed to have touched a nerve.

As an almost lifelong lover of physics, I didn't see a conflict. But maybe from the outside it was more obvious—as a cartoon in a review of my book in the *New York Times* (with a remarkably prescient headline) perhaps captured:

You Know That Space-Time Thing? Never Mind

According to Stephen Wolfram, computer programs — not mathematical equations — explain the way the universe works.

A NEW KIND OF SCIENCE
By Stephen Wolfram.
Illustrated. 1,263 pp. Champaign, Ill.: Wolfram Media. $44.95.

By George Johnson

AMONG a small group of very smart people, the publication of "A New Kind of Science," by Stephen Wolfram, has been anticipated with the anxiety aroused in literary circles by, say, Jonathan Franzen's recent novel, "The Corrections." For more than a decade, Wolfram, a theoretical physicist turned millionaire software entrepreneur, has been laboring in solitude on a work that, he has promised, will change the way we see the world. Adding to the suspense, the book has been announced and withdrawn as the artist returned to his garret to tinker, ignoring the bad vibes and hexes cast by jealous colleagues hoping to see him fall flat on his face.

Now, weighing in at 1,263 pages (counting a long, unpaginated index) and 583,313 words, the book could hardly be more intimidating. But that is the price one pays for a first-class intellectual thrill. While experimenting with a simple computer program 20 years ago, Wolfram stumbled on something rather eerie: "the beginning of a crack in the very foundations of existing science." Ever since, he has been following it deeper as it widens into a crevasse.

The normal thing would have been to dispatch regular reports from the field — unreadable papers published in fashionable zines like Physical Review Letters or Physica D. Instead, Wolfram decided to do what Darwin did (and he would not shun the comparison). He is springing loose his vision all at once, in a book intended for nonscientists and scientists alike.

From the very beginning of this meticulously constructed manifesto, the reader is presented with a stunning proposal: all the science we know will be demolished and reassembled. An ancient error will be corrected, one so profoundly misguided that it has led science down

around the sun or a weight falling from the Leaning Tower of Pisa. But as scientists try to explain systems of greater complexity — a hurricane, the economy of Portugal, a human or even a reptilian brain — the calculations become ever more elaborate until one is left with an unwieldy array of symbols that do not explain much at all.

Wolfram believes that even his own field, theoretical physics (he got a Ph.D. from Caltech when he was 20), suffers from the problem. Equations can capture characteristics of individual particles with breathtaking precision. But put three or four particles together and the complications begin to overwhelm. The problem, he proposes, is that equations are the wrong tool for the job. They should be replaced with computer programs — more specifically, the little snippets of software called algorithms.

That sounds absolutely ridiculous.

The universal operating system Wolfram imagines is not something horribly complicated like Windows. The key idea in the book is that simple, byte-size programs have the surprising ability to produce endlessly intricate behavior. His most basic example is a group of elegant little algorithms with a clunky name: cellular automata.

These have been kicking around in the popular science press for years. Start with a row of squares (the cells), some white and some black. Then transform the pattern according to a mindlessly simple rule. Here is an example: if either of a cell's neighbors is black, then make the cell itself black in the next round; otherwise, make it white. That is the whole program. Print each new generation below its progenitor and a pattern unfolds like a piano roll. Automate the procedure with a computer and watch what scrolls down the screen.

THOMAS ESCHER

Programs are just human inventions,

Most of these experiments — Wolf-

Other rules have the opposite effect: seed them with a random jumble of cells and, after a few iterations, they begin generating complex order. Some of the output resembles intricately varied stalactites; some looks like tracks of colliding particles in a high-energy accelerator lab. Think of stars and galaxies emerging from the confusion of the Big Bang, or life from the primordial sea.

Most pleasing to the eye are rules generating nested patterns like those of a crystal or a snowflake, or the markings on a seashell, the branching of a leaf, the spiral of a pine cone. Other patterns swirl like clouds, smoke or turbulent streams of water.

Wolfram believes he has clinched the deal with what, for many scientists, will be the meat of the book: a proof that a simple cellular automaton can be programmed to perform any conceivable computation (making it equivalent to what the British mathematician Alan Turing called a universal computer). If you buy all this, then a simple algorithm like those described in the book could constitute the machine code of the universe, the platform on which all the other programs run.

One idea after another comes spewing from the automata in Wolfram's brain. Maybe it is not evolution but algorithms that generate biological complexity. Maybe, if everything arises from computations, it makes perfect sense to think of the weather and the stock market as having minds of their own. Maybe free will is the result of something called "computational irreducibility" — the fact that the only way to know what many systems will do is to just turn them on and let them run.

All this is laid out clearly and precisely. Any motivated reader should be able to plow through at least a few hundred pages before the details become too burdensome. Then one can just marvel at the pictures. (It's evident why Wolfram, who adds depth to the term "control freak," published this work himself. Some illustrations contain hundreds of checkered cells per inch, requiring "careful sheet-fed printing on paper smooth enough to avoid significant spreading of ink.")

If social media had existed at the time, it would undoubtedly have been different. But as it was, it was a whole unchecked parade: from Nobel prizewinners with pitchforks, to a then-graduate-student launching their career by "proving" that my physics was "wrong". Why did they feel so strongly? I think they thought (and some of them told me as much) that if I was right, then what they'd done with their traditional mathematical methods, and all the wonderful things they'd built, would get thrown away.

I never saw it that way (and, ironically, I made my living building a tool used to support those traditional mathematical methods). But at the time—without social media—I didn't have a useful way to respond. (To be fair, it often wasn't clear there was much to say beyond "I don't share your convictions", or "Read what the book actually says... and don't forget the 300,000 words of notes at the back!".)

But there was unfortunately a casualty from all this: physics. As it now turns out (and I'm very happy about it), far from my ideas being in conflict with what's been done in physics, they are actually beautifully aligned. Yes, the foundations are different. But all those traditional mathematical methods now get extra power and extra relevance. But it's taken an additional 18 years for us to find that out. And it almost didn't happen at all.

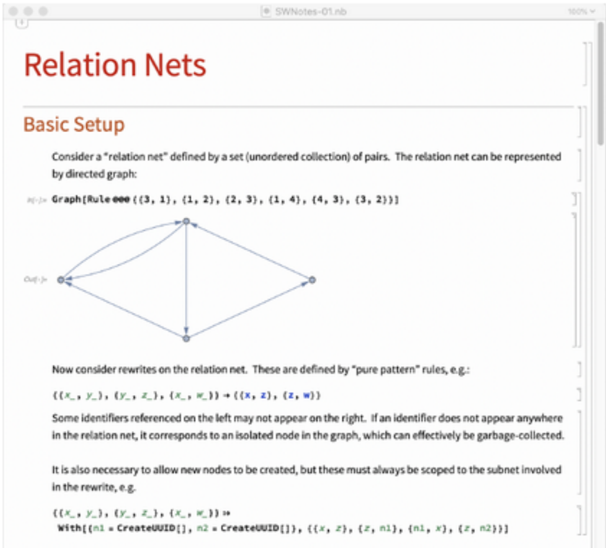
It's been interesting to watch the general progression of the ideas I discussed in *A New Kind of Science*. What's been most dramatic (and I'm certainly not solely responsible) has been the quiet but rapid transition—after three centuries—of new models for things being based not on equations but instead on programs. It's happened across almost every area. With one notable exception: fundamental physics.

Perhaps it's partly because the tower of mathematical sophistication in models is highest there. Perhaps it's because of the particular stage of development of fundamental physics as a field, and the fact that, for the most part, it's in a "work out the existing models" phase rather than in a "new models" phase.

Most of the projects I've ever done in my life—from my "Concise Directory of Physics" onward—I've done first and foremost because I was interested in them, and because I thought I would find them intellectually fulfilling. But particularly as I've gotten older, there's been another increasingly important factor: I find I get pleasure out of doing projects that I think other people will find useful—and will get their own fulfillment out of. And with the tools I've built—like Mathematica and Wolfram|Alpha and the Wolfram Language—as well as with A New Kind of Science and my other books and writings, that's worked well, and it's been a source of great satisfaction to me.

But then I had a little idea. I'd always been saying that I wanted models that are as minimal and structureless as possible. And then I'd say that networks were the best way I knew to get these, but that there were probably others. But even though I thought about lots of abstract structures through my work on the Wolfram Language, I never really came up with anything I was happy with. Until September 9, 2018.

But in September 2018 I think I was feeling more motivated by the abstract aesthetics than anything else. I realized there might be an elegant way to represent things—even things that were at least vaguely similar to the network-based models I had studied back in the 1990s. My personal analytics record that it took about 8 minutes to write down the basics:



There it was: a model defined by basically a single line of Wolfram Language code. It was very elegant, and it also nicely generalized the network models I had long thought about. And even though my description was written (for myself) in language-designer-ese, I also had the sense that this model had a certain almost-mathematical purity to it. But would it do anything interesting? Pretty soon I was doing what I basically always seem to end up doing: going out into the computational universe of possibilities and exploring. And immediately I was finding things like:

Of course, this is often the rhythm of science: some methodological advance sparks a golden age, and once everything easily accessible with that methodology has been done, one is faced with a long, hard slog that can last a century before there is some new methodological advance.

But going to the Summer School in June, I was again thinking about how to do my fundamental physics project.

Max was there. And so—as an instructor—was Jonathan Gorard. Jonathan had first come to the Summer School in 2017, just before his last year as an undergraduate in mathematics (+ theoretical physics, computer science and philosophy) at King's College London. He'd been publishing papers on various topics since he was 17, most recently on a new algorithm for graph isomorphism. He said that at the Summer School he wanted to work either on cosmology in the context of "Chapter 9", or on something related to the foundations of mathematics.

I suggested that he try his hand at what I considered something of an old chestnut: finding a good symbolic way to represent and analyze automated proofs, like the one I had done back in 2000 of the simplest axiom system for logic. And though I had no idea at the time, this turned out to be a remarkably fortuitous choice. But as it was, Jonathan threw himself into the project, and produced the seeds of what would become through his later work the Wolfram Language function **FindEquationalProof**.

Jonathan had come back to the Summer School in 2018 as an instructor, supervising projects on things like infinite lists and algebraic cryptography. And now he was back again as an instructor in 2019, having now also become a graduate student at Cambridge, with a nice fellowship, and nominally in a group doing general relativity.

It had been planned that Jonathan, Max and I would "talk about physics" at the Summer School. I was hopeful, but after so many years a bit pessimistic. I thought my little idea defined a new, immediate path about what one might do. But I still wasn't convinced there was a "good way to do the project".

But then we started discussing things. And I started feeling a stronger and stronger sense of responsibility. These ideas needed to be explored. Max and Jonathan were enthusiastic about

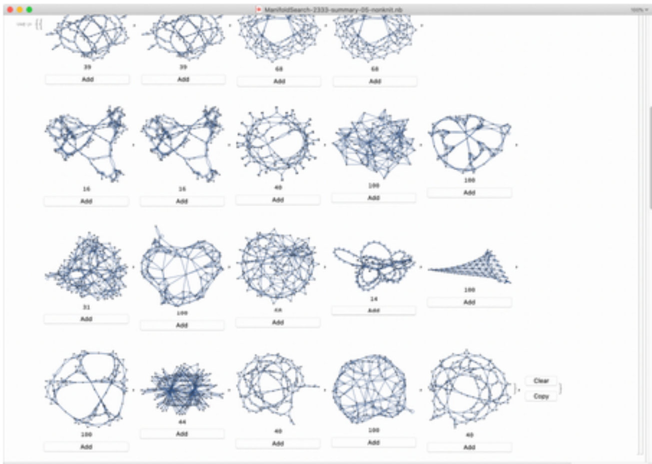
Jonathan had said a few times early in the Summer School that he didn't see why I was so concerned about causal invariance. I kept on pushing back. Then one day we went on a long walk, and Jonathan explained an idea he had (which, knowing him, he may have just come up with right there). What if the underlying rules didn't need to have causal invariance, because us observers would implicitly add it just by the way we analyze things?

What was this idea really? It was an application of things Jonathan knew from working on automated theorem proving, mixing in ideas from general relativity, and applying them to the foundations of quantum mechanics. (Basically, his concept was that we observers, because we're branching just like the system we're observing, effectively define "lemmas" to help us make sense of what we observe, and these lead to effective rules that have causal invariance.)

At first I was skeptical. But the issue with not finding enough causal invariance had been a blocker 16 years earlier. And it felt like a big weight lifted if that issue could be removed. So by the end of the walk I was convinced that, yes, it was worth looking at rules even if they were not explicitly causal invariant, because they could still be "saved" by the "Jonathan Interpretation of Quantum Mechanics" as I called it (Jonathan prefers the more formal term "completion interpretation", referring to the process of creating lemmas, which is called "completion" in automated theorem proving). As it turns out, the jury is still out on whether causal invariance is intrinsic or "in the eye of the observer". But Jonathan's idea was crucial as far as I was concerned in clearing the way to exploring these models without first doing a giant search for causal invariance.

It took another month or so, but finally on August 10 I sent back to Jonathan and Max a picture we had taken, saying "The origin picture ... and .... *I'm finally ready to get to work*!"

I generated thousands of screenfuls of visualizations:



I think if I had lived a century earlier I would have been a zoologist. And what I was doing here was a kind of zoology: trying to catalog the strange forms and habits of these rules, and identify their families and phyla. It was a glimpse into an unseen part of the computational universe; a view of something there was no particular reason that us humans would have a way to understand. But I was pretty sure that at least some of these rules would connect with things we already knew. And so I started to hunt for examples.

Most of what I do on a daily basis I can do on just one computer. But now I needed to search millions of cases. Conveniently, there's pretty seamless support for parallel computation in the Wolfram Language. So soon I'd commandeered about 100 cores, and every computation I could immediately parallelize. (I was also set up to use external cloud services, but most of the time I was

Given a better understanding of space in our models, we started looking more carefully at things like my old derivation of the Einstein equations for gravity. Jonathan tightened up the formalism and the mathematics. And it began to become clear that it wasn't just a question of connecting our models to existing mathematical physics: our models were actually clarifying the existing mathematical physics. What had been pure, abstract mathematics relying on potentially arbitrary collections of "axiomatic" assumptions one could now see could arise from much more explicit structures. Oh, and one could check assumptions by just explicitly running things.

Doing something like deriving the Einstein equations from our models isn't at some level particularly easy. And inevitably it involves a chain of mathematical derivations. Pure mathematicians are often a little horrified by the way physicists tend to "hack through" subtle mathematical issues ("Do these limits really commute?" "Can one uniquely define that parameter?" Etc.). And this was in many ways an extreme example.

But of course we weren't adrift with no idea whether things were correct—because at least in many cases we could just go and run a model and measure things, and explicitly check what was going on. But I did feel a little bad. Here we were coming up with beautiful mathematical ideas and questions. But all I could do was barbarically hack through them—and I just kept thinking "These things deserve a mathematician who'll really appreciate them". Which hopefully in time they'll get.

As we went through November, we were starting to figure out more and more. And it seemed like every conversation we had, we were coming up with interesting things. I didn't know where it would all go. But as a committed preserver of data I thought it was time to start recording our conversations, as well as my own experiments and other work on the project. And altogether we've so far accumulated 431 hours of recordings.