Declaration Of Independence: This article is based on many diverse sources in print, lecture and private rumor. Errors are however entirely and completely mine.

HUPAPP-94/102

Through Strings to Cosmic Strings And Why

Tristan Hübsch
Department of Physics and Astronomy
Howard University, Washington, DC 20059
thubsch@howard.edu

Contents

Dec. ’94; minor corrections: April 2003

–I prefer a clear statement subsequently disproved,
to a misty dictum . . . which can be
welcomed as a "great thought".
Bertrand Russell

1. The Name of the Claim

Physics may be defined as the discipline of understanding Nature. This definition is probably as good as any other of which I am aware, although (and perhaps precisely because) quite some of what will follow is needed merely to specify more precisely what exactly is meant by "*Nature*", "*understanding*" and "*discipline*", *i.e.*, what is the nature of disciplining our understanding.

To wit, true to the meaning of the Greek original word, physics *does* pertain to all aspects of Nature. Molecular phenomena are subject to chemistry, separate but well integrated with physics. When studying events of continental proportions, the study is called geology (but areology on Mars), while events which are orders of magnitude bigger (planetary, stellar, galactic, cosmic) become labeled astronomy. Things alive are subject to biology, but life itself may well be rooted in quantum physics [17]. Stretching this further, even thinking and feeling phenomena (a.k.a. psychology) may turn out to be caused and determined by physical processes, whereupon social events may be viewed as 'many-individual psychology', much as thermodynamics is 'many-body mechanics'. 1)

Of course, merely slapping a common name on all of it achieves little beyond annoying those who would prefer otherwise. So, hoping to have provoked the Reader to speculate along (or against) such all-encompassing unified avenues of human understanding of Nature, we turn towards more conservative but much better established topics.

A telegraphic review of some of the key issues in contemporary 'fundamental physics' will be attempted herein. To reach this 'fundamental' level, a journey through diminishing sizes and increasing energies will be undertaken. The splendid tourist guides (see [6,7,10,12,13,15,18,19], to name a few) are certainly a well known and excellent resource, the literary, educational and entertaining value of which this modest review cannot hope to approach. Instead, our utility should lie in marking *recent* additions to this framework, pointing out some of the less well-traveled paths, shortcuts and pitfalls and providing a general introductory tour for the newcomer and the casual surfer.

Specifically, this review aims to present the case of (fundamental) strings: what are strings, where does stringiness take place, why strings and not points or something else, and how strings thread into our evolving picture of Nature. First, however, we wish to review the (methodo)logical framework underlying these efforts.

–Insight, untested and unsupported,
is an insufficient guarantee of truth.
Bertrand Russell

1.1. Things may be not as they seem

Although an adept historian will promptly quote an earlier source of this thought, I should like to introduce this *Leitmotif* as the Copernician legacy. The willingness to abandon the 'obvious', the 'common wisdom', the 'plain reason' for an unorthodox thought is certainly the essential element of this.

---

1) This paragraph is obviously aiming to be provocative. Note, however, that it is not stated that any one field of research *reduces* to physics, but 'only' to 'derive' from it and that it is 'caused and determined' by it. Similarly, the behavior of water in a brook 'derives' from hydrodynamics, but the additional ideas of nonlinear dynamics, turbulence and chaos are necessary for its fuller understanding.

However, not just any unorthodox thought: a lunicentral or iovicentral system offers hardly an improvement. Rather crucially, heliocentricity *simplifies the concept* of the planetary system, by making it more uniform. Although Copernicus's system still assumed circular orbits and so needed corrections [2] (epicycles), the model *is* conceptually simpler; perhaps this can even be regarded as a variant of Occam's razor. While this is not yet Newton's universal law of gravity, it does have this unifying flavor. Also, the ultimate test for the model is recognized: the positions and movements of the planets as calculated (now more easily) from the heliocentric system agree with the astronomical observations. Underlying this is the sequence *observe-model-predict*, which may be thought of as the *Contrapunkt* to the above *Leitmotif*. Indeed, eyes look but the mind sees.

Examples to the above start at such elementary levels that the one hardly notices this:

1. The shadow of an object can easily be many times larger than the object itself and will more often than not have distorted proportions as compared to the original object. Yet, only very young children are frightened by the shadow of a wolf or a monster, however masterly produced by the shadow-puppet artist.
2. Standing in a large plain (without mountains in the horizon) the Earth does look flat. Yet, already the ancient Greek geographer Eratosthenes (c. 276–195 BC) has not only known that the Earth was round, but has also calculated its size (to within 10–15%!). The calculation was based on the shadow lengths at summer solstice noon in Syene and in Alexandria, the distance between them and elementary (by today's standards!) geometry. However, both Eratosthenes's results and reasoning have become 'politically incorrect', repressed and forgotten for some sixteen centuries to come, and have been rediscovered in the West only in the Renaissance. Although most of late 20th century humans have no difficulty accepting that the Earth is round, when (if?) humankind becomes truly space-faring, the once obvious flatness of the Earth may even become *incomprehensible*; much as once its roundness was.
3. Everyday experience makes it plain: the Sun and the Moon revolve around the Earth. And so had the mainstream ancient Greek school of astronomy maintained, as compiled by Ptolemy (c. 100–c. 165), and having repressed the unorthodox ideas of Aristarchus (c. 310–230 BC), who not only proposed the heliocentric system but also estimated that the Sun is 20 times further from the Earth and also 20 times bigger than is the Moon [3]. Again, it took a good sixteen centuries for the West to rediscover these.
4. To the 'naked eye', our blood looks pretty smooth and homogeneous. And so it was believed to be until 1683, when the Royal Society published the first detailed pictures of red blood cells, seen through a microscope and drawn by Antoni van Leeuwenhoek (1632–1723). In 1932, Ernst August Friedrich Ruska (1906–1988) designed the first electronic microscope, the modern versions of which make it possible to *see*—as directly as ever—the individual molecules and even atoms comprising the matter which surrounds us.

The Reader will undoubtedly have no difficulty in extending this list by many other and possibly quite more interesting and amusing examples, wherein our underlying *Leitmotif* becomes obvious. The standard human senses, so well adapted to daily routine do

---

[2] Only upon Kepler's *ad hoc* postulate of elliptic orbits (*a posteriori* explained by Newton) will heliocentricity acquire truly convincing technical simplicity and accuracy.

[3] The error in Aristarchus's result is entirely owing to poor measurements of the time, the reasoning and geometry being impeccable.

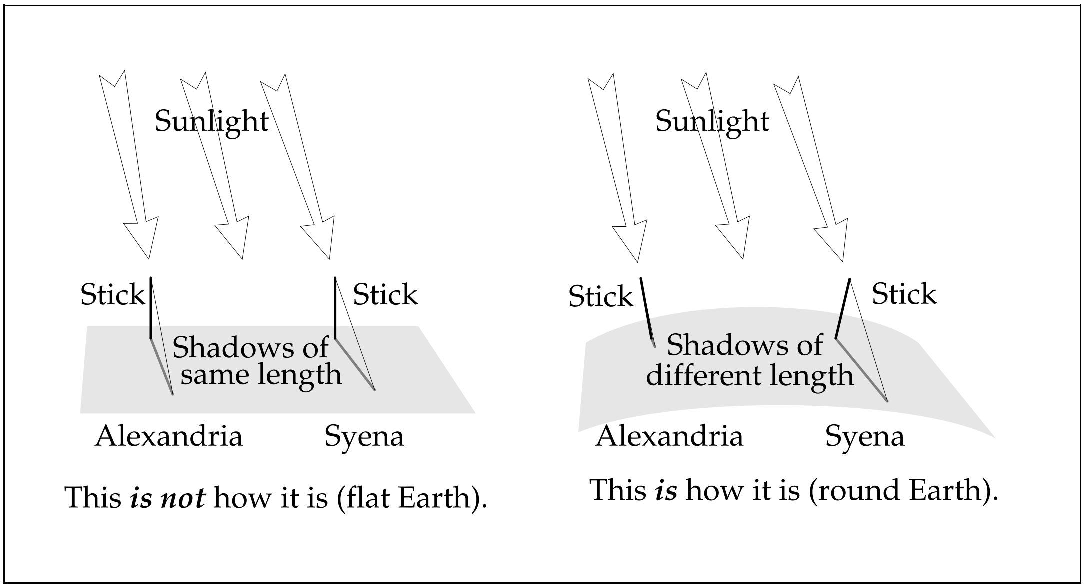

Figure 1: Eratosthenes's analysis which, with measurements of the angles and distances (not sufficiently accurate 200 BC), produces the size of our Planet.

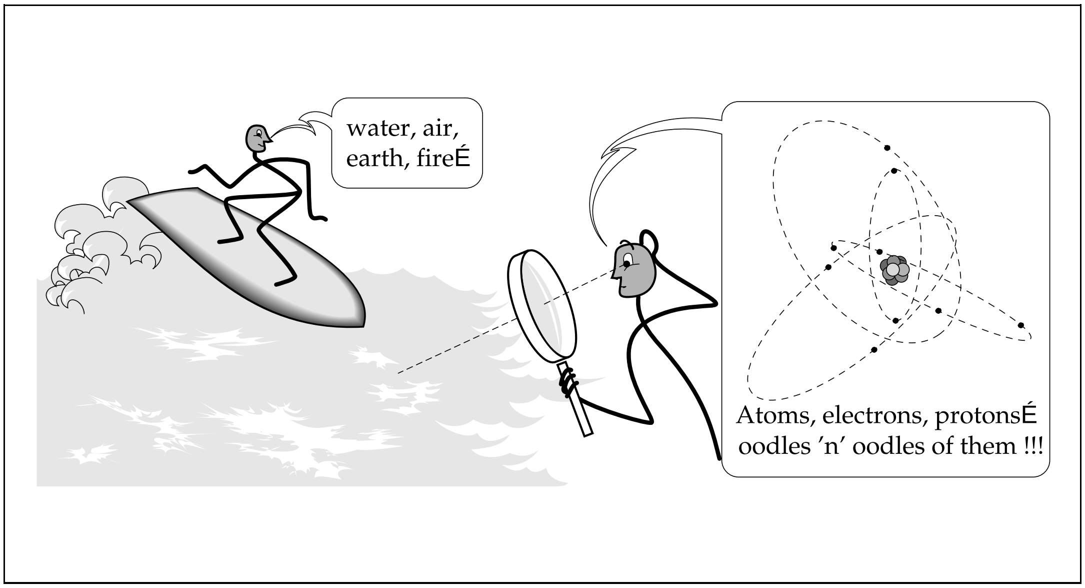

Figure 2: What on size-scales characteristic of human bodies may appear smooth and homogeneous, may look very differently under magnification.

not serve reliably when regarding proportions and perspectives which are not of daily routine. From the typical daily vantage point and human (characteristic) size, planetary and moreover stellar situations appear distorted and we must employ our (patiently trained) mind to correct the image. Indeed, having once learned, the Sun in the sky never looks the same again: we can always envision the Earth, on which we stand, to rotate about that star called the Sun. Similarly, having learned about the blood cells, our mind's eye has no difficulty "seeing" the erithrocites streaming with the blood plasma through our veins.

Yesterday's incredible and arrant nonsense may (and we tend to remember very well when it does) turn into today's plain and simple truth…

Surely, however, not *indiscriminately* so; for, the obvious questions are: which 'truths' to question and how to determine 'truthfulness'? Following Descartes's line of reasoning, *everything* that possibly (and self-consistently) can be questioned—must be questioned. However [4], physicists tend to be more pragmatical. Somewhat akin to "if it ain't broke, don't fix it", physics models and theories are being re-examined and questioned when they start predicting things which do not happen, or fail to predict things that do happen.

> *–Wovon Man nicht deutlich sprechen kann,*
> *Darüber muss Man schweigen.*
> Ludwig Wittgenstein

1.2. The black box paradigm of learning

To formalize this a little bit, let us represent the system under scrutiny as a *black box* and start out with not knowing anything about its contents. What follows may then be regarded as the three pillars of (exact, natural) science:

I To learn about the box's contents, some (controlled or otherwise known) *input* is sent onto the box and the *output* being observed. The 'input' may be something as simple as knocking, or shaking, or perhaps more sophisticated such as X-rays or ultrasound. The 'output' is whatever …well, output there is; for example, as the box was shaken, its weight might have moved about in a way that indicates that the weight is concentrated in several separately mobile subsystems within the box. Or, the box might have rang hollow to knocking. Or, the X-rays might have shown skeletons of three mice…

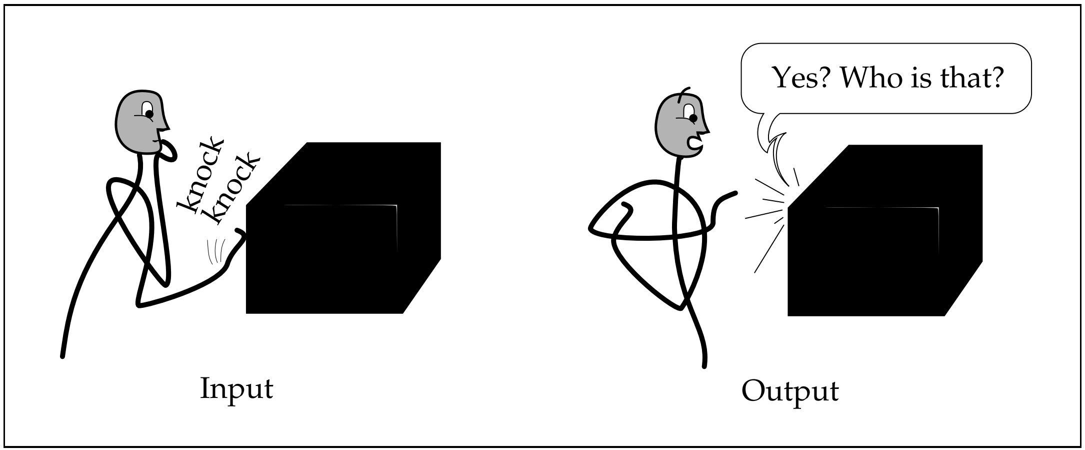

Figure 3: The Black Box paradigm of an experiment.

II Given the information about the box in the form of 'response-to-the-input', both of which suitably *quantified*, we develop a *mathematical model* which faithfully reproduces all of the registered outputs in response to the corresponding inputs. Note that both inputs and outputs must be measured and so will be known only to within some

---

[4] … and even without nitpickingly concluding that Descartes's *cogito, ergo sum* leads to solipsism, or recalling Hume's demonstration of just how detrimental such questioning tends to be…

error, defining the *resolution*. Clearly, the mathematical model can never be guaranteed to be more precise than this resolution, and we refer to it as the resolution of the model as a whole.

III The mathematical model is now used to make predictions: to calculate the response to new, never before tried inputs, which are then tested, if and however possible.

Herein then lies the rationale for 'what and when to question' and 'how to determine truthfulness'. A physical (and more generally, scientific) model needs to be re-examined, with one or more of its "ingredients" questioned and perhaps replaced if it fails to reproduce and correctly correspond to Nature—to within the resolution of the model. Amusingly perhaps, this reveals that what is commonly called 'exact science' is always in error, but exact about how much so.

Clearly, this three-step process of 'observe-model-predict' will be repeated on and on, as guaranteed by Gödel's incompleteness theorem 5), since the subject matter is sufficiently complicated and is not straightforwardly exhaustible (as, for example, tic-tac-toe *is*, viewed as a game of strategy) [9]; see also appendix A. Once a model is devised, predictions are being extracted, and as human ingenuity improves the technology and the techniques, these predictions are extracted and tested at ever increasing precisions. Sooner or later, these new tests highlight a shortcoming of the model, whereupon further data is gathered and the model is modified and extended so as to incorporate *also* this new data. Once the improved model satisfactorily reproduces this, yet further predictions are extracted and tested, which eventually indicate yet further directions in which the model needs to be modified and extended, whereupon yet newer and newer predictions can be extracted, and...

A little thought about this framework reveals something truly extraordinary! A scientific model 6), as described here, becomes modified and extended—never annihilated! The proverbial "back to the drawing board" often occurs in scientific research, but never means trashing established theories. Agreeably, the preceding sentence may be declared a tautology, for 'established' may be taken to mean 'those which never become trashed'. Yet, over the past three centuries of experimental physics, the *fundamental theories* were never trashed, only extended and frequently—merged.

The reasons for this may be found by comparing scientific models with earlier (mainstream) approaches and doctrines: science unifies the inspiration of (experimental) induction with the rigor and extensiveness of (mathematical) deduction.

Amusingly, both Eratosthenes's and Aristarchus's cosmological achievements mentioned above rely on quantitative measurements and analytical mathematics; rather before their time. Dormant for almost two millennia, this combination was tried again, and methodically and persistently by Galileo... and physics (the discipline of understanding Nature) engaged in warp drive. Roughly, the measurements are used to translate between

---

5) ... and barring the bleak possibility of the scientific spirit becoming extinct or exterminated...

6) By *scientific model*, we understand the mathematical model together with its concrete interpretation; that is, the formulae, flowcharts, program, together with the physical meaning of the symbols, *i.e.*, a dictionary between the symbols of the mathematical model and quantities in Nature.

quantities describing the natural phenomena under consideration and the corresponding quantities in the mathematical model. The mathematical model is then used as an impeccable and relentlessly rigorous tool for deduction which incorporates all the data and produces predictions. So, Newtonian mechanics is not invalidated but extended by Einstein's relativity: in the regime when all the speeds occurring in a system are much smaller than the speed of light in vacuum, the relativistic corrections to the Newtonian mechanics are negligible. As at least some of the speeds increases, the corrections become relevant and Newtonian mechanics is no longer a good approximation (the limits of error, about which physics must be exact, become intolerably big), and we might as well use "the relativistic formulae".

This insisting on the 'observe-model-predict' cycle immediately discards "theories" such as that of *phlogiston*, which was supposed to be the æthereal substance of heat; this being a "theory" which neither explained nor predicted any quantitative data, and so was a theory only in a… well, literary sense. Similar fate was met by the "plum pudding" model of the atom, which explained and predicted very little (most of it wrong), but which was even by its originator called humbly a 'model'—a hypothesis perhaps worth looking into and to be mercilessly discarded if wrong; which it was, both wrong and discarded.

The absolutely crucial point here is that what aspires to be called a (scientific) *theory*—not to speak of a *fundamental theory*—must be testable, at least in principle. This of necessity means that the theory must be *quantitative*, *i.e.*, that the theory explains and predicts experimental data, which then can be tested. Note also that the quantitative prediction may be something as simple as a yes/no outcome, a single bit of information; whether a single bit or a googolplex (7) of them—but: new information.

A word of caution, however: "can be tested" does not mean that you can go to your local experimental shop and have the results while-you-wait. Neither does it mean that our Planet's Budget will be allotted to perform the experiment (not that there will be a Planetary Budget any time soon). Nor does it mean that anyone has the faintest idea how to design and perform the actual experiment, even with a Pan-Galactic Budget and a post-*Star Trek* technology. However, the theory must be "testable, in principle". Needless to say, theories which can be readily tested can also be either readily promoted into "established"—as far as it is known (a piece of 'fine-print' rarely stated explicitly, but understood implicitly), or readily discarded should they fail to conform to Nature.

It definitely cannot be overemphasized:

Theories which can, in principle, be disproved are scientific.

Amusingly, *[in Chinese,] a word that can be negated is a verb.* (Diane Wolff [20])

Note also, that it is logically impossible for science to be exact without being quantitative. That is, 'exactness' must be taken to mean that at least a framework of questions can be developed each of which requires a yes/no answer, whereupon the pattern of answers (easily written as a binary number) may be regarded as the quantitative characterization of events which are to be modeled and predicted (*i.e.*, the predicted data). That these yes/no

---

(7) A *googol* equals 10^{100}–one followed by one hundred zeros; a *googolplex* equals $10^{10^{100}}$–one followed by one googol zeros. For a comparison, there are only about 10^{80} particles in the Universe.

answers may obey a statistical (probabilistic) distribution is merely a *technical* complication and does not at all diminish ‘exactness’ in this sense [8]. This is certainly always true of all branches of physics; if statistical physics and quantum mechanics are probabilistic, that merely complicates the techniques and dictates the style of questioning. In fact, it is in part the sophisticated (mathematical) techniques of statistics that specify which questions make sense and which of them do not, and among those that do make sense, which questions have a definite answer and which ‘merely’ a probabilistic one. Temperature, as an average of energy, can be predicted exactly, but the energy of a single molecule will be subject to (predictable) fluctuations and so probabilistic (assuming that a specific single molecule could somehow be tagged and followed unobtrusively).

Viewed from such a vantage point, physics in particular and science in general may be accused of being pragmatic; and, to quite some degree they are. But, it is this science which can bring Moon-rocks and pictures of Io’s and Ganymede’s surface to the Earth; it can produce artificial heart-valves which are not rejected by the human body; it can detect early signatures of hurricanes, cyclones and tsunamis and warn the endangered population. Unfortunately, it can also make our Planet glow in the dark of the Universe, for centuries to come [9]. However, science not only *works*, but in doing so, it shapes our ways of thinking (and so influences just about anything else), and so is considerably more than ‘just pragmatic’.

The foregoing also manifests the price we must pay: although physics is about Nature, it never describes Nature itself, it merely produces models which are sufficiently (and ever more) accurate. So, for example, the statement “Rutherford’s planetary model of atom consists of a nucleus at the origin and the electrons orbiting about the nucleus” does not state that an actual atom actually consists of an actual nucleus… Rather, it states, that a mathematical model based on such a contraption accurately reproduces the actual observations. As it turns out, the fact that atoms are stable requires that Rutherford’s planetary model be modified with additional ‘quantization’ rules, which lead to the Bohr model and the ‘Old quantum mechanics’. Later observations caused further developments, leading through what is now called ‘quantum mechanics’, then ‘quantum field theory’ and most recently, quantum (super)string theory. Most importantly, each of these contains the previous ones, but in all fairness, it ought to be said that the current candidate for *the fundamental theory*—the (super)string theory—has (by far) not been fully verified as a theory of Nature; it may be so, but (super)strings have not yet been shown in all detail to reproduce the “real world”.

1.3. Scientific predictions: powerful and inevitable

–Gravity couldn’t care less for your foot
under the falling stone.

A few more remarks about scientific theory and then we’ll be off to (super)strings. We have already met the “three-step cycle” of developing, testing and refining theories, and the logical possibility that this cycle may continue forever. Indeed, this may be viewed as a curse or a blessing—that there will not be an *ultimate and complete* theory of everything.

---

[8] Remember: exact science is always in error, but exact about how much so.

[9] Ironically, a Zen practitioner’s response to a solipsist’s question about the reality of the world might be a kick with a stick.

What is of recent called “theory of everything” refers to all fundamental (elementary) interactions of the Nature. But, supplying all the bricks and mortar does not yet a cathedral make. Even when the “ultimate” theory of all fundamental (elementary) interactions is known, there is a looooong way to go from there to atoms, to molecules, to… us and our ambitious thoughts, and beyond.

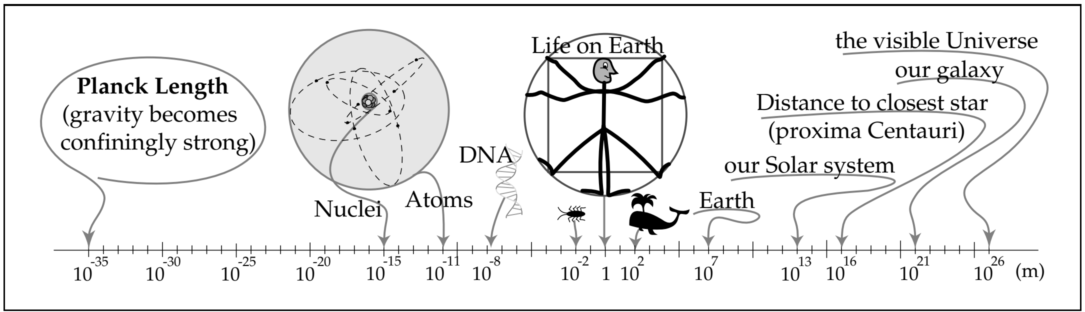

Figure 4: A logarithmic scale of sizes, from the Planck length where everything looks like a black hole (and from within which information cannot be extracted), to the largest distances, from which the light only now reaches us (and from beyond which information has not yet reached us).

Not only is there much room for ‘filling in the details’ whilst following this 1-dimensional arrangement by size, but very often tiny portions in this all-encompassing size scale produce ‘pockets of knowledge’ not infrequently of fantastic and baffling complexity—a characteristic perhaps not unlike fractals. Suffice it here just to mention that the complexity of collective phenomena (behavior beyond the ‘thermodynamic’ average, such as eddies, tornadoes, the shapes and the dynamics of clouds, market crashes…) has only recently been subject to serious scientific thought. Also, life as we know it—and so biology—occupies merely a few orders of magnitude, roughly between $(10^{-6}–10^{2})\,m$; chemistry occupies an even smaller niche around $10^{-9}m$.

Following this… ahem… glory road of scientific discovery, it is important to realize that *whatever can be deduced from the assumed axioms/postulates is a prediction of the model.* In other words, if a model reproduces perfectly the original input data, and produces a bunch of testable predictions, but if even only one of these predictions turns out to disagree with Nature—the model is wrong. It may happen that a minor modification of the model will both cure the glitch and retain its fidelity otherwise; if so, this modification must be incorporated as an integral element of the (revised) model, subject of course to any further test that can be conjured. If such a revision or extension cannot be found—off with its head 10).

The world, unfortunately, is real; I, unfortunately, am Borhes.
Jorge Louis Borhes [2]

---

10) Fortunately, theories unlike their human inventors can be resurrected, and this happens on occasion. The glitch which was formerly considered lethal may turn out to be ‘repairable’ at a later time, when a better understanding of the model and requisite techniques and methods of analysis is attained. Of course, it can (and does) also happen that the experiments are flawed in a way which is revealed only much later, and the corrected analysis of which turns out to agree with what was formerly thought of as a glitch.

All predictions are forced upon us as an inevitable consequence of the given assumptions and ensured by the rigor of mathematical deduction—sometimes foreseen, sometimes not, and sometimes spectacularly unexpected!

> *The Heitler-London bondage is a unique, singular feature of the [quantum] theory, not invented for the purpose of explaining the chemical bond. It comes in quite by itself, in a highly interesting and puzzling manner, being forced upon us by entirely different considerations.* Erwin Schrödinger [17]

The Heitler-London bond is one of the basic 'ingredients' in modern Chemistry. Similarly spectacular was P.A.M. Dirac's prediction of the anti-electron (and with it, as a logical consequence, anti-everything), and W. Pauli's prediction of the particle called *neutrino*. (The latter was confirmed experimentally only two decades later!)

Further remarks and details about this mathematical framework and its relevance to scientific models may be found in the appendix A; we now turn to our journey through decreasing sizes.

2. Towards Basics

> *–I think we may yet be able to understand atoms.*
> *But in the process we may have to learn*
> *what 'understanding' really means.*
> Niels Bohr, to W. Heisenberg [10]

Following the adage 'when eating an elephant, take one bite at the time', physics *analyses* natural phenomena (systems), identifying their subprocesses (subsystems). These are typically easier to understand, whereupon it remains to integrate the so understood subprocesses (subsystems) into the whole. In doing so, certain characteristics of the whole are identified as merely a conglomeration of those of the parts, whereas other properties are essentially 'collective'—unexplained by the characteristics of the parts themselves and inextricably rooted in the complexity of the whole. While analyzing the 'parts' says little about collective phenomena, it certainly allows specifying and discussing properties which are *not* collective, leaving thereby a cleaner approach to this next frontier.

2.1. Smaller and smaller and ...

> *–"What a curious feeling!" said Alice.*
> *"I must be shutting up like a telescope!"*
> Lewis Carroll

In good measure, subdivision of phenomena, processes and systems does proceed akin to the most obvious example of the black box paradigm: that of microscopy. Light is shined onto the object under scrutiny (the black box, figuratively) and the *reflected* light is then guided through a system of lenses and/or mirrors to form an enlarged image for the observer to see. The difference between the so reflected light and what would have reached the observer's eye if the object were removed is the image of the object in contrast to the background. As an enhancement and extension of our natural senses of vision, a microscope is used to (rather literally) look into the structure of various material objects. In doing so, an important limitation is realized. Standard optical microscopes cannot possibly resolve structures finer than $10^{-6}\,m$, regardless of the precision and perfection of the optical elements: the lenses, mirrors, *etc.* The reason for this is the wave nature of visible light, the wavelengths of which range between ~ 380–$760\,nm$ ($1\,nm = 10^{-9}\,m$).

When the object under scrutiny becomes of such or smaller size, the light bends (diffracts) around it, and the image blurs beyond recognition.

This is exactly analogous to the fact that sounds (commonly audible by humans) easily bend around human-size objects. Indeed, it is impossible to *hear* the presence of, e.g., a totally silent person who stands between us and another person talking to us. The wavelength of humanly audible sounds range within about $17\,mm$–$17\,m$ ($1\,mm = 10^{-3}\,m$) and all but the shortest wavelengths (which few humans can hear very well and which are even then typically masked by sounds of longer wavelengths) easily bend around human-size objects. One says that the *resolution* of a wave is of the order of its wavelength, meaning that only larger objects can be imaged well—*resolved*.

The reader will however recall that ultrasound *can* be used for imaging human-sized targets—it is routinely used to 'sonograph' fetuses *in utero*. As higher frequency corresponds to shorter wavelength, the resolution of ultrasound is better, *i.e.*, finer details can be imaged with ultrasound than with humanly audible sound. Now we recall that humanly visible light is but a narrow portion of the spectrum of electromagnetic radiation. In perfect analogy, then, electromagnetic radiation of frequencies higher than those of visible light (and so shorter wavelengths) should provide finer resolution in appropriately designed microscopes. Indeed, there are many types of electromagnetic radiation with wavelengths shorter than those of visible light (ultra-violet light, X-rays, ...) and these may be used to design more powerful microscopes. In practice, however, the design of such microscopes is hampered by the fact that there are few if any materials that could act as lenses: once the wavelengths are much smaller than those of visible light, usual optical lenses no longer bend the radiation.

To the rescue comes the quantum nature of Nature: matter particles, such as electrons, also exhibit wave-like characteristics, and the basic relation is that the wavelength of the probing beam is *inversely proportional* to its energy. (Even a single electron can exhibit wave-like behavior, so "beam" here refers indiscriminately to single and many particles, as the case may be.) Table 1 provides several objects and events in Nature, together with their characteristic sizes and the corresponding energies; that is, the listed energy is the minimum required of a probe to have so as to be able to resolve the details of the characteristic size.

Thus, for example, to any probe (beam, ray, test-particle, radiation, ...) with energy less than about $10\,keV$, typical atoms appear as indivisible, structure-less objects. Of course, a probe with much less energy would not even 'see' an atom, but instead only the (much) larger structure of which the atom is a part. One needs a probe with more than about $10\,keV$ energy, to which the structure of the atom will begin to show. Of course, with the increased energy, the probe is more and more likely to disrupt the observed target, or at least modify its characteristics, so that what is observed is not entirely the property of the target, but the target-probe interactive system.

This non-negligibility of the probe and its interaction with the target is essential is is a built-in feature of the quantum theory. In this sense, the probing and observing of the system alters the system irrevocably, which is sometimes expressed by stating that quantum observations—and so all knowledge—are *participatory*. This causes an *inherent* uncertainty in all possible forms of probing, and so also in knowledge; this is expressed in

Objects, Events	**Size (in m)**	**Energy (in eV)**
Crystalline lattice spacings	$\sim 10^{-10}$	$\sim 10^{3}$ ($\sim 1\,keV$)
Size of atoms (2×Bohr's radius)*	$\sim 10^{-10}$	$\sim 10^{1}$ ($\sim 10\,eV$)
Size of nuclei (roughly) (strong nuclear force range)	$\sim 10^{-15}$ ($1\,fermi$)	$\sim 10^{8}$ ($\sim 100\,MeV$)
Size of protons, *i.e.*, hydrogen nuclei (energy equiv. of the proton mass)	$\sim 10^{-16}$	$\sim 10^{9}$ ($\sim 1\,GeV$) ($= .936\,GeV$)
Electro-weak force range	$\sim 10^{-18}$	$\sim 10^{11}$ ($\sim 10^{2}\,GeV$)
so-called "Grand Unification"	$\sim 10^{-31}$	$\sim 10^{24}$ ($\sim 10^{15}\,GeV$)
Quantum gravity, (super)strings	$\sim 10^{-35}$	$\sim 10^{28}$ ($\sim 10^{19}\,GeV$)

*Bohr's diameter is $\frac{\alpha}{2\pi} \sim 10^{-3}$ times smaller than the de Broglie wavelength.

Table 1: Some 'landmark' objects and events, and their characteristic sizes and the corresponding characteristic energies. Compare also with Fig. 5. One *electronvolt* ($1\,eV$) is the energy needed to transport an electron against the electrostatic field across the potential difference of one volt.

Heisenberg's "uncertainty principle", which indeed may be regarded as the fundamental postulate of quantum theory.

The uncertainty principle is however extremely precisely stated, and moreover—in terms of quantities defined in classical (pre-quantum) theory; thus again, quantum theory does not obliterate but generalize the classical theory. To each variable used in the description of a physical system, classical theory assigns a precisely defined *conjugate momentum*; let q and p be one such pair. The uncertainty relation then states that

$$\triangle q \, \triangle p \quad \geq \quad \hbar \ ,$$

where $\hbar = 1.0507{\times}10^{-34}\,J{\cdot}s$, and where $\triangle q$ and $\triangle q$ is the uncertainty in observing and measuring q and p, respectively. Thus, if the position of a particle is measured to within, say, $10^{-15}\,m$, its momentum cannot possibly be determined better than $0.5254{\times}10^{-19}\,kg\,m/s$. Typically, errors induced by the imperfection of the apparatus will be bigger than this, but there do exist measurements where this essential uncertainty does show. Again, exact science is in error, but exact about how much so. Furthermore, measurements of another variable, q', which is independent 11) of q and p, does not affect the measurements of q and p. That is, having measured q' with whatever accuracy provided by the detecting apparatus, the simultaneous measurement 12) of *either q or p* is similarly

---

11) The technical requirement is that q' should *commute* with both q and p.

12) In the context of quantum mechanics, 'simultaneous measurement' does not mean 'at the same time'—in fact, that is most often impossible. Instead, it implies a successive measurements of two quantities, which is independent of the order in which the measurements are done.

restricted only by the resolution of the measuring apparatus.

2.2. Breaking up is hard to do

–The whole may be other than the sum of parts.

The alert Reader will have perhaps puzzled over the persistent use of 'structure' and 'sub-structure', but not 'divisibility'. Indeed, the latter is often taken as synonymous with 'composite', implicitly assuming that a composite system, one which exhibits sub-structure, can be somehow divided into its constituents. Alas! this is merely a prejudice born from daily experience. An egg, once broken into an omelette, cannot be put together into whole; a broken porcelain plate may perhaps be reconstructed with the aid of super-glue, but it will never be quite the same again.

However, something curious happens when something as small as an atom is divided. Consider ionizing an atom of hydrogen: separating its nucleus (a proton, with positive electric charge) from its negatively charged electron. This can be achieved, e.g., by applying a sufficiently strong electrostatic field (with $\geq 13.6\,eV$ potential energy). The respective proton and the electron can be removed from each other light-years and light-years (and at least in a *gedanken*-experiment such as this, the rest of the Universe may be neglected). Leave them by themselves for a while and… the electrostatic force reunites them! Owing to the unbounded action-range of the electrostatic force, the electron and the proton which used to constitute an atom were never really separated in the first place; they remained in interacting with each other (through the electrostatic field) throughout the experiment. It should be clear that no amount of complication will overturn this conclusion.

This point of view however immediately raises a new question. The forces which held the pieces of a broken porcelain plate together before it broke also have unlimited action-range! So, how come these do not reunite the pieces (regardless of how long the Reader is willing to wait)? The clue is not just in the action-range, but also in the dependence of the force strength on the distance. The magnitude of the electrostatic force decreases with distance as $1/r^2$, whereas the strength of molecular forces decrease much faster, perhaps as $1/r^6$ or faster. Now, consider testing the effects of some such force at a distance r and assume, for simplicity, that the force-field is spherically symmetric. Then, the same effect would be detected at any point on the sphere of radius r and with the center at the source. Since the surface-area of this sphere grows as r^2, the *flux* of the force-field through the surface is constant. By contrast, molecular forces which decay as $1/r^6$ (or faster) would produce a flux through such a sphere which decays as $1/r^4$ (or faster) and so vanishes at larger and larger distances from the source. Molecular forces are said to be localized, whereas (Coulomb) forces obeying the "inverse-square-law" are said to have infinite range.

Notice that the magnitude of both molecular and Coulomb forces *decrease* with distance. Far-away test-particles are hardly interacting, while the close ones are strongly interacting; that is, low-energy probes are perturbed hardly at all, while high-energy probes (small wave-length, close resolution) are strongly deflected. This is indeed the hallmark of Rutherford's experiments which established the existence of a positively charged nucleus.

This is not so with the force of an elastic spring, which *increases* with distance. Essentially the same type of experiment, this time with highly energetic electrons or protons as the probe and stationary targets of bulk matter or beams of electrons, positrons, protons, antiprotons… on a collision course with the first beam. At energies well above $10^4\,eV$, new phenomena are noticed, ascribed to what became called the strong and the weak nuclear forces. Over the distances where the effects of these forces can be measured, the magnitudes of these forces *increase* with distance, *i.e.*, decrease with the energy of the probe. Given the paradigm of the elastic spring force, this in itself is perhaps not counter-intuitive, however some of the consequences are.

To pull a spring, one must invest energy which increases the potential energy of the stretched spring. Beyond a point determined by elasticity, the spring simply breaks. However, the spring-like strong nuclear forces do not have that advantage of imperfect elasticity, instead, two particles (called quarks) bound by the strong nuclear force may be separated to increasing distances—at the price of increasing the potential (binding) energy. This could go on indefinitely and one could separate two quarks to an arbitrarily large distance and to the satisfaction of doubting Thomases, were it not for the fact that the binding energy soon becomes sufficiently big that it can convert into a particle–anti-particle pair. One of these freshly created particles then binds with one of the 'old' ones and the other with the other, and the attempt to separate two quarks to distances bigger than about $10^{-15}\,m$ fails.

Thus, quarks (to the best of experimental evidence and theoretical prediction) cannot be separated to arbitrarily large distances from each other, but stay confined as separate entities only within distances less than $10^{-15}\,m$—wherein they roam relatively unperturbed. Hence, 'divisibility' (at least, as usually understood) is definitely not synonymous with 'composite'; a proton exhibits its structure (being composed of three quarks) through the scattering patterns in sufficiently energetic collision experiments.

One remark here is in order. The strong forces between the quarks comprising a proton are related but different from the forces between two nucleons, such as a proton and a neutron (in, e.g., the deuterium nucleus). The latter forces are so-called 'residual' forces, just as (some) molecular forces are residual forces stemming from electric (and magnetic) forces. However, even the residual strong nuclear force is much stronger than the electrostatic one, since positively charged protons cohabit in the nucleus. The weak force also exhibits this elastic-spring like force law, but is crucially different from the strong force in that the force-field quanta, the $W^{\pm}, Z^0$ particles are massive; by contrast, the force-field quanta of the strong interaction, called gluons, are massless.

2.3. … and smallest *–Ατομον.*

It should appear that the (spatial) resolution of a measuring apparatus and may, at least in principle, be improved to an arbitrarily fine resolution, but this is not so. A glance at Fig. 4 shows that something unprecedented happens when spatial resolution should be refined to 'look into' details smaller than about $10^{35}\,m$, the Planck length. It is not difficult to see why this should happen. Recall that, in order to 'look into' smaller and smaller details, the probe must have more and more energy. In the process of interaction with the target, the probe-target temporarily form a combined system, the total mass of which is the sum of masses of the target and the probe, *and the mass-equivalent of the energy of the probe.* Thus, as the energy of the probe is increased, so does the total mass of the probe-target system during interaction.

Now, all would be well were it not for gravity and the fact that the gravitational field of a system is (linearly) proportional to the *total* mass of the system. As the gravitational field increases, so does the escape-velocity; moreover, the gravitational field is not constant but is proportional to $1/r^2$, so that the escape-velocity is much bigger near a gravitating center then farther away from it. So then, at a certain energy of the probe, its resolution becomes sufficiently small and it approaches the target so close that the escape-velocity becomes too big for the probe to escape. The probe therefore is swallowed by the target and brings out no information: the target 'looks' like a black hole.

Admittedly, this argument extrapolates many orders of magnitude in size, basing on the current understanding of gravity and quantum mechanics. However, note that the *qualitative* part of the argument only rests on the fact that the resolvable distances *decrease* with growing energy, while the distance at which the escape-velocity becomes too big *increases* with the total mass, and on the equivalence of mass end energy. If moreover the *quantitative* aspects of this argument are also reliable, the minimal resoluble size of $\sim 10^{-35}\,m$ emerges rather straightforwardly.

If Nature is built from elementary *particles* (with no sub-structure by definition), these should appear as miniature black holes. Their *event horizon*[13] should form a closed surface of which no detail smaller than $\sim 10^{-35}\,m$ should be discernible. So, massless elementary particles would 'look' like minimum size, $\sim 10^{-35}\,m$, spherical black holes. Massive ones would have a bigger 'event horizon' and may have more complicated shapes, the smallest details of which however must be larger than about $\sim 10^{-35}\,m$.

The inaccessibility of the 'inside' of the event horizon of an elementary particle indicates that it does not really make sense to treat these as point-like objects — willy-nilly, they acquire an extension in space.

This conflict between the point-like *model* of elementary particles and the extrapolation of the results of general relativity (describing gravity) and quantum mechanics is indeed a prediction of such a combined model, forced upon us by the model itself. To avoid the contradiction, we must abandon some aspect(s) of the model, but in such a way to keep experimentally verified features (up to $\sim 10^2\,GeV$, *i.e.*, down to $\sim 10^{-18}\,m$).

Another (admittedly, also heuristic and not at all rigorous) argument manifesting the incompatibility between general relativity and quantum theory of point-like elementary particles is as follows. Note that Heisenberg's uncertainty principle requires that both position and momentum in the same direction cannot be determined infinitely precisely. On the other hand, in general relativity, matter curves spacetime and so determines it. So, a single (massive) point-like particle curves and so determines spacetime in which the location of the particle is a precisely determined special point. Moreover, the particle is there at rest, whence both location and momentum are known infinitely precisely; hence a contradiction with quantum theory.

2.4. Quantum field theories

–Whatever is not forbidden, may happen;
and the rest, if no one watches.

Technically, the incompatibility between general relativity and quantum theory for point-like elementary particles brings about notorious divergences: when calculating physically

---

[13] The term 'event horizon' denotes the boundary in space enclosing a black hole, from within which nothing can escape on account of confiningly strong gravity.

observable quantities, the results are hopelessly divergent (infinite). By contrast, in the special relativity–quantum theory amalgam, "relativistic field theory", the formally divergent results can be eliminated through a process called 'renormalization'. Indeed, the quantum theories of electromagnetic, weak nuclear and strong nuclear forces (with all the known matter included) do incorporate special relativity, and form a logically consistent framework[14].

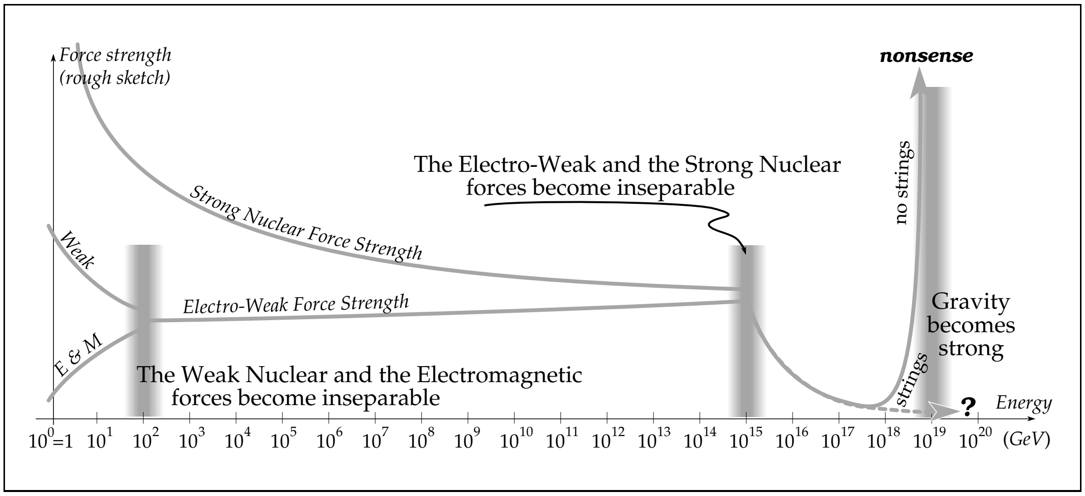

Figure 5: A logarithmic scale of energies, from $1\,GeV$—the energy equivalent of the mass of a proton ($\approx 100,000$ times more than the ionization energy of Hydrogen), to $10^{19}\,GeV$, where gravity becomes confiningly strong and point-field theories become nonsense, while string theories seem merely to undergo a phase transition, albeit not yet well understood.

Modeled on Quantum Electrodynamics (the predictions of which are verified to astounding ten decimal places), the quantum field theory of electro-weak interactions describes the observed electromagnetic, weak nuclear forces and their unification accurately (both theoretical and experimental precision here is rather more modest, but in agreement). Several quantum field theory models have been advanced which describe the predicted unification of the electro-weak force with the strong nuclear force (see Fig. 5), and only further experiments can decide which of those describes the 'real world'.

In all of these, the transitory regime (indicated by the blurred regions in Fig. 5) where the unification takes place is akin to (and is indeed called) a phase transition. Below $\sim 10^2\,GeV$, for example, there is a clear distinction between electromagnetic processes and weak nuclear processes and either of the two can occur without the other. Above $\sim 10^2\,GeV$ however, these processes merge inextricably. This is quite similar to the fact that electric and magnetic phenomena are well distinguished in static systems and frequently occur one without the other, but become inseparably united and give rise to electromagnetic radiation

---

[14] Fact is, the existence of the so-called top quark (for which very convincing experimental evidence has been recently produced) is predicted by so-called anomaly cancellation. That is, without the top quark, the standard model of elementary particles and interactions would not have been consistent!

when charges are set into (relative) motion. The difference between these two kinds of unification—electro-magnetic *vs.* electro(magnetic)-weak—is the *ordering parameter*.

In the case of electro-magnetic, whether the electric and the magnetic field are separable or not depends on the velocities involved, as compared to the speed of light in vacuum ($\approx 300,000\,km/s$). For example, it is well known that a current (a stream of electric charges) creates a surrounding magnetic field, which is constant if the current is. Now, the speed of individual charged particles is typically rather small as compared to the speed of light in vacuum. However, the speed of the *current as a whole* in fact (typically) equals the speed of light in vacuum, as seen from the speed with which a pulse or other type of modulation travels. With no moving charges, the ratio of this speed of the current with the speed of light in vacuum is an *ordering parameter* is zero and the electric field of static charges produces not magnetic field. When charges are moving, this ratio is (close or equal to) one, and the electric field of the moving charges produces a magnetic field. Moreover, if the current is not constant neither is the magnetic field, and its variation produces additional electric field, which in turn... This mutual feedback effect between the electric and magnetic field of a non-stationary current produces the new phenomenon: electromagnetic radiation, whereby part of the energy of the current is carried away.

With the electro-weak ('electro-magneto-weak' would have been more accurate, but is too much of a mouthful) unification, the corresponding ordering parameter is the energy involved, as compared to the mass of the $W^{\pm}$ and Z^0 particles. (By now, a decade after the first detection at CERN, these are routinely observed and studied; the energy-equivalent of their masses are close to $100\,GeV$.) By contrast, the mass[15] of (particles of) light is zero, meaning that the total mass of light is entirely owing to its (kinetic) energy. Clearly, if the experiment involves energies which are much smaller than $\sim 100\,GeV$, the $W^{\pm}$ and Z^0 particles cannot be produced and do not contribute to the processes studied. Weak nuclear processes (mediated by the $W^{\pm}$ and Z^0) then are entirely enacted by *virtual*[16] $W^{\pm}$'s and Z^0's. This is owing to Heisenberg's uncertainty principle and the fact that *energy* and *time* are conjugate variables, whence

$$\triangle E \; \triangle t \quad \geq \quad \tfrac{1}{2}\hbar \; .$$

Roughly, for times shorter than about $(10^{-8}$–$10^{-10})\,s$, the intrinsic uncertainty in energy becomes larger than about $\sim 100\,GeV$, and $W^{\pm}$ and Z^0 particles may be created freely. However, since they must decay within such a short time, the *probability* that they affect the processes is diminished. This provides for an unambiguous identification of any process as either electromagnetic or weak-nuclear. However, once the energies involved in the experiment become well bigger than $\sim 100\,GeV$, *real* $W^{\pm}$'s and Z^0's are produced as readily as electromagnetic radiation (photons). Owing to charge conservation, the W^{+}- and W^{-}-radiation does not mix with the rest, but the Z^0-radiation and the (electromagnetic) photon-radiation mix inseparably and form a new type of phenomenon, just as (variable) electric and magnetic fields combine into electromagnetic radiation.

---

[15] For the meticulous Reader: 'mass', by itself, refers to what is sometimes called the 'rest-mass'—that which is independent of other forms of energy that the particle might possess, or more accurately the mass as measured in a coordinate system where the particle is at rest. 'Total mass', however, refers to what is sometimes called the 'relativistic mass', which includes the rest-mass and also the mass-equivalent of whatever energy the particle might possesses otherwise.

[16] Unlike real particles, virtual ones –by definition– cannot be independently observed.

In both cases, the qualitative features are drastically different on one and the other 'side' of the transition region and we refer to this as a *phase transition*, akin to the transition between ice and water. In Fig. 5, these are represented by the blurred regions.

—o—

A similar phenomenon is expected to happen around $\geq 10^{15}\,GeV$, where the electro-weak interactions become inseparable from strong nuclear interactions. It should however be emphasized that all the graph in Fig. 5 is based on experimental data at currently accessible energies, $\leq 100\,GeV$, and so are necessarily extrapolations. The assumption that there are no new phenomena or particles to be detected between $\sim 100\,GeV$ and $\sim 10^{15}\,GeV$ is often called the 'great desert hypothesis'; in a way, this follows the principle of Occam's razor—no novelty is introduced unless absolutely necessary. The subsequent ideas and arguments rely on this at least in part, and may need to be reexamined once firm evidence is gathered that the 'great desert' is populated. In fact, a number of models are being developed in the attempt to sort out the possible phenomena and particles that might populate this region, yet in such a way that the 'below $\sim 100\,GeV$ physics' is undisturbedly conforming with experiments. Hopefully, such (and any other) models should produce a prediction 'just around the corner', that is, predictions which are experimentally verifiable (or disprovable) in some near future.

Instead of reviewing these models (which would have to be done, more or less, on a case by case basis), we turn to the 'other side' of Grand Unification, to energies approaching the Planck energy, at $\sim 10^{19}\,GeV$, and so sizes approaching the minimal, Planck length at $\sim 10^{-35}\,m$.

Before that, however, a few brief remarks are long overdue: just *what* do these weak and strong nuclear interactions *do*? Besides æsoteric phenomena of particle physics, these are in fact responsible for our own existence, as we know it. Electromagnetic radiation—and in particular light—is what brings the energy from the Sun to the Earth and makes life as we know it possible. The fundamental process which produces the immense energy of our Sun is *nuclear fusion*, in which the nuclei of *deuterium* and *tritium* (two heavier isotopes of hydrogen) fuse into helium and release a neutron and energy. The reason that there is surplus energy owes to the details of—strong nuclear interactions. Finally, note that nuclei of pure hydrogen would not fuse; instead, deuterium and tritium are needed. The nucleus of a hydrogen atom is a single proton; the nucleus of a deuterium atom consists of a proton and a neutron (held together by strong nuclear forces); the nucleus of a tritium atom consists of a proton and two neutrons. So, where did these neutrons come? Notably, the process of β-decay, $n^0 \to p^+ + e^- + \bar{\nu}$ is mediated by weak nuclear forces and its reversed version $p^+ + e^- \to n^0 + \nu$ can and does occur within stars and has also occurred long before the stars were formed. In addition to providing the required fuel for (strong nuclear) fusion, weak nuclear forces also moderate this process, whereby preventing our Sun to burn out in one brilliant explosion.

Thus, by making the Sun burn in the first place, making it burn at a steady pace that we are familiar with, and by bringing its energy to the Earth, the *strong nuclear*, the *weak nuclear* and the *electromagnetic* interactions bring about the conditions on the Earth which sustains our life and our asking about it. Finally, the fourth fundamental interaction—*gravity*—keeps the Earth from flying asunder and also keeps it in a comfortable orbit near the Sun. Were it not for these four, you would not be reading this.

> *–How often have I said to you that*
> *when you have eliminated the impossible,*
> *whatever remains, however improbable, must be the truth.*
> Sir Arthur C. Doyle

3. Strings Unraveled

Thus, after all these lengthy introductions, we turn to (super)string theory as a *model* of the fundamental interactions and processes in Nature.

3.1. Strings 101

> *–Fiat filum.*

The basic idea is rather simple: what has hitherto been regarded as elementary *particles*, will hereafter be replaced by objects of spatial extendedness, and concretely, by loops. So, while particles sweep out world-lines as time passes, strings sweep out cylindrical world-sheets. If a particle should decay into some other two, its world-line has to bifurcate into the world-lines of the resulting two particles (see Fig. 6). Note that this produces an exceptional point on the 1-dimensional manifold (the bifurcating line) representing the particle decay; whatever coordinate (proper time) one chooses on this 1-dimensional manifold, the bifurcating point remains special.

On the other hand, if a string should decay into some other two, its world-sheet splits into two, akin to the legs in pants, or a Y-joint in plumbing (see Fig. 6). Note that there is no special point on this 2-dimensional manifold; in fact, there is not even a special 1-dimensional subspace which would represent the special location of bifurcation. We may choose the time coordinate and the space coordinate in some particular fashion, which will then identify somewhere on the world sheet a figure-8, where the single string 'crosses over' to form two separate strings. However, almost any other choice of coordinates (time and space) along the world-sheet will move this figure-8 elsewhere. Since coordinates themselves have no intrinsic physical meaning (they are not measurable, only distances are), relativity guarantees that there is nothing physical to this figure-8 where the string appears to bifurcate.

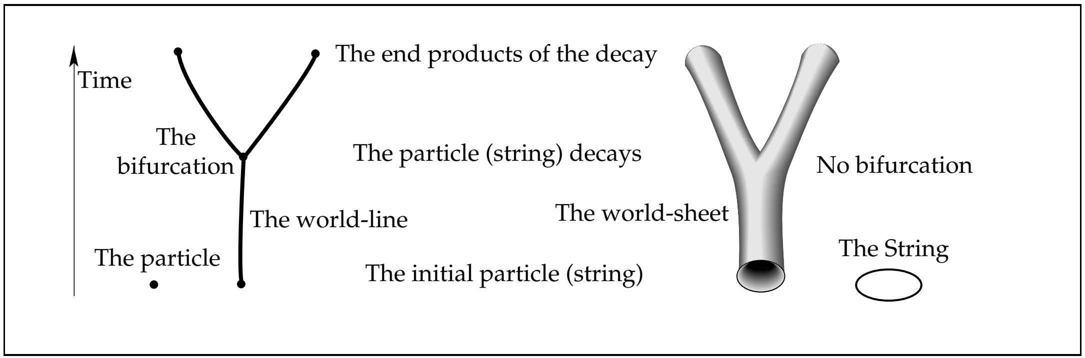

Figure 6: The simple process of one particle (string) decaying into two other particles (strings).

Another way to see this is to note that the bifurcating Y-shaped 1-dimensional manifold is not smooth, while the stringy 2-dimensional pants-shaped world sheet is.

Precisely because of this smoothness, the effect of virtual states (particle, strings) in the two theories is crucially different. Recall that virtual states are those which are allowed to be created within the time allotted by Heisenberg's uncertainty principle. That is, if the characteristic time of a process such as the decay in Fig. 6 is $\triangle t$, the uncertainty relation limits the precision of the measurement of energy to $\triangle E \geq \frac{1}{2}\hbar/\triangle t$: faster processes allow enough leeway in energy to create and swiftly destroy heavier and heavier particles. These temporarily created and swiftly destroyed particles are called *virtual.* Such virtual particles interact with the income and the outcome particles, just as regular particles do, except their interaction is limited to the short time of their existence.

Fig. 7 (on the right hand side) represents two such possibilities for a virtual particle to interact. On the far left, with just the income particle "long" before it will decay, and towards the middle of Fig. 7, interacting with all three, "real" particles. It is quite clear that these two possibilities are fundamentally different. As it turns out, the former diagram produces a harmless redefinition of the mathematical function describing the income particle. However, the latter one causes a redefinition of the decay constant (for the graph on the left hand side of Fig. 6). All would be well if this redefinition were a convergent quantity; trouble is, it isn't. The reason is seen from the fact that the size of the loop along which the virtual particle runs may be arbitrarily small. Since there is nothing to fix the size, continuously many sizes are possible and since the virtual particle is not observed, we must sum up the contributions from each possible size; this integration typically produces a divergent result. In the quantum field theories of electro-weak and strong interactions (with matter), there are numerous such divergent contributions and they fortuitously cancel against each other! This however does not happen with point-like theories of quantum gravity.

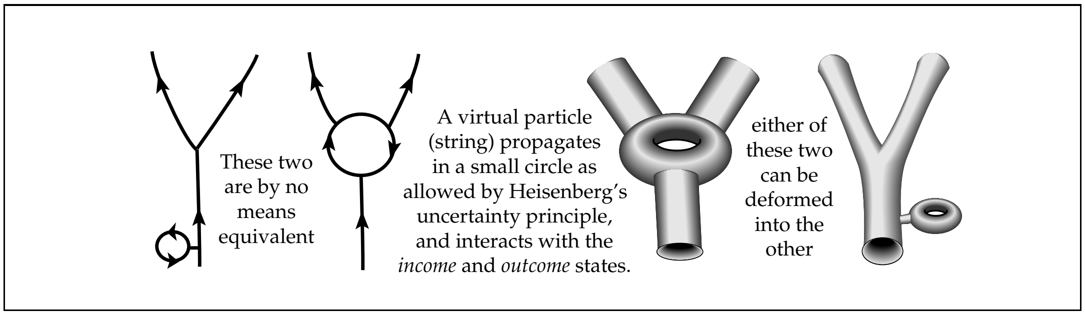

Figure 7: Virtual particles cannot be prevented from modifying the "tree-level" diagrams such as in Fig. 6, and here we include two simple 1-loop corrections.

Owing to the smoothness of the 'pants' diagram (to the right in Fig. 6), its 1-loop stringy correction is sketched on the right hand side of Fig. 7. It should be clear that the world sheet of one diagram can be smoothly distorted so as to become the other. Since such distortions merely reflect a different choice of local coordinates on the world sheet (and these coordinates are unphysical, in that they are unobservable), the two diagrams are in fact one and the same. The diagram to the far right of Fig. 7 represents a 1-loop redefinition of the mathematical function which describes the income state and this occurs well before the interaction (decay), so the stringy 1-loop diagram does not redefine the decay coupling constant of the diagram in the right hand side of Fig. 6.

This rather heuristic argument can indeed be made both more formal and more rigorous, since there is a well-defined dictionary of translating each of the diagrams in these figures into concrete calculations and the heuristic arguments are borne out by concrete mathematics.

Recall also the conflict between the ideas of general relativity and the uncertainty principle: that is no longer a conflict with strings, since strings are by definition not localized to a point. In fact, what becomes identified with "particles" in string theory are *vibrational modes of the string.* Vibrations of a loop in a plane are easily depicted as undulations (see Fig. 8). The energy of an individual mode grows with the mode number (counted as the half of the number of nodes, the points on the loop which are fixed during the particular vibrations).

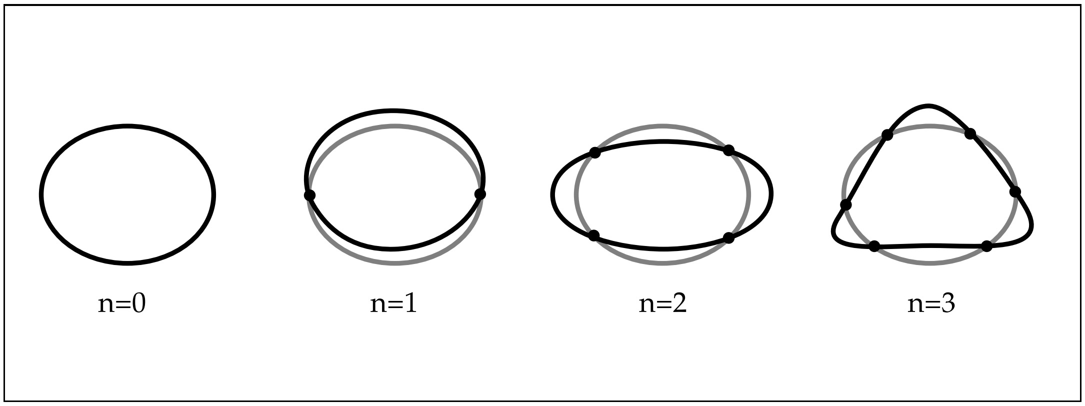

Figure 8: Vibrations of a planar loop; they may be counted by the number of nodes (black dots).

Now, all string theories *automatically* produce a type of interaction that can be identified with gravity, whence the characteristic size must be identified with the Planck length, $\sim 10^{-35}\,m$. This parameter is in fact the *only truly free parameter of string theory*! Thus, once this identification has been made, *everything else*, in principle, is a prediction of string theory [17]. In particular, having identified the one free parameter of string theory, the characteristic mass comes out to be the Planck mass, $\sim 10^{19}\,GeV$—a whopping seventeen orders of magnitude (100,000,000,000,000,000 times) bigger than the mass of $W^{\pm}$ and Z^0.

Depending on the complexity of the underlying theory (additional structure, the simplest of which is to have the string vibrate not in a plane but in a multi-dimensional spacetime), there will be one or several lowest-lying modes, which in a well-constructed string theory correspond to massless particles. All other modes correspond to particles the masses of which are multiples of the Planck mass. This immediately means that only the 'massless modes' of strings are identifiable as particles which may be observed at energies below $\sim 10^{19}\,GeV$ (all others are heavier and will never be detected directly).

---

[17] That is, if only our understanding of this theory were sufficient to calculate all the predictions. However, the present state of string theory is rather like Cavalieri's calculation of volumes before the invention of calculus; a patchwork at best, and without a fundamental, axiomatic framework.

The careful Reader will have noticed the emphasis on the freedom to re-coordinatize the world sheet. In fact, the so-called *reparametrization invariance* is the general requirement that physical laws and observables shall be independent of any particular choice of coordinates that may have been used to set up the framework in which to describe a given system. Under this name, it is applied to the coordinates of our (usual) 3+1-dimensional spacetime; the reparametrization invariance principle is in fact the principle of general relativity and produces the classical field theory of gravity. When applied to certain other physically unmeasurable degrees of freedom, it is known as *gauge invariance* and produces the classical field theories of electromagnetism, weak nuclear force and strong nuclear force.

Similarly, no particular choice of coordinates on the world sheet can possibly be measured, whence we *must* require *reparametrization invariance* on the world sheet. In addition to smooth changes of coordinates, there also exist discontinuous changes, which form what is called the *modular group*. These changes are generated by so-called Dehn twists, one of which is easily pictured as follows. Consider a portion of the world sheet which looks like a torus (surface of a donut). Cut this surface along one 'small circle', twist the two boundaries with respect to each other a full turn and glue them back on. Any smooth path that was drawn on this torus is again smooth, but will wind once (more than before) about the torus.

The requirements of world sheet *reparametrization invariance* and modular invariance (invariance with respect to modular transformations, such as the Dehn twists) restrict the possible string theories in a very important manner. In addition, there are other, more subtle and more technical restrictions. In particular, one such restriction is that string theories must be constructed to have a special type of symmetry called *supersymmetry*. More precisely, string theories without supersymmetry exhibit a rather subtle but deleterious type of divergence; string with supersymmetry *may* avoid this fate, and—as far as known to date—they do.

This all boils down to:

The only known theories void of any known inconsistency are superstrings.

—o—

The adventurous Reader might have long since remarked: why strings, why not membranes, or jelly-blobs, or... Suffice it here to offer again an oversimplified and heuristic answer: they wouldn't propagate. This may seem puzzling, since we are surrounded by objects with 3-dimensional spatial extendedness, and these indeed do propagate. The point is that we are looking for *elementary* extended objects. The surrounding objects are all conglomerations of many Avogadro numbers ($\sim 10^{26}$) of *particles* (which may actually be comprised of tiny, $\sim 10^{-35}\,m$, strings), and so the following argument would not apply.

Consider kicking an elementary particle: all the transferred energy goes into translating the whole particle. Now consider kicking an elementary string: not all the transferred energy goes into translating the whole string! Recall that strings have infinitely many modes, whence the transferred energy becomes distributed among the infinitely many modes and in such a way that (1) higher modes acquire less of the transferred energy (2) the sum of these partial energies converges and equals the originally transferred amount. The portion transferred to the 0-mode is then nonzero and translates the string as a whole. *Voila!* the string propagates. Consider now kicking a membrane: all the transferred energy is used up in exciting the vibrational modes of the membrane. The reason for this is that while a string possesses modes labeled by a single 'mode number', two such indices are required for the vibrational modes of a membrane. Thus, where the sum of all partial mode energies in a string was a simple sum. it is a *double* sum for a membrane 18); this produces an incurably divergent result and so whatever the distribution of the total transferred energy, none of it remains to translate the membrane.

> *–Principle XVI. That it is contrary to reason to say that there is a vacuum or space in which there is absolutely nothing.*
> René Descartes

3.2. Vacuum? Which vacuum?

The complete string theory is far from being solved (or even understood as a complete theory), so that the best one can do is parametrize the possible vacua and study them in a systematic fashion. That there might be more than one state to be called a vacuum, should not really be surprising. Recall that vacuum is not really empty in quantum theory, but is full of perpetually created and swiftly vanishing virtual particles, and is also full of antiparticles for each fermion 19) there is; these form the background against which measurements are performed.

The simplest superstring theory is the one which is set to propagate in flat spacetime. Alas! various consistency requirements then imply that this spacetime must have 9+1 dimensions!

Even if 'flat spacetime' is replaced by 'locally flat' spacetime, that is, a spacetime which is smooth although possibly curved, this critical dimension remains valid. Owing to a 75-year old idea by Theodore Kałuża, the extra six spatial dimensions may be imagined as curled up sufficiently tightly that no experiment can detect it. To an observer much bigger than the characteristic size of those six additional dimensions, spacetime will appear to have four dimensions (see Fig. 9). Whereas Kałuża, in 1919, contemplated one additional dimension and so was forced to conclude that this fifth dimension has the topology of a circle (is periodic) of a very small size, we face *six* extra dimensions which ought to be curled up somehow. Technically, by "curled up" one means "compact" and of small characteristic size. It should be obvious that there are many more 6-dimensional compact spaces, whence most of the variety in choosing the superstring vacuum.

Unlike in Kałuża's original idea, however, superstrings already have the complete bundled software of gravitation and other fundamental interactions. So, these will not

---

18) To the mathematically well-versed reader: yes, of course it is possible to convert the double summation into a single one using Cantor's diagonal trick. However, the summands still depend on the mode numbers in such a way that the summation diverges.

19) Named after the physicist Enrico Fermi, fermions are particles whose spin is a half-integral multiple of the reduced Planck constant, $\hbar$. The 'spin' of a particle or a system is the angular momentum observed by an observer with respect to whom the particle or the system as a whole does not move. The half-integrality of spin has for a consequence that fermions may be created from (enough) energy or be annihilated into energy only in pairs.

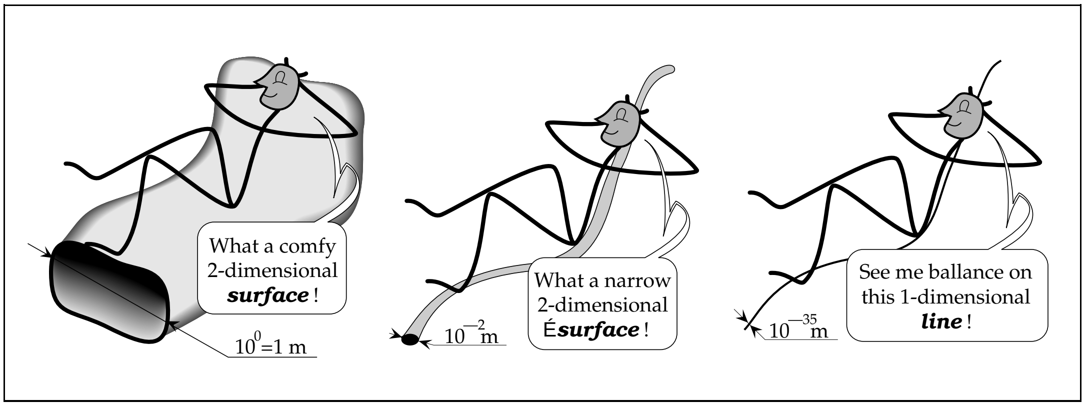

Figure 9: While the notion of dimension is mathematically an *exact* concept, in the 'real world' practice it depends on relative sizes and resolution.

be coming out of symmetries related to translations and rotations within those curled-up dimensions. However, the *geometry* of that curled-up compact 6-dimensional space turns out to affect the fundamental interactions to a certain extend, but rather more importantly, determines various parameters such as the mass of the electron, and other "matter particles", the strength of their direct, so-called Yukawa-type interactions, *etc.* In principle, this predicts the 26-odd undetermined parameters of the so-called Standard Model which has hitherto proven experimentally correct—all from geometrical properties of the curled-up 6-dimensional space.

Suffice it here to note that in a certain sense, the 6-dimensional curled-up factor of the 9+1-dimensional spacetime may be replaced with world sheet models the geometric interpretation of which eludes us. Nevertheless, the relevant calculations can sometimes be performed even to greater accuracy and much more ease than in the geometrical approach; such models of course have the same *a priori* right to be studied. Hopefully, some such "compactification model" (be it geometrical or geometry-less) will end up predicting all the fundamental features of the low energy ($\leq 100\,GeV!$) physics. The search is on.

–Mach did not believe in the existence of atoms,
on the grounds that they cannot be observed.
Wolfgang Pauli, to W. Heisenberg [10]

3.3. String experiments?

How can we perform stringy experiments? How can we see strings?

The naïve attempt of smashing probes with more and more energy until the resolution reaches Planck size inevitably fails. Not only because of the above argument, whereby details of Planck-length size are hidden by event horizons, but also because such colossal energies were available *per particle* only during the Big Bang. Surely, we are not looking forward to a mad scientist re-creating the Universe from scratch. Fortunately, the Reader need not be alarmed: even a project at a 10,000,000,000,000,000 times smaller scale has been recently axed by the U.S. Government; the extra 16 zeros do not seem to be available in any budget, any time soon.

Another avenue should be sought. Perhaps, a "waiting experiment" can be designed. For example, some grand-unified theories predict that the proton should decay, and into a pi-meson and a positron. Now, the proton mass being almost 6–7 times bigger than the mass of the pi-meson and the electron together, about 85% of its mass is transferred to surrounding matter, predominantly in the form of radiation. That is some 800 MeV of radiation per decayed proton. In living tissue, sufficient amounts of radiation are lethal, so that by looking into the mirror I can testify with confidence that the protons that might have decayed in my body (or immediate vicinity, and neglecting cosmic and round radiation) have certainly not yet reached this lethal dose (and are rather far from it, as far as my health seems to indicate). This places a lower limit on the probability of the proton decay. To improve on such an experiment, one observes a much bigger bulk of potentially reactive matter than my own body, and laces it with detectors of much finer resolution (and ethically more appropriate) than the radiation vulnerability of human body. *Voila!* an abandoned gold mine in India serves precisely such a rôle. While such experiments by design produce only upper or lower limits, their results may be combined with other experiments, bracketing the desired quantity and eventually either observing and measuring it or proving that the considered process does not happen. Notice that this type of experiments is designed to monitor processes involving virtual particles!

It is much more likely that stringiness will be detected (if at all) through such an indirect, virtual, *i.e.*, waiting experiment. However, no one has so far come up with a concrete experiment, foremost because no one has so far come up with a hallmark phenomenon for strings (as is proton decay for grand-unification).

Finally, let me mention an amusing prediction of a very general class of superstring models. So far, we have not dwelled much upon the choice of the curled-up 6-dimensional space; whatever is chosen, is taken to be the same at every point of the 3+1-dimensional spacetime. This choice is called a 'direct product' and one says that the 9+1-dimensional spacetime is a direct product of the 3+1-dimensional Minkowski spacetime and the curled-up 6-dimensional space (whatever that happens to be).

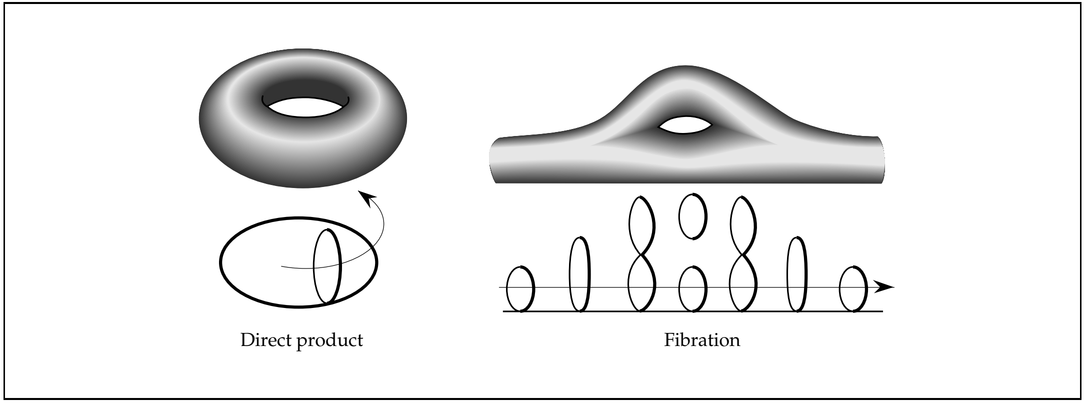

Figure 10: The torus as a direct product of two circles, and a more variable construction, called fibration.

The left hand side of Fig. 10 presents such a direct product space: take a circle and sweep it unchangingly along another one. Indeed, $T^2 = S^1 \times S^1$. Note that the torus may also be thought of as fixing several copies of the small circle, as they occur distributed along the torus, and then sweeping the big circle along the small ones, changing meanwhile the radius of the big circle so as to fit the surface of the torus. On the right hand side of Fig. 10, there is another, rather more variable construction called a fibration. As the initial circle is sweeped along the straight line, it is allowed to grow and deform, until it crosses over into a figure-8, then splits into two circles, then reattaches into a figure-8, smoothes into a single circle and finally returns to its original size and shape. (The imaginative Reader may wish to glue the two ends together into the boundary-less object, called the pretzel.) Note that the figure-8 1-dimensional manifolds are not smooth, but that the total space, depicted on the top, is. The “horizontal” space in a fibration is called the base space and the “vertical” space is called the fibre. As should be clear from this example, the an isolated sub-collection of the fibres is allowed to become singular in a fibration; whether or not the total space also singularizes is not restricted.

Returning to the shape of the 9+1-dimensional spacetime in superstring models, an obvious possibility is to let the curled-up 6-dimensional space vary along the 3+1-dimensional spacetime. Amusingly, even if one requires that this happen in an analytic fashion, the curled-up 6-dimensional fibre is forced to singularize over certain points in the 3+1-dimensional spacetime, somewhat like the circle becoming the figure-8 in Fig. 10. The points in 3+1-dimensional spacetime where this happens form a 1+1-dimensional surface and so look as strings at any given point in time. These strings stretch throughout the Universe and curve the local spacetime; they accrete matter and so may cause the galaxies and meta-galaxies order into filamentary fashion. Eventually, through interaction with matter and radiative decay, these cosmic strings decay and leave just the filamentary structure on the galactic and meta-galactic scales. Combined with several other effects (quite a few of them again deriving from the physics of ultra-small), this indeed paints a rather correct picture of stellar distribution.

> –*Mathematics may be defined as the subject in which*
> *we never know what we are talking about,*
> *nor whether what we are saying is true.*
> Bertrand Russell

Appendix A. Nitpicking

Mathematical concepts lack ‘reality’, in the sense that one cannot whack one’s opponent with a *Klein bottle*, or with a *derivative*. However, a physical model based on these mathematical models can be very real: a well disguised (3-dimensional and so self-intersecting immersion of the) Klein bottle provides for a magician’s jug into which liquid can be poured, but won’t come out (and then, with a twist of the wrist, pours effortlessly); the derivative, with respect to time, of distance traveled—is speed.

Now, broadly, mathematics may be thought of as a language, and involves

1. a (growing) set of *words*, of which a relatively small number are indefinable (elementary) concepts;
2. syntactic rules, by which to form *sentences*, *i.e.*, statements;
3. rules of logical deduction, by which to relate statements as one being the logical consequence of another 20).

---

20) The use of singular here is not a mistake! For, if a statement was a logical consequence of several other statements, these can just as easily be concatenated into a single one.

Since this structure is being used in scientific models, mathematics may be regarded as the language of science. However, this view is incomplete, at least because it omits the very important impact of mathematics on the scientific model: the rigor of mathematical deduction and framework is used to determine not only which statements make sense but also which follow from the experimentally known statements. That is, the rôle of mathematics in a scientific model may also be likened to that of a 'sausage machine' into which experimental observations are translated, and out of which experimental predictions are being deduced with confidence and certainty. So, simultaneously a language and a 'sausage machine', mathematics must be more than either of these—especially in view of the obvious fact that humans are quite capable of doing 'pure', or perhaps better named *unapplied* mathematics. For our present purposes, however, (scientifically applied) mathematics may be thought of as a language and a 'sausage machine', simultaneously.

It is straightforward that the concept of 'consequence' enables an 'ordering' of mathematical statements. Those that are not consequences are called *axioms* or *postulates*, while all others are *theorems*. However, this ordering is not unique: sometimes, the relation of 'logical consequence' works both ways. Sometimes, not only is statement A a logical consequence of statement B, but also the other way around; A and B are said to be equivalent. We are then free to choose whether A shall be called a postulate and B a theorem or the other way around. Usually, one strives to reorganize this hierarchy until the number of (independent) postulates is minimized. Notice that the notion of 'logical consequence' is in principle different from the 'cause-and-effect' relation (which is unambiguously unidirectional), although they may coincide at times.

Given a handful of postulates and the powers of logical deduction, there are infinitely many theorems which can be derived from them. Clearly either a statement or its negative is taken to be the derived theorem. If somehow both a statement and its negative would be logically deducible from a given set of postulates, then this set of postulates is *self-contradicting* and one or more of these postulates must be dropped. When this is applied in a scientific model, experimental observations are translated into mathematical statements—*which must be retained as true.* Keeping these statements all the time, other statements are being derived as predictions (and then tested), but also a small set of postulates is sought such that all the known (true) theorems would follow therefrom. These postulates will then become the foundation of the model, the theory, or the whole science.

Clearly, since new experiments are being performed, there is always the possibility that a true statement will be discovered which cannot be derived from the hitherto accepted postulates; the new statement will be neither provable nor disprovable from the hitherto established model or theory. As an example of this, recall that electromagnetic phenomena *cannot* be derived from (reduced to) mechanics; they involve genuinely new concepts, ideas and... well, postulates. As a counter-example, heat and heat transfer phenomena *can* be derived from mechanics, using the mathematical methods known as statistics.

So, one naturally wonders if there is a 'complete set of postulates' to describe Nature. That is, if all the (infinitely many!) true statements about Nature (theorems) can be derived from a finite set of postulates. It may be that humans will never uncover all of them, but it may nevertheless be good to know if this smallest possible list of absolutely necessary assumptions is finite. Fortunately for the natural inquisitiveness of the human kind, the answer is negative: Kurt Gödel's "Incompleteness Theorem" states that every mathematical system of a sufficient sophistication is also incomplete [9]. That is, in such systems, there is always a statement which can be made, and neither this statement nor its negative can be derived from the given postulates. This incompleteness is not a flaw of human abilities or lack of time; it is an essential and inherent characteristic of mathematics.

Besides defining ‘sufficient sophistication’ precisely (which is way beyond the scope and level of this articlet), Gödel also notes that already arithmetic is ‘sufficiently sophisticated’ and so incomplete. Since the mathematics which is needed in science certainly uses arithmetics (and considerably more), it follows that science cannot escape this essential incompleteness. One will always be able to write down a statement which cannot possibly follow from the assumed postulates, and the negative of which is likewise undecidable. Mathematicians at that point have a choice, to include the statement or its negative as a new axiom. Physicists at that point translate the statement and its negative into a prediction and its negative and check which one is true in Nature (assuming that the relevant experiment can be done).

To provide an example, consider the early history of the theory of β-decay (wherein a neutron was observed to decay into a proton and an electron), which stood baffled at the experimental fact that the total energy of the neutron (before decay) was more than the total energy of the proton-electron system (after decay). Some physicists, together with Niels Bohr, were ready to abandon the law of conservation of energy as a completely universal law (deferring a specification of the circumstances in which it could be violated), while Wolfgang Pauli noted another the (implicit) assumption: that all particles in the decay were observed. Flipping this assumption, Pauli postulated the existence of a new particle, the neutrino. So, the neutron was decaying into a proton, an electron and a neutrino (today renamed into anti-neutrino, but that is merely a more convenient convention): $n^0 \to p^+ + e^- + \bar{\nu}$, and the (anti)neutrino was carrying the missing energy. From the fact that it was not observed and assuming electric charge conservation, the $\bar{\nu}$ had to be electrically neutral and a limit on its mass could be derived. Postulating the $\bar{\nu}$, Pauli had restored (and enriched) physics and made its models once more conform with Nature. Except that the neutrino was experimentally detected only 20 year later.

Lest the alert Reader accuse me of falsifying history, I hasten to point out that the actual turn of events is not an example of stumbling upon a statement not decidable within the given set of axioms and then checking whether it is true or false in Nature. Rather, the statement was made by Nature and observed by the physicists. However, the statement *could have easily been made* and, given previous experience with conservation of energy, would have probably run something like “Total energy is conserved in the decay $n^0 \to p^+ + e^-$”. In the early 1930’s, everybody who understood these words would have agreed that this makes sense but does not follow from any physics known up to that point, although is consistent with all of it. Its negative would have had equal chance of being true in the real world, since strictly speaking, conservation of energy was not proved to be upheld by these than new and unknown processes. This then would be an example of a Gödelian undecidable statement. Experiments show that the statement in the quotation marks is false (as stated) in Nature, and its negation would have to be accepted as a true statement and a new postulate. Pauli “merely” identified the implicit assumption which to flip instead of a conservation law [21].

To a certain extent, Einstein’s unification of mechanic and electromagnetic phenomena is also an example to this: Newton’s laws are invariant under Galilean transformations of space-time coordinates, while Maxwell’s equations respect Lorentz transformations. Now, either the two types of phenomena exhibit the same sort of invariance, or they don’t; it matters little that the former choice seems aesthetically more pleasing: beauty is in the eyes of the beholder. Choosing the former postulate—that mechanic and electromagnetic phenomena are invariant with respect to the same transformations of space-time—led Einstein to modify Newtonian mechanics. This led to new and exciting (well, downright hilarious) predictions. Had it turned out that atomic clocks are not slowed down when set into motion and then returned, the deductions would have to be backtracked and the postulate not otherwise tested would have to be dropped or changed. This backtracking is not at all a simple process as the Reader might conclude from this admittedly simplifying description; there can easily be implicit and untested assumptions, and also at times the reasoning can be subtly (or less so) flawed. While mathematical rigor is perfect, humans who practice (applied or unapplied) mathematics are not.

Finally, two remarks are in order. First, let me note that Gödelian undecided statements are expected to be rather more complicated than the two examples given above [22], and can easily be fantastically unpractical. In addition, Gödelian undecidable statements are somewhat like prime numbers: infinite in number and with no algorithm for listing them. So, given the fact that persistent experimentation already produces an abundant wealth of experimental data, few if any new postulates are likely to be discovered by hunting for Gödelian undecidable statements and then devising experimental tests to see which way the Nature votes. This method would be both unwieldy and rather akin to random jumping—even if there existed a systematic algorithm of listing Gödelian statements, that algorithm would probably have nothing to do with the subject matter, the study of Nature.

Thus we are back to the friendly coexistence and communication between experimentalists and theorists and the many shades of physicists in between, a *modus operandi* which has so far proven quite a prolific and fertile arrangement. It is however well worth it to understand both the powers and the limitations of the framework and the *milieu* wherein scientific models are developed and studied.

> *To the problem of imperfect knowledge [Science] suggests a new and unprecedented solution—honest work.* David Brin [3]

Acknowledgments: I am indebted to the ΦBK for inviting me to speak at the Nicolaus Copernicus workshop, May 1994 at Howard University, Washington DC; mine was the second part of a joint lecture “What is the Universe Made Of?” with Prof. C. Kumar, who covered the astronomy aspects. This work was supported by the DOE grant DE-FG02-94ER-40854.

---

[21] Besides conservation of energy, the process $n^0 \rightarrow p^+ + e^-$ would also violate the conservation of angular momentum, as well as several other less well known conservation laws.

[22] The Reader should note that the simplicity of the two examples is deceiving: both of them include the unwritten physics background on which they are based!

Further Reading and References

[1] E.A. Abbott: *Flatland* (Dover, New York, 1992; orig. 1884).

[2] J.L. Borhes: *A New Refutation of Time*, in *A Personal Anthology* (Grove Press, Inc., N.Y., 1967).

[3] D. Brin: *The Commonwealth of Wonder*, in *Otherness* (Bantam Books, New York, 1994).

[4] D.B. Cline: *Low-Energy Ways to Observe High-Energy Phenomena Sci. Am.* **271** *(9/94)40–47.*

[5] S. Dimopoulos, S.A. Raby and F. Wilczek: *Unification of Couplings Phys. Today* **44** *(10/91)25–33.*

[6] G. Gamov: *One Two Three... Infinity* (Bantam, New York, 1971; orig. 1947).

[7] G. Gamov: *Mr. Tompkins in Wonderland* and *Mr. Tompkins Explores the Atom* (Cambridge University Press, New York, 1965; orig. 1940 and 1945, resp.).

[8] H. Georgi: *Flavor* $SU(3)$ *Symmetries in Particle Physics Phys. Today* **41**(4/88)29–37.

[9] K. Gödel: *On Formally Undecidable Propositions of Principia Mathematica and Related Systems* (Dover, New York, 1992; orig. 1962).

[10] W. Heisenberg: *Physics and Beyond* (Harper and Row, New York, 1972).

[11] P. Langacker and A.K. Mann: *The Unification of Electromagnetism with the Weak Force Phys. Today* **42**(12/89)22–31.

[12] H. Pagels: *The Cosmic Code* (Bantam, New York, 1983; orig. 1982).

[13] H. Pagels: *Perfect Symmetry* (Bantam, New York, 1986; orig. 1985).

[14] W.H. Press and D.N. Spergel: *Cosmic Strings: Topological Fossils of the Hot Big Bang Phys. Today* **42**(3/89)29–35.

[15] B. Russell: *The ABC of Relativity* (Signet, New York, 1969; orig. 1925).

[16] B. Sadoulet, J.W.Cronin: *Particle Astrophysics Phys. Today* **44**(4/91)53–57.

[17] E. Schrödinger: *What is Life?* with *Mind and Matter* (Cambridge University Press, Cambridge, 1992; orig. 1944 and 1958, resp.).

[18] S. Weinberg: *The First Three Minutes* (Basic Books Inc., Publishers, New York, 1977).

[19] S. Weinberg: *Dreams of a Final Theory* (Random House, Inc., New York, 1993; orig. 1993).

[20] D. Wolff: *Chinese for Beginners* (Barnes and Noble Books, New York, 1985; orig. 1974).

[21] F.D. Peat: *Superstrings and the Search for the Theory of Everything* (Contemporary Books, Chicago, 1988)