1 The nature of observing Nature

1.1 Fundamental physics as a natural science

The ultimate aim of this course is to present the contemporary attempt to perceive the (fundamental) structure and nature of Nature. First, however, we must examine the (methodo-)logical framework at the foundation of this aim.

1.1.1 Not infrequently, things are not as they seem

Although an erudite historian will certainly and readily cite earlier quotations of the thought expressed in the title of this section, I should like to introduce this *leitmotif* as a Copernican legacy. The readiness to abandon the "obvious," "generally accepted" and "common sense" for unusual insights – those we can actually check – is certainly an essential element. This motif permeates the development of our understanding of Nature, and reappears in its contemporary form as *duality* [☞ Section 11.4].

Of course, not just any unusual insight will do: a *luni*centric or an *iovi*centric system, for example, would offer no advantage over the *geo*centric cosmological system. Most significantly, heliocentricity simplifies both the conceptual structure and the practical application of the planetary system, and makes it more uniform. Although still assuming circular orbits and so in need of corrections,[1] Copernicus' model is *essentially* simpler; maybe this could be regarded as a variant of Ockham's principle.

This idea is not yet Newton's universal law of gravity, but already contains its germ, its unifying motif: all planets follow the same type of regular motion and only appear to wander randomly (as their original Greek name implies). Also, the ultimate test of this model is easily identifiable: the positions and the motions of the planets determined by (the simpler) computations within the heliocentric system agree with astronomical observations.

Examples of this leitmotif begin at such a simple level that they are rarely noticed:

1. The shadow of an object is often distorted and many times larger than the object itself. Nevertheless, only very little children are afraid of the shadow of a wolf or a monster, however aptly conjured by the artists in a puppet theater.

[1] Only after Kepler's *ad hoc* postulate of elliptical orbits (which Newton explained *a posteriori*) did heliocentricity achieve its really convincing technical simplicity and precision.

https://doi.org/10.1017/9781009291507.003 Published online by Cambridge University Press

2. Viewed from a large plateau (without mountains on the horizon), the Earth does look flat. Yet, Eratosthenes (*c*. 276–195 BC) not only proved that the Earth was round, but even computed its size (to about 10–15% of the modern-day value!). This computation was based on the length of the summer solstice noon shadows in Syene (a.k.a. Aswan) and in Alexandria, the distance between these cities, and using geometry that is two millennia later regarded as elementary. In time, Eratosthenes' results and reasoning became "politically inconvenient," were suppressed and forgotten for some sixteen centuries, and were rediscovered in the West only centuries later, in the Renaissance. Although by now few people doubt that the Earth is round, when (if?) humankind expands into Space, the once obvious flatness of the Earth will become unthinkable; just as once its roundness was.

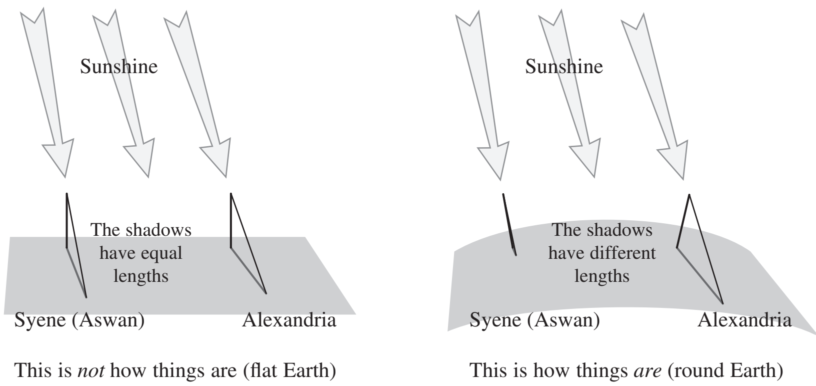

Figure 1.1 Eratosthenes' analysis which, by means of measuring angles and distances, gives (depending on the precise value of ancient units he would have used) the size of our planet Earth to about 16% at worst and 2% at best! (The shadows in the illustration are exaggerated.)

3. Everyday experience convinces us: the Sun and the Moon revolve around the Earth. This was indeed known to the ancient Greek science, as reported in Claudius Ptolemy's (*c*. AD 90–165) *Almagest*. This suppressed the teachings of Aristarchus (*c*. 310–230 BC), who not only advocated the heliocentric system, but also estimated that the Sun is about 20 times further away from Earth than the Moon and about 20 times bigger.[2] It took sixteen centuries for the West to rediscover this.
4. To the naked eye, our blood seems homogeneous and continuous. So it was believed to be until 1683, when the Royal Society published the first detailed pictures of red blood cells, as seen through a microscope and drawn by Antoni van Leeuwenhoek (1632–1723). In 1932, Ernst August Friedrich Ruska (1906–88) designed the first electronic microscope, the modern versions of which permit us to see – in the most direct way possible – individual molecules and even atoms, of which all matter around and within us is composed.

This last insight (quite literally!) is due to technical development, and it fully convinces us of the finite divisibility of *things* around us. Seemingly continuous things: fluids, air, metals… in fact consist of an enormous number of teensy particles! Whence stems the conviction that there exist "elementary particles" – the smallest building blocks of which everything else consists. Although

[2] The 20-fold error in Aristarchus' result stems from insufficient precision in angular measurements of the time; his reasoning and geometry were essentially correct. Also, the ratio of the diameters of the Sun and the Moon indeed does equal the corresponding ratio of their average distances from the Earth, but is ≈ 400, not 20.

https://doi.org/10.1017/9781009291507.003 Published online by Cambridge University Press

this idea is fantastically successful in explaining Nature and even predicting its behavior, it behooves us to keep in mind that the "particulate nature" of Nature mirrors our gradually improving understanding of Nature, and that this insight is subject to verification and periodic audits.

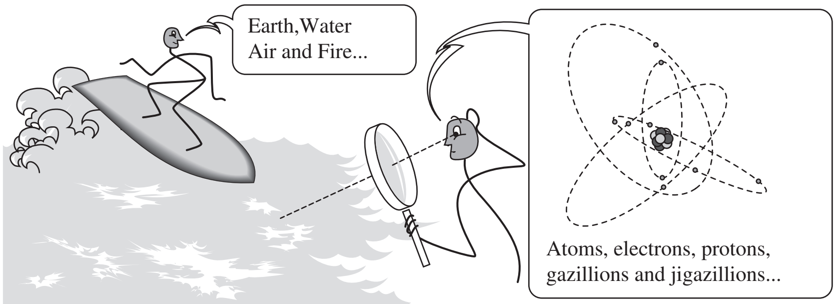

Figure 1.2 What at humanly characteristic scales seems smooth, homogeneous and continuous, may well look completely different under sufficiently closer scrutiny.

The Reader will certainly have no difficulty extending this list with many other and possibly more interesting and amusing examples, evidencing our basic leitmotif. Standard human perception, so well adapted to our daily routine, does not serve us well when concerning scales and proportions that are not as commonplace. From the typical, everyday vantage point and at characteristic human scales, planetary and stellar events appear warped. We must apply our (patiently educated and disciplined) mind to correct this picture. Indeed, once so educated, the Sun in the sky never again seems the same! In our mind's eye, we can actually *see* the Earth upon which we stand, as it rotates around the star we call the Sun. Similarly, once educated about the blood cells, our mind's eye has no difficulty *seeing* the erythrocytes as they stream through the blood plasma in our veins, and the leukocytes as they attack the blood-borne bacterial invaders.

Yesterday's unbelievable and ridiculed "nonsense" (that diseases are caused by germs too tiny to be visible was indeed widely ridiculed) may well turn out to become an evident truth of today – and such realizations turn out well remembered. So-called "evident" truths must not be exempt from verification just because they are considered evident: not infrequently, "evident" is simply that which is familiar and what are we used to. Not yet having doubted something is no guarantee of its truth.

However, we must then inquire which claims should we doubt and how do we establish the truth of any particular claim if everything is to be doubted? Following Descartes' rationale, *everything* that may be doubted without self-contradiction should be doubted. However, physicists are usually more pragmatic than that.[3] With a nod to the principle "if it ain't broke, don't fix it," physics models and theories are doubted and re-examined when they start predicting things that *are not*, or fail to predict the things that *are*. . . And, predictions are derived from a model as much as technically and practically possible.

In fact, it is our *duty* to "churn out" everything one possibly can from every scientific model. This is both for the sake of economy (the predictions of a model are its "products") and in order to establish if the model is in as full an agreement with Nature as it is possible to determine at any given time.

[3] . . . and even without the persnickety conclusion that Descartes' motto *cogito, ergo sum* leads into solipsism, or recalling Hume's demonstration just how destructive such infinitely regressive doubting may be. . .

https://doi.org/10.1017/9781009291507.003 Published online by Cambridge University Press

1.1.2 The black box: a template of learning

To formalize our approach, let us picture the scrutinized system as a *black box*, representing the lack of knowledge about its contents. What follows may then be regarded as the three pillars of (exact, natural) science.

I. To learn something of the contents of the box, an *input* (controlled or otherwise known) is directed at the box, and we observe the *outcome*. The input may be something as simple as knocking, shaking, or maybe something more technical, such as X-rays or ultrasound. The outcome is whatever *emerges* from the box in response. For example, as the box is shaken, its weight might move in a way suggesting that it is concentrated in several distinct subsystems inside the box. Or, the box may ring hollow to knocking. Or, X-rays may show the image of Thumbelina's skeleton. . .

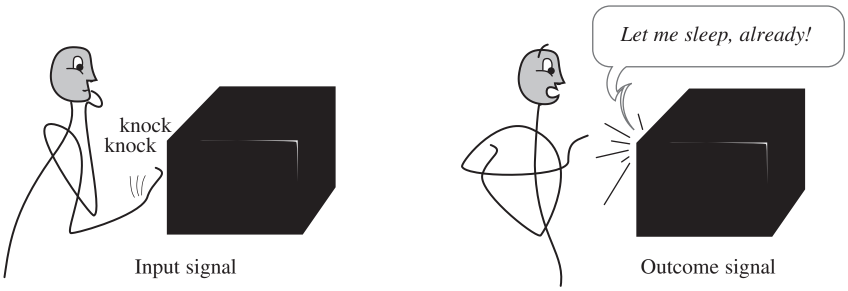

Figure 1.3 The black box experiment template.

II. Using the information about the box in the form of a "response to the input," where both input and outcome are adequately *quantified*, we develop a *mathematical model* that faithfully reproduces all received outcome signals as a response to the corresponding input signals. Needless to say, both input and outcome signals must be measured, and will therefore be known only up to measurement errors. This defines the resolution/precision/tolerance of the model. Of course, a resolution of the mathematical model cannot be guaranteed to be better than this; and this must then be understood as the resolution of the model as a whole.

III. This mathematical model is then used to derive the consequences of the conceptual model: One computes the response of the system (as represented by our model, in the role of the black box) to new, as yet untested input signals. These responses then need to be tested, if and when that becomes possible.

Herein then lies the clue as to "what and when to doubt" and "how to test truthfulness." Physics (and, more generally, scientific) models must be re-verified, wherein one or more of the "ingredients" are doubted and perhaps even replaced, if the model does not reproduce and correctly correspond to Nature to within the resolution of the model [☞ also Comment 10.5 on p. 388]. This shows that:

Conclusion 1.1 *Exact science always errs, but is exact about how much.*

Comment 1.1 *"Physics students learn this very quickly, through a shock, when they proudly obtain the required results of the first lab exercise, and the teaching assistants quiz them about the errors at least as much as about the obtained results." D. Kapor*

https://doi.org/10.1017/9781009291507.003 Published online by Cambridge University Press

This three-step process, "observe–model[4]–predict," repeats iteratively and infinitely, in counterpoint to the above-cited leitmotif, and guaranteed by Gödel's incompleteness theorem,[5] since the research subject is sufficiently complex and is not easily exhaustible (unlike, e.g., "tic-tac-toe," which is exhaustible as a game of strategy) [211]; see the lexicon entry in Appendix B.1, as well as Appendix B.3. When the model is constructed, the predictions of the model are derived and checked experimentally, as well as possible in practice. As human ingenuity incessantly improves the technology, and new techniques and methods (both experimental and theoretical) are being continually developed, new predictions are being continually derived and checked with an increasing precision. Sooner or later, these new checks (both experimental and theoretical) indicate the shortcomings and uncover statements that can be neither proven nor disproven within the given theoretical system.[6] If such a new statement can be experimentally checked as true or false, the model needs to be extended so as to include this new fact about Nature. When the so-extended model successfully reproduces all (known) "new" facts, additional predictions of the now extended model are derived and checked, and these typically indicate further directions of extension and improvement, upon which yet more additional predictions may be derived, and so on.

Comment 1.2 *To illustrate, the phenomena we now label as* ***electrodynamics*** *are describable by equations that are easily written down within the theoretical system of classical mechanics of particles and fields, but can be neither proven nor disproven within this system. The Maxwell equations (5.72) and the electrodynamics laws that they represent, provide* ***new axioms*** *to the theoretical system of the classical mechanics of particles and fields applied to charged particles and electromagnetic fields. In turn, Section 5.1 shows these equations to follow from the gauge principle, which therefore is the one (overarching)* ***new axiom****; see also Appendix B.3.*

1.1.3 Philosophers are not scientists

A second glance at this framework of thought reveals something extraordinary! The scientific models[7] described here, and systems of such models forming theories and theoretical systems, are improved and extended, but not *literally falsified*, i.e., proven to be unconditionally false! (For the most part, it is rather our mental imagery and philosophical "underpinnings" of the scientific model that are taken too seriously, and may have to be abandoned as false.) Radical revisions of course do occur in scientific research – and not so infrequently – but that does not *falsify* established models and theories, only perhaps an unwarranted trust that those models and theories would be exact and absolute truths. Properly understood within their qualifications, models and theories of fundamental physics have not been falsified throughout the past three centuries, but have been and continue to be refined, extended and often united.

Reasons for this are found in comparing scientific models with earlier efforts and doctrines. Scientific models unify the inspiration of (experimental) induction with the rigor, self-consistency and persistence of (rigorous mathematical and logical) deduction.

[4] In this context, the verb "to model" encompasses the creation of the mathematical model that describes the scrutinized phenomenon, and that can be summarized into an applicable formula. Whence stems the *law* for the system wherein the phenomenon is observed, and "to model" then includes "to introduce as a law of Nature." However, this is not an absolute and inviolable law by decree, but one that is subject to verifications in comparisons with Nature, and adaptations to this one and ultimate arbiter.

[5] . . . barring the dismal logical possibility of the scientific spirit dying out or becoming exterminated. . .

[6] These are essentially undecideable statements; see the lexicon entry on Gödel's incompleteness theorem, in Appendix B.1, and Appendix B.3 in particular.

[7] A *scientific model* includes the mathematical model together with its concrete interpretation: formulae, algorithms, programs, together with their physical meaning, i.e., a dictionary between the symbols of the mathematical model and the corresponding quantities in Nature. In this sense, a "model" then also implies a "law" – in the sense of Newton's, Ampère's or Gauss's law, not in the sense of a decree of some legislative body. The notion of "natural law" is thus integrally woven into the scientific modeling of Nature, far from it having been abandoned, as sometimes opined [533].

This complementary combination of *quantitative measurements* and *mathematical modeling* is often attributed to the revolution in the philosophical approach in studying Nature, and is most often linked to Galileo and Newton. However, Eratosthenes' and Aristarchus' above-cited planetological results were clearly based on this same combination of methods. This idea is therefore over two millennia old. Suppressed through most of the past two millennia, this same combination of measurements and mathematics was methodically and consistently revived by Galileo, Newton and their followers. With the development of mathematics – and especially of calculus, invented for that purpose by Newton, Leibniz and contemporaries[8] – physics engaged into *warp* drive (the superluminal propulsion from the sci-fi series *Star Trek*).

Roughly, measurements translate quantities describing observed natural phenomena into corresponding quantities in a mathematical model. This model is then used as a faithful (as best as known) representative and replacement of the natural phenomenon. It is also a persistently rigorous tool for deductive predictions about that natural phenomenon. Those predictions are then checked in turn, the model adapted, corrected and improved, if and when the predictions turn out to differ from what is observed in Nature.

Thus, Einstein's theory of relativity does not *falsify* Newton's mechanics but *extends* it: When all relative speeds in a system are much less than the speed of light in vacuum, relativistic corrections to Newton's mechanics are negligible and Newton's mechanics yields a perfectly usable model of reality. If some of the relative speeds increase, the corresponding corrections become relevant, Newton's mechanics is no longer a good enough approximation (the errors, about which physics always must be precise, become unacceptably large), and we must use the relativistic formulae. In turn, Einstein's relativistic physics cannot be claimed to be absolutely true/exact either, but merely that it is more accurate than Newton's. After all, we already know that quantum physics may well force us to revise the structure (and perhaps even the nature) of spacetime itself when approaching Planck-length scales. Science can only make qualified statements, the "truth" of which will always depend on precision (resolution) – and which continues to improve in ways that no one can foresee.

Insisting on the iteration of this precision-sensitive "observe–model–predict" cycle immediately discards "theories" such as the one about *phlogiston*, the supposed intangible substance of heat. That "theory" neither explained nor predicted quantitative data, and may be called a "theory" only in common, non-technical parlance. A similar fate befell the so-called "plum pudding" model of the atom, which explained and predicted very little (and incorrectly), and which its Author humbly called a "model" worth exploring, and mercilessly abandoning if found faulty; which it was – both faulty and abandoned.

It is absolutely crucial that what we intend to call a *scientific* theory must be subject to verification through comparison with Nature, at least in principle. This implies that a theory must be *quantitative*, i.e., a theory must explain and predict experimental data, which can be checked. Quantitative predictions may be as simple as "yes/no" results; whether one predicts a single bit of information or an entire googolplex[9] of them – predictions must contain *new* information.

A word of warning: "subject to testing" does not mean that we can simply call up the local lab, order some results, and expect a twenty-minute delivery. Nor does it mean that even a planetary budget could fund the required experiment (not that there will be a planetary budget any time soon). Nor does it mean that anyone has even the faintest hint of an idea for a concrete experiment, even with a pan-galactic budget and a post-*Star Trek* technology. However, the theory must be

[8] It has recently been discovered that Archimedes knew about the concepts of limit and the principle of exhaustion [382], but that this knowledge has been neglected and forgotten for the better part of two millennia.

[9] *Googol* (which must not be confused with *Google*) is the number 10^{100}; *googolplex* is the number $10^{10^{100}}$. For comparison, there are only about $\mathbf{N} := 10^{80} \ll 10^{100}$ particles in the universe, but the number of all their k-fold relations is immensely larger than googol, $\sum_k \binom{\mathbf{N}}{k} = 2^{\mathbf{N}} \ggg 10^{100}$, and the number of all relationships between those relationships (as a second-order estimate of complexity) is much larger than googolplex, $2^{2^{\mathbf{N}}} \ggg 10^{10^{100}}$.

"subject to testing, in principle": thought experiments may be envisioned rigorously, and their execution is obstructed by neither political economy nor practical "minutiae" such as magnetizing a mountain-size apparatus. Of course, the models that may be tested may be either demonstrated as *tentatively* established,[10] or discarded if they can be shown to disagree with Nature.

It cannot be over-emphasized (see, however, also Digression 1.1 below, as well as Sections 8.3.1 and 11.1.4 and Appendix B.3):

Conclusion 1.2 *Models that can (in principle) be refuted are scientific.*

Interestingly, a verb (in Chinese) is, by definition, a word that can be negated [578]. However, the correct application of this criterion, so simply stated, supposes a detailed understanding of the structure of scientific systems, to which we return in Section 8.3.

Digression 1.1 The principle of Conclusion 1.2 reminds us of the principle of falsifiability, popularized by Karl Popper [443, 444]. Intending to describe the historical process of the evolution of science, he concluded that experiments about atoms *falsify* classical physics, which is then *substituted* by quantum physics since that successfully describes atoms. So understood, the *principle of falsifiability* harbors at least two equivocations: (**1**) the naive version equates it with the related "testability" and presupposes direct and unequivocal experimental testing, and (**2**) equivocation in categories. Both equivocations are dangerous to the socio-political status of science. Also, the tacit assumption that all statements of a model are necessarily either confirmable or falsifiable, which simply need not be the case [☞ the lexicon entry on Gödel's incompleteness theorem, in Appendix B.1, and also Appendix B.3].

The first equivocation is based on a restriction of physics as a science to a "directly empirical" science, whereby a theory that we cannot experimentally test is being denied its "scientificity." However, there exist (in the scientific and the sci-fi literature and media) effects that contradict no known science, but for the experimental testing of which [☞ also Refs. [171, 505]]:

1. the resources are too expensive (e.g., a synchrotron around the Earth or around the Sun and Proxima Centauri, not to dream of a tokamak from here to Andromeda),
2. the requisite procedures are prohibited by moral or ethical reasons (e.g., cloning, bionic, and certain educational, behavioral and nutritional experimentation),
3. the resources require an as yet unknown technology (e.g., painting the ceiling of a room with neutronium would cancel gravity in the room – if "only" we knew how to produce neutronium paint and how to paint the ceiling without it caving in),
4. a new concept and/or methodology is needed (e.g., for a direct measurement of an *upper* limit of the proton's lifetime).

It is already intuitively clear that not one of these obstructions for experimental testing should take away from the "scientificity" of a theory. And, even simpler, it is clear that experiments with stars, positions of the constellations and the development of our own universe cannot be performed at will, nor is setting up an experimental control group

[10] Being forever subject to future and additional testing, "established" can in this context only ever be understood as *tentative*; this is a "small" detail that is rarely stated explicitly, but must always be understood.

https://doi.org/10.1017/9781009291507.003 Published online by Cambridge University Press

possible even in principle! Nevertheless, it is just as clear that astrophysics, astronomy and cosmology are no less "scientific" for this.

— ❦ —

The second possible equivocation is more subtle, and so also more dangerous. Also, it has at least two aspects. On one hand, there is the danger of confusing the category to which a certain *theoretical structure* belongs. For example, "classical physics" is not a particular model with particular predictions that may be experimentally tested, but a scientific system of assumptions (axioms) and procedures of derivation; this then may be applied to concrete phenomena, such as a pendulum, a bob on a spring, or the atom. The incorrectness of any one concrete model – as in the case of the classical model of the atom (see however Footnote *11* on p. 310 as well as example B.2) – may imply an error in the *application* of classical mechanics or in classical mechanics *itself*, or perhaps even elsewhere in the underlying complete chain of reasoning. We must explore precisely which of the assumptions lead to the observed disagreement with Nature. In fact, the application itself may turn out to harbor an error for various reasons, from a minor technicality to a fundamental inappropriateness. That, after all, is the usual advisory about all proofs by contradiction. However, it would evidently be silly to deny the "scientificity" of classical physics as a whole because of its inability to model the atom.

On the other hand, the very idea that a scientific theory *falsifies* another is a dangerous equivocation. Both in common parlance and in legal practice, the verb "to falsify" implies that the statement being falsified is being shown to be a falsehood. This, in turn, implies the tacit expectation of a binary true/false value. However, it is – or should be – very well known that the relation between quantum and classical physics is *continuous* and depends on the context and "resolution." For any process under scrutiny, we must compute the ratio of $\hbar$ with all characteristic actions and all other commensurate physical quantities.[11] If each of these ratios is *sufficiently smaller* than 1, the numerical errors in the results computed using classical physics are *negligible*. It is evident that "sufficiently small" here implies a finite and an a-priori established tolerance. Therefore, the answer to questions such as "is classical physics applicable even to a single particular event?" essentially depends on at least one *continuous* parameter, and the answer cannot possibly be an unconditional "yes/no." Classical physics is therefore extended/generalized and not *falsified* by quantum physics. The situations with relativity, field theory, and even superstring theory are analogous.

Generally, physicists understand that quantum physics does not simply *falsify* classical physics, but *extends* it into a domain where classical physics is *not sufficiently precise*. Unfortunately, philosophers of science are not physicists. This pragmatic approach should be compared with a similar vantage point of philosophers of natural sciences such as Thomas S. Kuhn [323], where one needs to know that Kuhn obtained his BS (1943), MS (1946) and PhD (1949) degrees in *physics* at Harvard, where he lectured on history of physics 1948–56. However, Kuhn opines that theories (and paradigms) are chosen by the group of researchers that is more successful than others, and assigns this choice a degree of socio-politically pliable subjectivity. This seems all too alien to most physicists I know,

[11] In elementary particle physics one uses so-called natural units, based on the natural constants $\hbar$ and c, whereupon these are not written explicitly, and formally one says that "$\hbar = 1 = c$." This practice may well be used in any complete unit system: once in agreement to use SI units, "length of 10" may only mean "10 m," "force of 5" may only mean "5 N=5 kg m/s^2," etc. However, as the purpose of this book is to *introduce* the Reader into this practice, factors of $\hbar$ and c are herein written explicitly, but in gray ink.

https://doi.org/10.1017/9781009291507.003 Published online by Cambridge University Press

and which I myself (perhaps all too naively) cannot accept for physics itself, nor any other science, but only and at most for the admittedly capricious socio-historical process of development in a particular subfield.

Suffice it here then to just assert without a historically and statistically justified argument – as a manifesto, if need be: Theories and theoretical systems in physics are chosen by Nature itself, through our long and patient communication with it (in the sense of the caricature in Figure 1.3 on p. 6). Albeit extremely challenging and difficult at times, this is always well worth the effort and ardor.

Of course, it is logically impossible for a science to be exact without being quantitative. That is, "exactness" must be accepted as the requirement that it must be possible to develop a system of questions that can be answered by precise yes/no answers. Subsequently, these answers (easily written as a binary number) may quantitatively characterize the events to be modeled – and to be predicted. If these yes/no answers follow a statistical (probabilistic) distribution, this is only a *technical* complication and does not take away from the "exactness" in this sense.[12] This is always true of all branches of physics. While statistical physics and quantum mechanics are probabilistic, this only complicates the techniques and dictates the style of research. In fact, many fine (mathematical) techniques of statistics specify precisely which questions are meaningful to ask, which are meaningless, and which among the former ones have a definitive answer, which "only" a probabilistic one. For example, the temperature (as the average kinetic energy) of a fluid may be predicted precisely, but the kinetic energy of a single molecule is subject to fluctuations; the distribution of these fluctuations is predictable precisely, but not their individual values in practice, owing to the too large number of contributing factors, such as the repeated collisions with 10^{26}'s of molecules. The kinetic energy of a single molecule may in practice then only be known probabilistically – even if it were possible to mark and follow a single molecule without disturbing and changing it.

From this point of view, physics and science in general may be accused of being pragmatic, which they indeed are to a considerable degree. However, it is pragmatic physics and science that brought us Moon rocks, pictures of the surface of Jupiter's satellites and of distant galaxies and nebulae, and which can find extrasolar planets; that produce artificial heart valves that the human immune system accepts; that can provide early signs of hurricanes, cyclones and tsunamis so as to warn the endangered population. Unfortunately, ethically and morally wrong, and just plain uninformed application of science may also lead to our planet radioactively glowing in the dark of the universe, or "only" to lose all ice and heat up to a point where life as we know it is no longer possible. Through this feedback, science also affects our thinking, our opinions and convictions, and so influences almost everything else, thereby being far more than "just pragma."

The foregoing also uncovers the price to pay: although physics is about Nature, it describes Nature indirectly, by way of the models that are sufficiently (and ever more) precise.

Example 1.1 The statement "in Rutherford's planetary model, the atom consists of a nucleus at the center and the electrons orbiting around it" does not mean that atoms literally exist in such simple tinkertoy form. More precisely, one means that the mathematical model developed from this *mental caricature* nevertheless reproduces the so far addressed real observations with sufficient accuracy.

[12] Let's recall: exact science always errs, but knows precisely how much [☞ Conclusion 1.1 on p. 6].

https://doi.org/10.1017/9781009291507.003 Published online by Cambridge University Press

In fact, the stability of atoms requires that Rutherford's planetary model is amended by additional "quantization" rules, which in turn lead to Bohr's model and the "old quantum mechanics." Subsequent observations caused a further development of this and subsequently developed models, leading through "quantum mechanics" to "quantum field theory," and even to "quantum theory of (super)strings." In this development, each stage completely contains the previous. Of course, it must be admitted that the current favorite for the *fundamental theory* – (super)string theory – is (by far) not confirmed experimentally as a theory of Nature: It has not even been shown that (super)string theory really can reproduce all the details of the "real World" as known so far, but it is the first one for which no indication to the contrary is currently known.

To be fair, (super)string theory is not one concrete theory but a theoretical system, just as classical mechanics is not limited to the description of a concrete physical system, but is a systematic approach to describing a class of concrete physical systems. The surprising abundance of unexplored possibilities in the theoretical system of (super)strings and the fact that no contraindication has been found, together provide hope that amongst the (super)string models an optimal candidate for the so-called Theory of Everything [☞ Chapter 11.5] can be found – with the requisite warning: this will continue to require a lot of hard work.

1.1.4 Scientific predictions: useful and inevitable

We are already acquainted with the three-step, observe–model–predict, as well as the logical inevitability of repeating this iterative cycle infinitely while developing, testing and asymptotically improving scientific models. Indeed, we may regard it both as a curse and as a boon that the idea of an *ultimate and complete* Theory of Everything is a vanishing point: a theory to which all scientific endeavors aim, asymptotically [☞ Section 11.5].

Apart from this asymptotic (un)reachability, Theory of Everything is also a misnomer, since it refers only to fundamental interactions in Nature. However, neither does a pile of rocks (and other building blocks) make a palace, nor does a few pounds of protein, lipids, some fat and calcium make a Schrödinger's cat. Even if the "ultimate" theory of all fundamental interactions were known, a hazy road would remain to lead from there, through atoms and molecules, to… us, our ambitious thoughts, and beyond.

Not only is there much room for "filling in the details" even if we stick to this 1-dimensional arrangement by size, but very often tiny portions in this all-encompassing size scale produce "pockets of knowledge" of fantastic and baffling complexity, a characteristic perhaps not unlike fractals. Suffice it here just to mention that the complexity of collective phenomena (behaviors 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} and

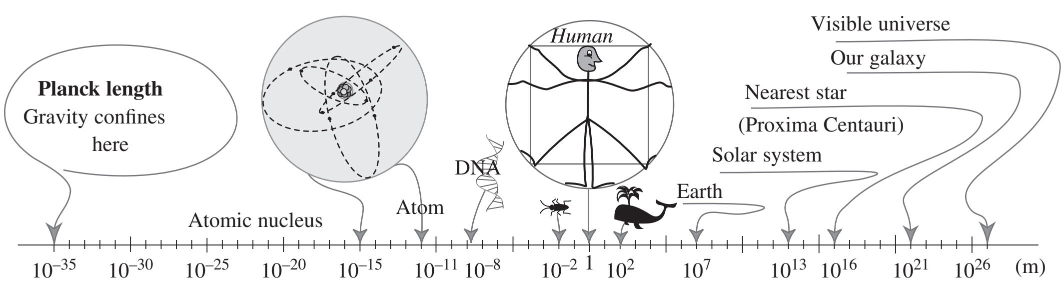

Figure 1.4 A logarithmic scale of sizes, from the Planck length, where everything looks like a black hole (from within which no information can be extracted), to the largest distances, from which the light only now reaches us (and from beyond which information has not yet reached us).

https://doi.org/10.1017/9781009291507.003 Published online by Cambridge University Press

10^2 m; chemistry occupies an even smaller niche around 10^{-9} m. Their complexity and richness are, however, obvious to every Student who has taken exams in these subjects.

Heeding the adage "when eating an elephant, take a bite at a time," physics *analyzes* natural phenomena (systems), identifying their sub-processes (sub-systems). These are usually more easily grasped and understood, upon which it however remains to (re-)integrate them. Along this journey, certain characteristics of the whole are recognized simply as a sum of these parts, while others are identified as intrinsically "collective" – unexplained by the characteristics of the integral parts and inextricably rooted in the complexity of the whole rather than the nature of the constituents. Whereas the analysis of the "parts" says little about the collective phenomena, it certainly permits a better specification and discussion of the properties that are *not* collective, thereby leaving a clearer path towards this complementary front in understanding Nature.

Following this "glory road" of scientific discovery, it is important to realize that:

> ***Claim 1.1*** *Whatever follows logically from the assumed axioms/postulates of a model is a prediction of the model.*

That is, if a model reproduces perfectly the original input/output data, and produces a number of testable predictions even just one of which turns out to disagree with Nature, there is something wrong about the model. It may happen that its minor modification will both fix the glitch and retain the model's 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.[13]

All predictions are derived as unavoidable consequences of the given assumptions, and are ensured by the rigor of mathematical deduction. Those very consequences and predictions are sometimes precisely the goal of developing the model; at other times, those are byproducts or afterthoughts. Once in a while, however, they are quite spectacularly unexpected discoveries:

> *The Heitler–London bond 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 [477]

The Heitler–London bond is one of the basic "ingredients" in modern chemistry, and we may rightly understand chemistry as based upon quantum statistical physics. Similarly spectacular was Dirac's prediction of the anti-electron (positron), and with it – as a logical consequence – an antiparticle for each other type of matter particle, as well as Pauli's prediction of the particle that Fermi named "neutrino," and which was confirmed experimentally only two decades later!

1.2 Measurement units and dimensional analysis

An exhaustive and detailed description of most real-life physical phenomena is often complex and requires technically demanding computations. However, satisfactory estimates may often be obtained by comparison with a similar and well-known case, by means of "dimensional analysis."

The next few examples in Section 1.2.1 illustrate this practice, using straightforward similarity in the form of simple proportions (scaling) of physical dimensions. Sections 1.2.3 through 1.2.5 show that dimensional analysis and a few other general properties of the quantities that interest

[13] Fortunately, and unlike their human inventors, theories can be resurrected, and this happens on occasion. The glitch that had formerly killed off a model may turn out to be "repairable" at a later time, when a better understanding of the model and requisite techniques and methods of analysis are attained [☞ Section 11.1, for example]. In turn, it can (and does) also happen that the experiments were carried out or analyzed in error, and this is revealed only much later. The corrected analysis may well turn out to agree with what was formerly thought of as a glitch of the model.

us importantly limit the possible answers, and may well at times even fully determine the form of those answers.

1.2.1 Lilliputians

In Jonathan Swift's *Gulliver's Travels*, Lemuel Gulliver meets the people of Lilliput, who are identical to humans, but smaller: their average height is 45 mm. In other words, Lilliputians are about $\lambda = 40$ times smaller copies of ordinary humans.

1. **How much weight (in units of their own weight) can a Lilliputian lift?**
   (a) The weight of the burden is determined by the force available for lifting it, and force is proportional to the area of the muscle cross-section. As the Lilliputian is built just like a normal human, the cross-section of his muscle is $\lambda^2 = 1{,}600$ times smaller.
   (b) The weight of the Lilliputian himself is $mg = \rho V g$, where g is the gravitational *constant*, ρ the density which equals that of ordinary humans, and V is the volume. Since the Lilliputian is 40 times shorter, his volume is $\lambda^3 = 64{,}000$ times smaller than in ordinary humans, so that his weight is also 64,000 times smaller.

   It follows that a Lilliputian can lift a 1/1,600 of the burden an ordinary human can, which is however $\lambda^{-2}/\lambda^{-3} = \lambda = 40$ times more – in units of his own (64,000 times smaller) weight – than an ordinary human. *Proportionally*, Lilliputians are $\lambda = 40$ times stronger than ordinary humans: if an ordinary human can lift his own weight, a Lilliputian can lift 40 times that much!
2. **How fast does the heart of a Lilliputian beat?**
   The frequency with which the heart pumps blood is determined by the quantity of blood it moves in a single beat, and by the quantity of blood needed to circulate in a unit of time. In warm-blooded beings, one of the main functions of blood circulation is to maintain the temperature and life in tissues by carrying oxygen. (When the circulation fails, tissues die and cool.)
   (a) The volume of a Lilliputian's heart is $\lambda^3 = 64{,}000$ times smaller than the volume in an ordinary human. The same holds for the volume of blood that the heart pumps in a single beat.
   (b) The body cools through the surface of the skin, which is $\lambda^2 = 1{,}600$ times smaller in a Lilliputian than in an ordinary human.

   It follows that a Lilliputian's heart must beat $\lambda^{-2}/\lambda^{-3} = \lambda = 40$ times faster than that in an ordinary human, so it would achieve the circulation of λ^{-2} times smaller volume with its (λ^{-3} times smaller) "pumping units." That is about $40 \times 70 = 2{,}800$ times per minute, or about 46.67 times every second. That is the tone of 46.67 Hz frequency, very near the second "F-sharp" from left, on a standard piano: A Lilliputian's heart thus *hums* – a little deeper than the humming of an AC/DC adapter. The skin of most small warm-blooded animals is covered by fur to reduce the heat loss, among other things, also so that their heart will not have to beat so fast.
3. **How high can a Lilliputian jump?**
   The jump-height is determined by the energy available for lifting the body: Energy E lifts a body of mass m to a height of $h = E/mg$.
   (a) The mass of a Lilliputian is $\lambda^3 = 64{,}000$ times smaller than the mass of an ordinary human.
   (b) The energy available for the jump stems from the work invested by the muscle force F that contracted $\triangle L$: $E = F \triangle L$. The muscle force is proportional to the area of its cross-section, which is $\lambda^2 = 1{,}600$ times smaller in a Lilliputian than in an ordinary

https://doi.org/10.1017/9781009291507.003 Published online by Cambridge University Press

human. The change in the muscle length is $\lambda = 40$ times shorter, making the energy available for the jump $(\lambda^2)(\lambda) = 64{,}000$ times smaller.

(c) Since both the energy and the mass in a Lilliputian are $\lambda^3 = 64{,}000$ times smaller than in an ordinary human, and the gravitational constant is a *constant*, it follows that the height of the jump is also a constant: a Lilliputian can jump just as high as an ordinary human. In units of his own height, however, an ordinary Lilliputian can jump $\lambda = 40$ times as high as an ordinary human.

— ❦ —

Only the physical proportions of the human body were considered in these examples – height and width, in correlation with the basic function of some of the body parts and the survival of the whole. Rightly, the above examples may seem like a pastime and are indeed too naive for a complete account [☞ Exercise 1.2.3, then e.g. Ref. [561], to begin with], but it should be clear that they indicate some of the basic principles behind the fact that there are no insects as big as storks (or horses), nor can warm-blooded animals be as small as an ant, nor can land animals grow (akin to King Kong) to the size of the largest whales.[14]

Let's turn, however, to more "concrete" applications, and with a more detailed application of physical "dimensions," i.e., units. Students of natural sciences are familiar with unit systems based on specifying the units for some three "basic" physical quantities; by a conventional standard, these are mass (M), length (L) and time (T). Table 1.1 on p. 24 gives numerical data for the unit systems that we use. Suffice it to say, every physical quantity may be measured in units that are of the form $M^\alpha L^\beta T^\gamma$, for some (α, β, γ) [☞ also Table C.4 on p. 527]. In the next examples, our goal is to determine the triple (α, β, γ) for the desired physical quantities and thereby, to a large extent, the physical quantities themselves. A more general treatment of this "dimensional analysis" may be found in the book [208], and the books [244, 415, 416] abound with examples from everyday life where a little critical and mathematical analysis leads to sometimes unexpected results.

1.2.2 Characteristic scales

Take, for instance, a pendulum of length ℓ and mass m. Without writing down and solving the equations of motion, we can estimate the frequency as follows:

1. Neglecting dissipative forces, the frequency may depend only on the physical properties of the pendulum, length ℓ and mass m, and the gravitational constant g.
2. Using dimensional analysis, we have

$$[\ell] = L, \quad [m] = M, \quad [g] = \frac{L}{T^2}, \quad \text{we need} \quad [\nu] = \frac{1}{T}. \tag{1.1}$$

Assuming that the frequency ν is an analytic function of ℓ, m and g, we seek solutions of the form $\nu \propto k^\alpha\, m^\beta\, g^\gamma$, and find

$$\frac{1}{T} = [\nu] = [\ell]^\alpha\,[m]^\beta\,[g]^\gamma = L^\alpha\, M^\beta \left(\frac{L}{T^2}\right)^\gamma = \frac{L^{\alpha+\gamma}\, M^\beta}{T^{2\gamma}}. \tag{1.2}$$

[14] A mature blue whale (*Balaenoptera musculus*) can reach 30 m in length. Fossils indicate that some of the prehistoric, swamp-dwelling animals could reach the length of ~ 60 m (*Amphicoelias fragilimus*). However, the build of such animals was mostly horizontal, with a long and massive neck and tail, considerably different from modern warm-blooded land animals; *T. Rex*, built akin to a kangaroo, was no longer than about 13 m.

This implies the system of equations

$$\left.\begin{array}{r}\alpha+\gamma=0,\\ \beta=0,\\ 2\gamma=1,\end{array}\right\} \quad\Rightarrow\quad \left\{\begin{array}{l}\alpha=-\frac{1}{2},\\ \beta=\ \ 0,\\ \gamma=\ \ \frac{1}{2}.\end{array}\right. \tag{1.3}$$

Note that this result indicates that the frequency of oscillations is independent of the mass of the swinging bead, and specifies

$$\nu \propto \sqrt{\frac{g}{\ell}}. \tag{1.4}$$

Up to a numerical factor that depends on the definition of "frequency" (ν *vs.* $\omega = 2\pi\nu$), the expression (1.4) is in fact the exact formula, even for large oscillations (while the oscillation angle remains within $\frac{\pi}{2}$ to either side of the equilibrium point)! The result (1.4) merely acquires an overall $O(1)$ multiplicative numerical correction, growing monotonously from 1 to about 1.18 as a function of the amplitude. The fact that the frequency of (small) oscillations does not depend on the mass of the pendulum ($\beta = 0$) may come as a surprise, but is easily verified by simple experiments.

— ❦ —

A similar, but in some ways more interesting, example is presented by a bead of mass m, oscillating at the end of an elastic spring, well described as producing a linear restoring force, $F = -kx$, when stretched or compressed a length x. Now consider having this spring with the bead hanging vertically, and we proceed as before:

1. The frequency may depend only on the physical properties of the hanging spring, (k, m), and the gravitational constant g, where we again neglect dissipative forces.
2. To use dimensional analysis, we must first determine the dimensions (physical units) of the spring constant k. To this end, we may use the restoring force law, $F = -kx$, knowing that a force must have units in agreement with Newton's 2nd law:

$$[F] = [m\,a] = M \cdot \frac{L}{T^2}, \quad [F] = [-k\,m] \quad \text{so} \quad [k] = \frac{[F]}{[x]} = \frac{M}{T^2}. \tag{1.5}$$

3. To use dimensional analysis, we again list

$$[k] = \frac{M}{T^2}, \quad [m] = M, \quad [g] = \frac{L}{T^2}, \quad \text{we need} \quad [\nu] = \frac{1}{T}, \tag{1.6}$$

and again seek solutions in the form $\nu \propto k^\alpha m^\beta g^\gamma$:

$$\frac{1}{T} = [\nu] = [k]^\alpha\,[m]^\beta\,[g]^\gamma = \left(\frac{M}{T^2}\right)^\alpha M^\beta \left(\frac{L}{T^2}\right)^\gamma = \frac{L^\gamma\,M^{\alpha+\beta}}{T^{2\alpha+2\gamma}}. \tag{1.7}$$

This implies the system of equations

$$\left.\begin{array}{r}\gamma=0,\\ \alpha+\beta=0,\\ 2\alpha+2\gamma=1,\end{array}\right\} \quad\Rightarrow\quad \left\{\begin{array}{l}\alpha=\ \ \frac{1}{2},\\ \beta=-\frac{1}{2},\\ \gamma=\ \ 0.\end{array}\right. \tag{1.8}$$

This result clearly indicates that the frequency of oscillation of the bead on the spring is independent of the gravitational acceleration ($\gamma = 0$). This implies that the bead on the

https://doi.org/10.1017/9781009291507.003 Published online by Cambridge University Press

spring oscillates with the same frequency regardless of the orientation of the spring in the gravitational field – and even in the absence of any gravitational field! The result (1.8) implies

$$\nu \propto \sqrt{\frac{k}{m}}, \tag{1.9}$$

which is, up to the numerical factor 2π, again the exact formula and holds even for large oscillations as long as the spring "constant" is an analytic function of the displacement x.

Comment 1.3 *Note that no combination of ℓ, m, g and of k, m, g, respectively, is dimensionless. [✎ Verify.] With the uniqueness of the solutions (1.3) and (1.8), this implies the exactness of the results (1.4) and (1.9), respectively. Compare this with the situation described in Section 1.2.4.*

Comment 1.4 *Mathematically, the results (1.3) and (1.8) are very similar. Physically, however, the facts that the result (1.4) does not depend on the mass of the pendulum nor the result (1.9) on the gravitational acceleration both imply a much wider applicability of these results than initially conceived – but in two different ways. The first lets us freely swap the bobs of a fixed-length pendulum near the surface of the Earth, while the latter lets us predict the oscillations of a spring in Skylab, on the surface of the Moon or Mars, or anywhere else where the gravitational acceleration is approximately constant!*

This realization – that a given mathematical model may be far more widely applicable than originally intended – tends to be extremely important in practice.

1.2.3 Larmor's formula

Larmor's formula for the energy per unit time (therefore, power) that an electric charge q loses during deceleration $\vec{a}$ is

$$P = \frac{2}{3}\frac{q^2\,a^2}{c^3}\ \text{(cgs)}, \qquad P = \frac{q^2\,a^2}{6\pi\epsilon_0\,c^3} = \frac{2}{3}\frac{1}{4\pi\epsilon_0}\frac{q^2\,a^2}{c^3}\ \text{(SI)}, \qquad \text{where} \quad a = |\vec{a}|. \tag{1.10}$$

Dimensional analysis We are interested in *power* – energy loss per unit time, for which the units are

$$[P] = \frac{[\triangle E]}{[\triangle t]} = \frac{M\,L^2/T^2}{T} = \frac{M\,L^2}{T^3}. \tag{1.11}$$

Energy that changes (decreases) in the process of electromagnetic radiation is certainly of electromagnetic origin, and for electrostatic (Coulomb potential) energy V_C the units are

$$\frac{M\,L^2}{T^2} = [V_C] = \left[\frac{1}{4\pi\epsilon_0}\frac{q_1\,q_2}{r}\right], \quad \Rightarrow \quad \left[\frac{q}{\sqrt{4\pi\epsilon_0}}\right] = \frac{M^{1/2}\,L^{3/2}}{T}. \tag{1.12}$$

The electric charge may thus be expressed in "mechanical" units, $\sqrt{\text{kg}\,\text{m}^3}/\text{s}$. The so rescaled quantity $q' := \frac{q}{\sqrt{4\pi\epsilon_0}}$ is sometimes called the *rationalized* electric charge [☞ [407, 29, 339] for an application in quantum mechanics].

The power lost through electromagnetic radiation through deceleration must depend on the acceleration, $\vec{a}$. As this is a vector and power is a scalar, the power may depend only on the magnitude of the acceleration, $|\vec{a}|^\beta = a^\beta$. Other than this, the power may only depend on the *universal constant*, c (speed of light in vacuum) – and, of course, on the electric charge[15]:

$$[q^\alpha\,a^\beta\,c^\gamma] = \left(\frac{M\,L^3}{T^2}\right)^{\alpha/2}\left(\frac{L}{T^2}\right)^{\beta}\left(\frac{L}{T}\right)^{\gamma} = \frac{M^{\alpha/2}\,L^{3\alpha/2+\beta+\gamma}}{T^{\alpha+2\beta+\gamma}} \stackrel{!}{=} \frac{M\,L^2}{T^3}. \tag{1.13}$$

[15] The symbol "$\stackrel{!}{=}$" denotes an equality that is *required* to hold [☞ Tables C.7 and C.8 on p. 529].

https://doi.org/10.1017/9781009291507.003 Published online by Cambridge University Press

Comparing, it follows that

$$\left.\begin{array}{r}\tfrac{1}{2}\alpha \qquad\qquad = 1,\\ \tfrac{3}{2}\alpha + \ \ \beta + \gamma = 2,\\ \alpha + 2\beta + \gamma = 3,\end{array}\right\} \quad\Rightarrow\quad \left\{\begin{array}{r}\alpha = \ \ 2,\\ \beta = \ \ 2,\\ \gamma = -3,\end{array}\right. \tag{1.14}$$

so that

$$P \propto \frac{q^2\,a^2}{c^3}. \tag{1.15}$$

The numerical factor $\frac{2}{3}$ in Larmor's formula (1.10) cannot be determined by dimensional analysis, whereas the presence or absence of the $\frac{1}{4\pi\epsilon_0}$ factor is determined by the choice of units – SI or cgs, for example.

1.2.4 Perturbations of stationary states in quantum mechanics

Assume a 1-dimensional (non-relativistic) quantum system specified with the Hamiltonian[16] H_0, for which the stationary solutions are known:

$$H_0|n\rangle = E_n^{(0)}|n\rangle, \qquad U_0|n\rangle = e^{-i\omega_n t}|n\rangle, \qquad \omega_n := E_n^{(0)}/\hbar, \tag{1.16a}$$

$$H_0\Psi_n(x,0) = E_n^{(0)}\Psi_n(x,0), \qquad \Psi_n(x,t) = e^{-i\omega_n t}\,\Psi_n(x,0). \tag{1.16b}$$

In addition, using the Gram–Schmidt orthogonalization procedure, we may always arrange the space of solutions so that

$$\mathscr{H} := \Big\{ |n\rangle \;:\; H_0|n\rangle = E_n^{(0)}|n\rangle,\ \langle n|m\rangle = \delta_{n,m},\ \sum_n |n\rangle\langle n| = \mathbb{1} \Big\} \tag{1.16c}$$

is the Hilbert space of states of the system. The notation must be understood symbolically: Here, n represents any system of numbers: one or more, discrete and/or continuous, including also hybrid value-sets. The latter is the case with the familiar hydrogen atom, where the symbol n in the results (1.16) stands for the familiar collection of "quantum numbers" (n,ℓ,m,m_s). For bound states with negative energies, all four of these numbers vary over discrete values; for scattering states with positive energies, one of those numbers *[✐ which one?]* varies over continuous values, while the other three remain quantized.

Consider now a similar quantum system, differing from (1.16a)–(1.16c) by a *perturbation* Hamiltonian, $H' = H - H_0$. To begin with, let H' be independent of time and let H' – as an operator! – be *small*. That is, the effect of the change $H_0 \to H$ on the energies of stationary states and on the states themselves is, we assume, small. More precisely, we assume that these changes are analytic, i.e., may be expanded in a power series, which has been named after Brook Taylor since 1715. We then have that

$$E_n^{(1)} = \langle n|H'|n\rangle; \tag{1.17}$$

$$|n\rangle^{(1)} = -\sum_{m\neq n}\!\!\!\!\!\!\int \frac{\langle m|H'|n\rangle}{E_m^{(0)} - E_n^{(0)}}\,|m\rangle; \tag{1.18}$$

$$E_n^{(2)} = -\sum_{m\neq n}\!\!\!\!\!\!\int \frac{|\langle m|H'|n\rangle|^2}{E_m^{(0)} - E_n^{(0)}}; \qquad \text{etc.} \tag{1.19}$$

[16] After the Irish mathematician William Rowan Hamilton.

https://doi.org/10.1017/9781009291507.003 Published online by Cambridge University Press

A "shoestring" explanation Perturbation corrections of kth order must be proportional to the kth power of the perturbation operator H' – were it not for H', there would be no corrections. Thus:

1. $E_n^{(1)}$ is a real quantity that must be proportional to the first power of H'. Other than this, $E_n^{(1)}$ may depend only on the nth state, and so must be the expectation value of H' in the nth state, as given in equation (1.17). Besides, however $|n\rangle$ and $\langle n|$ may be normalized, $\langle n|m\rangle$ must be dimensionless; thus $\langle m|A|n\rangle$ must have the same dimensions (physical units) and physical character (scalar, vector… time dependence…) as does A, so that the dimensions (physical units) in equation (1.17) also agree.
2. The first correction of the $|n\rangle$ state cannot be proportional to that same state, as an addition of such a correction to a state would change the norm:

$$\big\||n\rangle\big\|^2 = 1 \quad \to \quad \big\||n\rangle + \epsilon|n\rangle\big\|^2 = \langle n|n\rangle + 2\epsilon\langle n|n\rangle + \epsilon^2\langle n|n\rangle \\ = 1 + 2\epsilon + O(\epsilon^2) \not\approx 1. \tag{1.20}$$

   Whence "$m \neq n$" in the sum/integral (1.18), and

$$\big\||n\rangle\big\|^2 = 1 \quad \to \quad \big\||n\rangle + \epsilon|m\rangle\big\|^2 = \langle n|n\rangle + \epsilon\big(\langle m|n\rangle + \langle n|m\rangle\big) + \epsilon^2\langle m|m\rangle \\ = 1 + O(\epsilon^2) \approx 1, \tag{1.21}$$

   since $\langle m|n\rangle = 0$ for $m \neq n$. Furthermore, since $|m\rangle$ form a complete basis (cf. Sturm–Liouville theorem for eigen-problems of Hermitian operators), $|n\rangle^{(1)}$ must be expandable in $|m\rangle$'s, as in equation (1.18). Comparing the left- and the right-hand side, the coefficients in the sum must be proportional to a matrix element of H'. Since $|n\rangle$ is on the left-hand side of the equation, it must also occur on the right-hand side, so that a linear change in the basis $|n\rangle$ would change both sides of the equation equally.

   It follows that the coefficient in the right-hand sum must depend linearly on $\langle m|H'|n\rangle$, which is the amplitude of probability that H' will change $|n\rangle \to |m\rangle$. As that matrix element has the dimensions of energy as does H', it must be divided by some energy – whence "$E_m^{(0)} - E_n^{(0)}$" in the denominator, which is the energy of the transition $|n\rangle \to |m\rangle$ described by the matrix element $\langle m|H'|n\rangle$.
3. The result (1.19) follows on applying first H' and then $\langle n|$ to equation (1.18).

Assume now that the perturbation Hamiltonian depends on time, $H' = H'(t)$. The probability amplitude for the transition from the initial (i) into the final ($f \neq i$) state[17] is then

$$a_{fi}^{(1)}(T) = \frac{1}{i\hbar}\int_{t_0}^{T>t_0} \mathrm{d}\tau\, \langle f|H'(\tau)|i\rangle\, e^{i\omega_{fi}\tau}, \qquad \omega_{fi} := |E_f - E_i|/\hbar, \tag{1.22}$$

to first order in perturbation theory. This result may be – up to the $1/i$ factor – explained by the same type of "shoestring" arguments and dimensional analysis as the above results for stationary perturbative corrections.

Note that for transitions with very large differences in energies, the frequency is very large, the $e^{i\omega_{fi}\tau}$ factor oscillates very fast, and this causes an effective cancellation in the integral. By contrast, for transitions with very small energy difference, the frequency is very small, and the $e^{i\omega_{fi}\tau}$ factor a priori does not diminish the contribution to the integral. A very similar behavior in integration occurs in the Feynman–Hibbs method of quantization [☞ Procedure 11.1 on p. 416, and Ref. [165]].

[17] More precisely, this is the probability amplitude that if the system was prepared in the initial system i at time t_0, at a later time $T > t_0$ it may be detected in the state $f \neq i$.

1.2.5 And, caution!

Consider now a hydrogen-type atom. The binding energy of such an atom must depend on the (reduced) electron mass, m_e, the electron charge, $-e$, and the charge of the nucleus, $+Ze$. The Coulomb force, which holds the atom together, is proportional to the product of charges, for which the relation (1.12) holds. Notice that the atomic number Z always accompanies e^2, going with the one factor of e that stems from the electric charge of the nucleus, which has Z protons. It then follows that the combination $(m_e)^\alpha(Ze^2)^\beta$ has units of $M^{\alpha+\beta}\,L^{3\beta}\,T^{-2\beta}$ and there is no choice of α,β for $(m_e)^\alpha(Ze^2)^\beta$ to have the dimensions (units) of energy, $\frac{ML^2}{T^2}$. For a formula that specifies the energy of a hydrogen-type atom, we need at least one more characteristic quantity, the dimensions (units) of which differ from those of all monomials $(m_e)^\alpha(Ze^2)^\beta$.

Such a characteristic quantity may well be provided by the natural constant c; its units indeed differ from those of $(m_e)^\alpha(Ze^2)^\beta$. More importantly, however, although the electron and the proton may be moving non-relativistically within the atom, they are connected by the electromagnetic field, which certainly does propagate relativistically. We thus seek a Coulomb solution to

$$[E_C] = \frac{M\,L^2}{T^2} = [(m_e)^x][(Ze^2)^y][c^z] = M^x\left(\frac{M\,L^3}{T^2}\right)^y\left(\frac{L}{T}\right)^z = \frac{M^{x+y}\,L^{3y+z}}{T^{2y+z}}, \tag{1.23}$$

i.e.,

$$\left.\begin{aligned} x + y \quad\;\; &= 1,\\ 3y + z &= 2,\\ 2y + z &= 2, \end{aligned}\right\} \quad\Rightarrow\quad \left\{\begin{aligned} x &= 1,\\ y &= 0,\\ z &= 2, \end{aligned}\right. \tag{1.24}$$

and obtain

$$E_C \propto m_e c^2 \approx 0.511\,\text{MeV}. \tag{1.25}$$

This is, of course, incorrect: $E_C = 0.511\,\text{MeV} \gg |E_1| = 13.6\,\text{eV}$. Besides, this estimate for E_C turns out to be independent of the electric charge, which is just plain wrong for the binding energy of the atom: were it not for the electric charge, there would be no atom as a bound state. Moreover, it is impossible to construct a dimensionless quantity

$$[(m_e)^x][(Ze^2)^y][c^z] = M^x\left(\frac{M\,L^3}{T^2}\right)^y\left(\frac{L}{T}\right)^z = \frac{M^{x+y}\,L^{3y+z}}{T^{2y+z}} \stackrel{!}{=} \frac{M^0\,L^0}{T^0}, \tag{1.26}$$

i.e.,

$$\left.\begin{aligned} x + y \quad\;\; &= 0,\\ 3y + z &= 0,\\ 2y + z &= 0, \end{aligned}\right\} \quad\Rightarrow\quad \left\{\begin{aligned} x &= 0,\\ y &= 0,\\ z &= 0. \end{aligned}\right. \tag{1.27}$$

The binding energy of a hydrogen-type atom therefore *must* depend on a *fourth* characteristic quantity! Only then will it be possible to construct a dimensionless monomial from these four characteristic quantities. A suitable power of this dimensionless monomial could then rescale $E_C = 0.511\,\text{MeV} \to |E_1| = 13.6\,\text{eV}$.

Bohr's postulate introduces just such a "new" characteristic quantity: $\hbar$, a unit of angular momentum, with physical dimensions $[\hbar] = \frac{ML^2}{T}$. With this new quantity, we have

$$\begin{aligned} [E_{\text{H}}] = \frac{M\,L^2}{T^2} &= [(m_e)^x][(Ze^2)^y][c^z][\hbar^w] = M^x\left(\frac{M\,L^3}{T^2}\right)^y\left(\frac{L}{T}\right)^z\left(\frac{M\,L^2}{T}\right)^w \\ &= \frac{M^{x+y+w}\,L^{3y+z+2w}}{T^{2y+z+w}}, \end{aligned} \tag{1.28}$$

and it follows that

$$\left.\begin{aligned} x + \;y \quad\;\; + \;\; w &= 1,\\ 3y + z + 2w &= 2,\\ 2y + z + \;\; w &= 2, \end{aligned}\right\} \quad\Rightarrow\quad \left\{\begin{aligned} x &= 1,\\ y &= -w,\\ z &= 2 + w, \end{aligned}\right. \tag{1.29}$$

https://doi.org/10.1017/9781009291507.003 Published online by Cambridge University Press

or

$$E_{\rm H} \propto m_e\,(Ze^2)^{-w}\,c^{2+w}\,\hbar^w = \left(\frac{\hbar c}{Ze^2}\right)^w m_e c^2. \tag{1.30}$$

Owing to the fact (1.12), the quantity raised to the wth power is dimensionless. Leaving Z to vary with the atom, we evaluate $\frac{(4\pi\epsilon_0)\hbar c}{e^2} \approx 137.036$ and see that the value $w = -2$ would indeed provide the desired $E_C = 0.511\,\text{MeV} \to |E_1| = 13.6\,\text{eV}$ rescaling.

Indeed, equation (1.30) is remarkably close to Bohr's formula, which may be written as

$$E_n = -2\alpha_e^2\,(m_e c^2)\,\frac{Z^2}{(2n)^2}, \qquad \alpha_e := \frac{e^2}{(4\pi\epsilon_0)\hbar c} \approx \frac{1}{137.036}. \tag{1.31}$$

The so-called "fine structure constant" (coupling parameter) α_e is indeed the dimensionless monomial predicted by equation (1.30). To be precise, dimensional analysis can predict only that

$$E_n \propto f(\alpha_e; n)\,(m_e c^2), \tag{1.32}$$

with no information about the *arbitrary* dimensionless function $f(\alpha_e; n)$.

It is also known that the "fine structure" corrections to the energy levels depend on relativistic corrections and the spin–orbital interaction – neither of which introduces a new physical quantity:

$$\triangle E_{\rm fs} = -\alpha_e^4\,(m_e c^2)\,\frac{1}{(2n)^2}\left(\tfrac{2n}{j+\frac{1}{2}} - \tfrac{3}{2}\right), \qquad j := \ell \pm \tfrac{1}{2} \quad \to \text{ degeneracy.} \tag{1.33}$$

This degeneracy refers to the spectrum (the collection of all eigenvalues) of the Hamiltonian, as there exist two states with the same energy for every ℓ, having $j = \ell \pm \frac{1}{2}$. Since the rest energy of the electron is $m_e c^2$, the sequence

$$|m_e c^2| \;:\; |E_n| \;:\; |\triangle E_{\rm fs}| \quad = \quad \alpha_e^0 \;:\; \alpha_e^2 \;:\; \alpha_e^4 \tag{1.34}$$

suggests that the hydrogen atom energy is an analytic function of the formal variable "α_e^2" and not of α_e:

$$E_n(\alpha_e) = m_e c^2 \sum_{k=0}^{\infty} C_k\,\alpha_e^{2k}, \tag{1.35}$$

$$C_0 = 1, \quad C_1 = -\tfrac{1}{2n^2}, \quad C_2 = -\tfrac{1}{4n^2}\left(\tfrac{2n}{j+\frac{1}{2}} - \tfrac{3}{2}\right), \quad \text{etc.} \tag{1.36}$$

Indeed, Sommerfeld's relativistic formula[18] [407] from 1915:

$$E_{nk} = \frac{m_e c^2}{\sqrt{1 + \left(\frac{\alpha_e}{n-k+\sqrt{k^2-(Z\alpha_e)^2)}}\right)^2}}, \qquad k = 1,2,\dots,n, \quad n = 1,2,\dots \tag{1.37}$$

gives an excellent description of the bound states of hydrogen-type atoms and supposes elliptic orbits for the electron, where n, k quantify the size and ellipticity of the classical orbit. It is easy to see that Sommerfeld's expression (1.37) depends analytically on α_e and that the Taylor expansion only has even powers.

[18] Sommerfeld's derivation assumes that $k = \ell + 1$ measures the electron's angular momentum. Subsequent derivations [122, 121, 216] based on Dirac's relativistic theory of the electron obtain the same final result, but with $k = j + \frac{1}{2}$, where $j = \ell \pm \frac{1}{2}$ owing to the electron's spin, which explains the residual and observed two-fold degeneracy of the bound states.

https://doi.org/10.1017/9781009291507.003 Published online by Cambridge University Press

Digression 1.2 However, the conclusion that the energy is an analytic function of α_e^2 is not *completely* true of the hydrogen-type atoms: There exists the so-called Lamb shift, for which

$$\triangle E_{\text{Lamb}} \approx \alpha_e^5\,(m_e c^2)\,\frac{1}{(2n)^2}\,\frac{1}{n}\Big(E_L(n,\ell) \pm \frac{1}{\pi(j+\frac{1}{2})(\ell+\frac{1}{2})}\Big) \tag{1.38a}$$

is an adequate approximation, with $|E_L(n,\ell)| < 0.05$ and where an odd power of α_e appears manifestly. Also, this "hyperfine" structure of the energy levels is further complicated by contributions from the interactions with the nucleus. These contributions then depend also on the proton mass m_p, through its magnetic moment:

$$\vec{\mu}_p := \gamma_p \frac{e}{m_p c}\vec{S}_p \quad \text{as compared to} \quad \vec{\mu}_e := \frac{e}{m_e c}\vec{S}_e; \quad \gamma_p = 2.7928. \tag{1.38b}$$

The ratio $(m_e/m_p) \approx 1/2{,}000$ provides a new dimensionless constant, whereby the energy formula complicates additionally:

$$\triangle E_{\text{hfs}} = \Big(\frac{m_e}{m_p}\Big)\,\alpha_e^4\,(m_e c^2)\,\frac{4\gamma_p}{2n^3}\,\frac{\pm 1}{(f+\frac{1}{2})(\ell+\frac{1}{2})} + \cdots, \tag{1.38c}$$

where $f(f+1)\hbar^2$ is the eigenvalue of the operator $(\vec{L}+\vec{S}_e+\vec{S}_p)^2$.

Comment 1.5 *Expanding on the result (1.30), a general property of hydrogen-type atoms is worth noting: The binding energy of hydrogen-type atoms must depend on* ***four*** *characteristic constants of the system:*

1. *the reduced mass[19] of a sub-system that is bound to the other – here, m_e;*
2. *the interaction coefficient – here, the product of charges, (Ze^2);*
3. *the speed with which the interaction travels between the sub-systems – here, c, for the electromagnetic interaction;*
4. *the unit (quantum) of the interaction action (since the classical atom is unstable, and is stabilized by angular momentum quantization) – here, $\hbar$.*

The existence of more than three characteristic constants of the system $(m_e, e, c, \hbar)$ guarantees the existence of a dimensionless characteristic constant $\alpha_e = \frac{e^2}{(4\pi\epsilon_0)\hbar c}$, since the system (1.29) consists of only three linear equations in four unknowns. The existence of the dimensionless α_e then permits a formula such as (1.32), which we may expand:

$$E_n = E_{n,0} + \alpha_e\,E_{n,1} + \alpha_e^2\,E_{n,2} + \cdots \tag{1.39a}$$

and notice that for the binding energy of a hydrogen-type atom:

$$E_{n,0} = 0, \qquad E_{n,1} = 0, \qquad E_{n,2} \neq 0. \tag{1.39b}$$

Only the coefficients of the second (and fourth, then fifth...) order in the α_e-expansion differ from zero!

[19] By "mass" of a particle, we always mean the relativistic-invariant quantity, which is also (needlessly) called the "rest mass" [☞ Section 3.1.3, and especially the result (3.36)].

Definition 1.1 *Bound-state systems may be roughly classified as:*

$$\left.\begin{array}{r}\textit{weakly bound}\\ \textit{strongly bound}\\ \textit{very strongly bound}\end{array}\right\}\ \textit{if}\ \left(\frac{\textit{binding energy}}{\textit{rest energy}}\right)\ \left\{\begin{array}{l}<1,\\ \approx 1,\\ >1.\end{array}\right. \begin{array}{r}(1.40a)\\(1.40b)\\(1.40c)\end{array}$$

Conclusion 1.3 *Since $\alpha_e \approx \frac{1}{137} < 1$, the result (1.39b) implies that the hydrogen atom is a weakly bound system. Indeed, the ratio $13.6\,\text{eV}/m_e c^2 = \alpha^2/2 \approx 2.67 \times 10^{-5}$.*

1.2.6 Exercises for Section 1.2

✎ **1.2.1** *Taking into account that both the load-bearing ability of bones and the muscle force is proportional to the cross-section area, and the height of a Lilliputian is $\lambda = 40$ times smaller than the height of an ordinary human, estimate:*

1. *the width of the legs and torso in a Lilliputian body for which the strain in the bones and muscles are about the same as in ordinary humans;*
2. *the weight of a typical Lilliputian;*
3. *the ensuing corrections in the previous estimates of weight-bearing, heartbeat and height of jump.*

✎ **1.2.2** *In his subsequent travels, Lemuel Gulliver found himself in Brobdingrag,*[20] *where the population is about $\Lambda = 40$ times taller than ordinary humans. Estimate:*

1. *How much weight (in units of their own weight) can a Brobdingragian lift?*
2. *How fast is a Brobdingragian's heartbeat?*
3. *How high can a Brobdingragian jump?*
4. *If a Brobdingragian is $\Lambda = 40$ times taller than an ordinary human, estimate:*
   (a) *the width of the legs and torso in a Brobdingragian body for which the strain in the bones and muscles are about the same as in ordinary humans;*
   (b) *the weight of a typical Brobdingragian;*
   (c) *the ensuing corrections in the previous estimates of weight-bearing, heartbeat and height of jump.*

✎ **1.2.3** *If humankind ever colonizes Mars, one would expect that the native generations will in time adapt to the four times weaker gravity. (Suppose that the breathing equipment is of negligible weight.) Estimate the changes in the height : width ratio in the body of a fully adapted* Homo Aresiensis, *and from there the other characteristics mentioned in the previous questions.*

✎ **1.2.4** *Estimate the lifetime of the hydrogen atom caused by electron bremsstrahlung, using the Larmor formula to estimate the radiation energy loss. The atom may be regarded as collapsed when the electron "falls" into the nucleus,* i.e., *when the radius of the electron's orbit reduces from $\sim 10^{-10}$ m to about $\sim 10^{-15}$ m.*

✎ **1.2.5** *Prove the statement made in Comment 1.3 on p. 17.*

[20] Complete editions of Jonathan Swift's novel contain a note supposedly added after the first printing wherein the fictitious Lemuel Gulliver explains to his cousin Sympson (who mediated the publishing of the novel) that the printers erroneously printed the name of the land as "Brobdingnag."

✎ ***1.2.6*** *Considering the discussion around the equations (1.16) in the first paragraph of Section 1.2.4, identify which of the quantum numbers* (n, ℓ, m) *in the Bohr model of the hydrogen atom becomes continuous for scattering states and which must remain discrete. Prove this by re-examining the familiar wave-function (4.8c) upon changing* $(E < 0) \to (E > 0)$*; the discussion in Appendix A.3 should be helpful for the complementary part of the question.*

✎ ***1.2.7*** *Compute what exactly changes in the formulae (1.31) and (1.33):*

1. *if the electron in a hydrogen atom is replaced by a muon:* $m_\mu \approx 207\, m_e$*;*
2. *if the electron in a hydrogen atom is replaced by an antiproton:* $m_p \approx 1{,}836\, m_e$*;*
3. *if the proton in a hydrogen atom is replaced by a positron,* e^+*:* $m_{e^+} = m_e$*.*

✎ ***1.2.8*** *Would the formulae (1.31) and (1.33) hold for a hypothetical* $Z > 137$ *atom? Why? (Hint: consider the consequences of the relations (1.40a)–(1.40c).)*

1.3 The quantum nature of Nature and limits of information

Nature is both quantum and relativistic; the constants $\hbar$ and c are *universal*. Also, Newton's law of gravity – *extended* by Einstein's general theory of relativity – is also universal, and so also is Newton's constant, G_N. Its units are

$$F_G = G_N \frac{m_1\, m_2}{r^2}, \quad \Rightarrow \quad [G_N] = \frac{[F_G][r^2]}{[m^2]} = \frac{M\, L\, L^2}{T^2\, M^2} = \frac{L^3}{T^2\, M}. \tag{1.41}$$

Table 1.1 Natural (Planck) units and their SI equivalent values

Name	Expression	SI equivalent	Practical equivalent
Length	$\ell_P = \sqrt{\frac{\hbar G_N}{c^3}}$	$1.616\,25 \times 10^{-35}$ m	
Mass	$M_P = \sqrt{\frac{\hbar c}{G_N}}$	$2.176\,44 \times 10^{-8}$ kg	$1.220\,86 \times 10^{19}$ GeV/c^2
Time	$t_P = \sqrt{\frac{\hbar G_N}{c^5}}$	$5.391\,24 \times 10^{-44}$ s	
El. charge[a]	$q_P = \sqrt{4\pi\epsilon_0\, \hbar c}$	$1.875\,55 \times 10^{-18}$ C	$e/\sqrt{\alpha_e} \approx 11.706\,2\, e$
Temperature	$T_P = \frac{1}{k_B} M_P c^2$	$1.416\,79 \times 10^{32}$ K	

[a] $\alpha_e \approx 1/137.035\,999\,679$ in low-energy scattering experiments, but grows to about $1/127$ near $\sim$200 GeV energies [☞ Section 5.3.3].

From this, we define the Planck, i.e., *natural* units, listed in Table 1.1. A comparison with SI equivalents makes it clear that the natural units are in no way reasonable when describing everyday events of typical human proportions: it would be hilariously ridiculous to try buying milk in units of Planck volume ($1\,\text{gal} = 8.964 \times 10^{98}\, \ell_P{}^3$), hot dogs in units of Planck mass ($16\,\text{oz} = 2.084 \times 10^7\, M_P$), or measure the time to the recess bell in units of Planck time ($45\,\text{min} = 5.008 \times 10^{46}\, t_P$). However, natural units do indicate certain limiting values, and this is worth exploring when considering ever smaller systems.

In fact, the natural units in Table 1.1 are not very convenient even for typical contemporary elementary particle physics: the electron mass is $m_e = 4.185\,45 \times 10^{-23}\, M_P$! Therefore, one frequently uses units such as "MeV/c^2," so $m_e = 0.510\,999$ MeV/c^2. In this system of "particle physics" units, we formally state $\hbar = 1 = c$ – that is, we use the unit system where $\hbar$ and c are two of

https://doi.org/10.1017/9781009291507.003 Published online by Cambridge University Press

the basic three units, *and then do not write them*. All physical quantities can then be expressed as various powers of one particular unit of measurement, for which the usual choice is energy. For this, one typically uses the "eV" unit, with the usual SI prefixes. Table 1.2 lists some relations useful in typical calculations.

Table 1.2 Some typical physical quantities, expressed in "particle physics" and SI units

Quantity	Particle physics	SI equivalent
Energy	$x\,\mathrm{MeV}$	$= x \times 1.602\,18 \times 10^{-13}\,\mathrm{J}$
Mass	$x\,\mathrm{MeV}/c^2$	$= x \times 1.782\,66 \times 10^{-30}\,\mathrm{kg}$
Length	$x\,\hbar c/\mathrm{MeV}$	$= x \times 1.973\,27 \times 10^{-13}\,\mathrm{m}$
Time	$x\,\hbar/\mathrm{MeV}$	$= x \times 6.582\,12 \times 10^{-22}\,\mathrm{s}$

1.3.1 Smaller, and smaller, and . . .

To a great extent, the division and analysis of phenomena, processes and systems happens just as it does in the most obvious application of the black box paradigm, e.g., in microscopy. Light hits the object under scrutiny (figuratively, the black box) and the *reflected* light is guided through a system of lenses and/or mirrors to form a magnified image for the observer to see. The difference between the so-reflected light and that which would have arrived at the observer's eye had the observed object not interfered is the image of the object contrasted with its background. As an amplification of our natural eye, a microscope is used to (quite literally) see into the structure of various material objects. In this, it is worth noting the important limitation. Standard optical microscopes cannot resolve structures finer than 10^{-6} m, regardless of the precision and perfection of the optical elements: lenses, mirrors, etc. The reason for this is the wave nature of visible light, with wavelengths in the range of about 380–760 nm. When considering an object that is smaller than that, light diffracts around it. The image is so fuzzy that no detail smaller than $\sim$380 nm can be discerned.

In perfect analogy, the sounds that humans normally hear easily circumnavigate objects of sizes smaller than about 17 mm. It is therefore humanly impossible to *hear* a marble that stands between us and the sound source. Humanly audible sounds have wavelengths within the 17 mm–17 m range, and all but the shortest wavelengths (which only a small number of people can hear well and which are also typically masked by sounds of longer wavelengths) easily circumnavigate objects of typical hand-held sizes. We say that the *resolution* is of the order of magnitude of the wavelength, understanding that only objects larger than the wavelength of the probing wave may be successfully *resolved*.

The alert Reader will, however, recall that *ultrasound* can be used to image objects of human size and smaller – and is routinely used to make a *sonogram* of, e.g., a fetus inside the womb. As a higher frequency corresponds to a shorter wavelength, the resolution of ultrasound is better, i.e., ultrasound may be used to image finer details than one can do with humanly audible sounds. Recall that the humanly visible light is but a tiny portion of the spectrum of electromagnetic waves. In full analogy, electromagnetic waves of a frequency higher than those in visible light (and so of shorter wavelengths) should produce finer resolution in appropriately constructed microscopes. Indeed, there are many types of electromagnetic waves with wavelengths that are shorter than those in visible light (ultraviolet light, X-rays, etc.), which could be used to construct stronger microscopes. In practice, however, the construction of such microscopes is hampered by the fact that very few if any materials can be used for lenses: ordinary optical lenses do not refract X-rays, the wavelengths of which are much smaller than those in visible light.

A solution is presented by the quantum nature of Nature: Material particles, such as electrons, also can behave like waves, and the basic relationship is that the wavelength of the probing beam

Table 1.3 Some "landmark" objects and events, and their characteristic sizes and the corresponding characteristic energies. Compare with Figure 1.4 on p. 12; $1\,\text{eV} \approx 1.6\times10^{-19}$ J.

Objects, events	Size	Energy (in eV)	
Crystalline lattice spacing	$\sim10^{-10}$ m	$\sim10^{3}$	(~1 keV)
Typical size of atoms[a]	$\sim10^{-10}$ m	$\sim10^{3}$	(~10 eV[a])
Typical size of atomic nuclei	$\sim10^{-15}$ m	$\sim10^{8}$	(~100 MeV)
Proton radius	$\sim10^{-16}$ m	$\sim10^{9}$	(~1 GeV)
Range of weak nuclear interaction	$\sim10^{-18}$ m	$\sim10^{11}$	(~100 GeV)
So-called "Grand unification"	$\sim10^{-31}$ m	$\sim10^{24}$	($\sim10^{15}$ GeV)
Quantum gravity, strings	$\sim10^{-35}$ m	$\sim10^{28}$	($\sim10^{19}$ GeV)

[a] The Bohr radius is $\alpha_e^{-1} \approx 137$ times smaller than the naive estimate $\sim\hbar c/E_C$ [☞ Section 2.4].

is *inversely proportional* to the energy of the probe. (Even a single electron can exhibit wave-like behavior, so that by a "beam" we mean herein one or arbitrarily many particles, as the case may be.) Table 1.3 lists a few objects and events in Nature, together with their characteristic size and corresponding energy; that is, the listed energies provide a minimum that a probe must have to resolve the details of the given size. Thus, to any "probe" (beam, ray, test-particle, radiation, etc.) with energy less than about 10 keV, typical atoms appear to be indivisible, structureless and featureless, point-like objects. Of course, a probe with (much) less energy would not even "see" an atom, but instead only the (much) larger structure comprised of atoms. To "see" the structure of the atom, one needs a probe with more than about 10 keV energy (per particle). This principle – that seeing ever smaller structures requires ever bigger energies – is the reason for the dual name of the game: "elementary particle physics" is rightfully also called "high energy physics."

With increased energy, the probability that the probe will change the scrutinized object (or at least some of its characteristics) also grows. What is observed is then not the exclusive property of the scrutinized object, but of the interacting object–probe system. This non-negligibility of the probe and its interaction with the scrutinized object is of essential importance and is a basic fact of quantum theory – and especially of atomic and sub-atomic systems. In this sense, testing and observing of a system irreversibly changes this system. This is sometimes expressed by saying that quantum observation – and so all empirical knowledge – is achieved with active participation of the observer. This causes an *essential* indeterminacy in all kinds of empirical exploration, and therefore also in all empirical knowledge. This quality is expressed in Heisenberg's "indeterminacy principle," which may be regarded as one of the fundamental principles of quantum theory.

The principle of indeterminacy is very precisely stated starting with quantities defined in classical (pre-quantum) theory. Again, quantum theory does not falsify but rather *extends* classical theory. To every degree of freedom and its corresponding variable (coordinate) that is used in the description of a physical system, classical theory corresponds a precisely defined *conjugate momentum*: let q and p denote such a pair. The indeterminacy relation then reads:

$$\triangle q\,\triangle p \;\geqslant\; \tfrac{1}{2}\hbar\,, \tag{1.42}$$

where $\triangle q$ and $\triangle p$ are the indeterminacies in observation and measurement of q and of p, respectively. Thus, if the position of a particle is measured to, say, 1.00×10^{-15} m precision, its (linear) momentum in the same direction cannot be measured better than to 0.525×10^{-19} kg m/s. The errors caused by apparatus imperfections are typically bigger than this, but there do exist measurements where this essential indeterminacy is detectable. To repeat: exact science always errs, but it knows precisely how much [☞ Conclusion 1.1 on p. 6]. Furthermore, the measurement of

https://doi.org/10.1017/9781009291507.003 Published online by Cambridge University Press

another quantity q', which is independent of q and p,[21] does not affect the measurements of q and p. That is, once we have measured q' with arbitrary resolution of the measuring instrument, the precision of the simultaneous measurements[22] of *either q or p* is limited only by the precision of the measuring instrument. For the general and precise statement, see Digression 2.7 on p. 73.

1.3.2 Breaking up is hard to do

The careful Reader may have questioned the persistent use of "structure" instead of "divisibility." The latter term is often taken to be a synonym of "being a composite," i.e., that the object "has/shows structure." This tacitly implies that a system that shows structure is in fact composite, and furthermore that it may somehow be divided into its constituent parts. Unfortunately, this is only a prejudice, borne out by everyday experiences: once broken, an egg can no longer be put back together whole; it may be possible to glue a broken glass goblet together with superglue, but the cracks remain, however fine.

In turn, something unusual happens when we divide something as teensy as an atom. Imagine ionizing a hydrogen atom, separating its nucleus (a single, positively charged proton) from its negatively charged electron. This may be accomplished, for example, by applying a sufficiently strong electrostatic field (with $\geqslant$13.6 eV potential energy). That proton and that electron can thus be moved away from each other arbitrarily many light years; at least in a *thought experiment* such as this, the rest of the universe may be ignored. Leave them so separated for some time, and... the electrostatic force will reunite them! Owing to the unbounded distances to which electrostatic forces reach, the electron and the proton that once formed an atom are never truly separated; their mutual interaction (via the electrostatic field) remains present through the whole "separation" experiment, so that this "separation" is quite fictitious.

This brings up another question. The forces that held the parts of the glass goblet together before it was broken in fact also reach to infinite distances. So, how is it that these forces do not reunite the parts of the broken goblet (however long the Reader is prepared to wait)? The answer is not only in the *distance* of the action, but also in the dependence of the force intensity on the distance. The intensity of the electrostatic force decays with the distance as $1/r^2$, while the intensity of "molecular forces" decays much faster. Imagine now testing the action of such a force at a distance of r, and assume, for simplicity, that the force field is spherically symmetric. That is, we observe the same action at every point of a radius-r sphere centered at the source of the force. As the surface area of the sphere grows as r^2, the *flux* (the product of the surface area and the electrostatic field) through the whole surface of the sphere remains unchanged. Through the same sphere, however, molecular forces that decay as $\sim 1/r^6$ (or faster) produce a flux that decays as $\sim 1/r^4$ (or faster) and so quickly fades at ever larger distances from the source.

> ***Conclusion 1.4*** *Molecular forces are said to be* ***localized*** *and have* ***finite range*** *(although the force need not in fact vanish at arbitrarily large distance). Coulomb-like forces that obey the "inverse square law" are said to have* ***infinite range****.*

So, the intensity of both molecular and Coulomb-like forces decays with distance. Distant particles interact weakly, whereas those near each other interact strongly. Thus, low-energy probes are deflected gently from their initial direction, while high-energy probes (with a small wavelength and fine resolution) are sometimes deflected at a large angle – as much as 180°! Precisely this correlation of the angular distribution and probe energy is the "hallmark" of Rutherford's experiments that confirmed the existence of a positively charged nucleus within the atom.

[21] The technical requirement is that q' *commutes* with both q and p; see Digression 2.7 on p. 73.

[22] In the context of quantum physics, "simultaneous measurements" do not mean "at the same time" – most often, this is trivially impossible. Instead, it implies successive measurement of two quantities, which is independent of the order in which the measurements are done.

https://doi.org/10.1017/9781009291507.003 Published online by Cambridge University Press

However, this is not so in collision experiments that are essentially the same as Rutherford's, but where the probes have >100 MeV energy [☞ Table 1.3 on p. 26]: These interactions differ significantly from Coulomb-like forces, and may be ascribed to so-called strong nuclear interactions. At distances where the action of these strong nuclear forces may be measured, their intensity *stagnates* with the distance between the colliding centers – as if the connection between those could be represented (modeled) by a string [☞ Chapter 11]! By itself, this may not seem unusual, but some of its consequences definitely are; see Chapter 6.

To stretch a string, one must invest work, and this increases the potential energy of the stretched string. At a certain point, determined by the string elasticity, it simply breaks. Analogously, two particles (so-called quarks) bound by the strong nuclear interaction may be separated to ever larger distances only by incessant investment of ever more energy. This could be done arbitrarily long, and the two quarks could be separated arbitrarily far from each other, except that the invested work sooner or later becomes sufficient to create a particle–antiparticle pair. Each one of these newly created particles then binds with one of the "old" ones, so that the attempt to separate two quarks to more than about 10^{-15} m fails. Instead of having separated one quark from the other, the quark we were trying to move becomes bound with the newly created antiquark, and the other "old" quark is joined by the newly minted quark replacing the old one. These quark–antiquark pairs form two new systems (so-called mesons) that really can be separated arbitrarily far, but the original quarks remain confined within these newly minted mesons [☞ Figure 1.5].

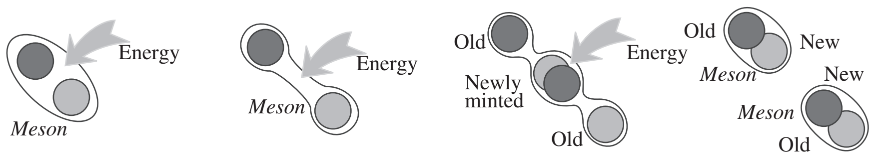

Figure 1.5 Inseparability of quarks and antiquarks in spite of investing ever more energy.

Thus, quarks (to most precise experimental verification and theoretical prediction) cannot be extracted arbitrarily far from one another, and remain "confined" – either in the original system, or in a newly minted system, joined with (anti)quarks created by investing ever more energy.

However, as long as the distance between the quarks is less than about 10^{-15} m, their binding energy is sufficiently small and they move effectively freely. Thus, the concept of "divisibility" (as it is usually understood) is definitely not synonymous with the concept of "compositeness," and those two notions must be clearly distinguished:

1. In all experiments performed to date, the electron behaves as a point-like particle, i.e., it shows no structure.
2. The proton shows structure (three quarks) through the complexity of the angular dependence in scattering, i.e., through deviations from Rutherford's formula – and does so when the collision energy surpasses a well-defined threshold; however, the quarks cannot be extracted arbitrarily far without creating new quark–antiquark pairs [☞ Figure 1.5].
3. Atomic nuclei show structure in collisions: they may be broken into smaller nuclei and/or their constituents, protons and neutrons (jointly called *nucleons*); the resulting parts may be permanently separated, i.e., the restoring forces have a limited, finite range.
4. Atoms show structure in collisions: they may be ionized by extracting one or more electrons; the restoring force between the positive ion and the extracted electron, however, has an infinite range, and the separation is not permanent: if left alone, the atom recombines.

https://doi.org/10.1017/9781009291507.003 Published online by Cambridge University Press

5. Molecules show structure in collisions: they may be broken (dissociated) into smaller molecules and/or atoms; the resulting parts may be permanently separated, i.e., the restoring forces have a finite range.

A remark is in order. Strong nuclear interactions that bind quarks into a proton are related to but not the same as the forces between the proton and the neutron (within atomic nuclei). The latter forces are called “residual,” just as some of the molecular forces between otherwise neutral atoms stem from electromagnetic interactions between somewhat separated constituent parts of these atoms (electrons and nuclei). However, even these residual forces are much stronger than the electromagnetic ones, as they overpower the Coulomb repulsion between positive protons within the nucleus. The weak nuclear interaction also has the characteristics of a string-like restoring force, but its characteristic short range and weakness stem from being mediated by the massive particles $W^{\pm}$ and Z^0. In turn, both the strong nuclear force and the electromagnetic force are mediated by massless particles, called gluons and photons, respectively.

1.3.3 . . . and smallest: limits of information

It would seem that the (spatial) resolution of measuring devices could, at least in principle, be made arbitrarily fine, but this is not the case. A glance at Figure 1.4 on p. 12 shows that something unusual should happen if the spatial resolution were to improve to the point of detecting details smaller than about 10^{-35} m, around the Planck length; see Ref. [518] and Ref. [99] for a recent discussion. Recall that, for detecting ever smaller details, the probe must have an ever larger energy, since the de Broglie wavelength of the probe is

$$\lambda_p = \frac{2\pi\hbar}{p_p} = \begin{cases} \frac{2\pi\hbar}{\sqrt{2m_p T_p}}, & \text{(non-relativistic)} \\ \frac{2\pi\hbar c}{T_p}, & \text{(relativistic)} \end{cases} \tag{1.43}$$

where

$$T_p = \begin{cases} \frac{p_p^2}{2m_p}, & \text{(non-relativistic)} \\ \sqrt{p_p^2 c^2 + m_p^2 c^4} - m_p c^2, & \text{(relativistic)} \end{cases} \tag{1.44}$$

is the kinetic energy of the probe of mass m_p [☞ Section 3.1.3].

During interaction with the “target,” the probe and the target temporarily form a combined system. The total mass of this combined system is greater than or equal to the sum of the target mass and the probe *mass-equivalent* of the total energy of the probe: $m_{t+p} = m_t + (m_p + T_p/c^2)$. Thus, as the kinetic energy of the probe grows, so does the total mass of the temporary target+probe system during the interaction.

Now, the situation becomes interesting owing to gravitational effects, and the fact that the gravitational field of the system grows (linearly) with the *total* mass of the system. Since the gravitational field grows, so does the “separation speed,” v_{sep}.[23] Furthermore, the gravitational field is not constant but grows unboundedly, as $1/r$ when $r \to 0$, so that the separation speed is much larger near the gravitating center than further away from it. If the target and the probe are both smaller than the distance between them, at the Schwarzschild distance between their centers,

$$r_S = \frac{2\,G_N\,m_{t+p}}{c^2}, \qquad \text{so that} \qquad v_{\text{sep}} := \sqrt{\frac{2G_N\,m_{t+p}}{r}} \xrightarrow{r \to r_S} c. \tag{1.45}$$

[23] This speed is often called the “escape velocity,” as it pertains to the successful launching of a projectile with a speed that suffices for the projectile to escape from the gravitational field of a (much larger) planet. The principle is, however, perfectly general, and applies equally to the separation of two objects after their collision.

That is, the separation speed $v_{\rm sep}$ becomes equal to the speed of light in vacuum, and no probe can separate from the target it has hit. This result holds (both the target and the probe are "sufficiently small") if neither the target nor the probe has any structure larger than the distance r_S as given in equation (1.45).

Alternatively, recall that a probe may be treated as a wave with the de Broglie wavelength $\lambda = 2\pi\hbar/p$, owing to the quantum nature of Nature. This wavelength is smallest for ultra-relativistic probes with $p_p \approx E_p/c$, so $\lambda_p \lesssim 2\pi\hbar c/E_p$. When the de Broglie wavelength becomes as small as the Schwarzschild radius, i.e., when the probe energy grows so that its resolution equals the Schwarzschild radius, the target "swallows" the probe and it cannot extract any information from within a sphere of radius r_S: the target+probe system now looks like a black hole:

$$\lambda_p \sim r_S \quad \Rightarrow \quad \frac{2\pi\hbar c}{E_p} \sim \frac{2\,G_N\,(m_t c^2 + E_p)}{c^4}, \tag{1.46a}$$

$$\Rightarrow \quad E_p \sim \tfrac{1}{2}\big[\sqrt{4\pi M_P^2 + m_t^2} - m_t\big]\,c^2, \tag{1.46b}$$

$$\xrightarrow{m_t\to 0} \quad E_p \sim \sqrt{\pi}\,M_P\,c^2, \tag{1.46c}$$

where we used that $M_P = \sqrt{\hbar c/G_N}$ [☞ Table 1.1 on p. 24]. The formal limit notation $m_t \to 0$ may also be obtained using the leading term in the expansion of the equation (1.46a) when $m_t c^2 \ll E_p$, or as the leading term in the expansion of the result (1.46b) when $m_t \ll M_P$. Indeed, for an ultra-relativistic probe, $E_p/c^2 \gg m_t, m_p$, and $p_p \approx E_p/c$. (In the non-relativistic limiting case, when $E_p/c^2 \ll m_t, m_p$ and $p_p \approx \sqrt{2m_p E_p}$, the probe has insufficient resolution to approach the target and test its structure.)

Of course, this argument extrapolates over many orders of magnitude in distances, and is based on current understanding of gravity and quantum mechanics. However, note that the *qualitative* part of the argument relies on the facts:

1. the minimal size of resolved details *decreases* (the resolution improves) with increasing probe energy,
2. the distance where the "separation speed" becomes inaccessibly big *increases* with the total mass of the source of gravity,
3. the mass (source of gravity) and the rest energy (measure of the ability to do work *not* owing to motion) are proportional to each other.

This already implies that there exists a minimal object (system) size or distance in Nature. If we can furthermore rely on the quantitative details of the argument, the minimal resolvable distance of about $\ell_P \sim 10^{-35}$ m follows unambiguously.

If Nature consists of elementary *particles* (which, by definition, have no constituents), then they must seem like miniature black holes. Their *event horizon*[24] must form a closed surface of which no detail smaller than $\sim 10^{-35}$ m can be resolved. This suggests that massless elementary particles would have to look like minimal, $\sim 10^{-35}$ m, spherically symmetric black holes. Massive particles could have a bigger horizon and a more complicated shape, but again the resolvable details must be bigger than about 10^{-35} m [☞ Digression 9.5 on p. 340].

> ***Conclusion 1.5*** *The unknowability of the "inside" of the event horizon of elementary particles indicates that there is no sense in regarding them as ideal point-like objects. Willy-nilly, elementary particles are extended in space.*

[24] The term "event horizon" denotes the border (bordering surface) in space that fully encloses a black hole, and from within which nothing can come out because of confiningly strong gravity. It is *believed* that all naturally occurring black holes are so wrapped by an event horizon; see Comment 9.4 on p. 337.

This conflict between (**1**) the extrapolated results of the general theory of relativity (which describes gravity) for *point-like* elementary particles, and (**2**) quantum mechanics (whereby the points cannot really be smaller than about 10^{-35} m in diameter) is of course a prediction of such a combined model, and a result of the model itself. To avoid the conflict, we must leave behind some aspects of this model of gravitating point-like quantum elementary particles, but retain enough of it so as to continue reproducing its experimentally verified properties, for energies $\lesssim 10^2$ GeV, i.e., for distances $\gtrsim 10^{-18}$ m.

There is another (also intuitive and not formally rigorous) argument that indicates the incompatibility of the general theory of relativity and the quantum theory of point-like particles: Heisenberg's principle of indeterminacy implies that the position and the linear momentum in the same direction cannot both be determined with infinite precision. On the other hand, in the general theory of relativity, the presence of matter curves spacetime and so defines a class of coordinate systems: a massive point-like particle curves spacetime in which the position of this particle is determined with perfect precision. Furthermore, this particle is at perfect rest in this class of coordinate systems, so that both the position and the linear momentum are determined with infinite precision. This is in direct and unavoidable contradiction with quantum theory.

— ❦ —

In turn, Conclusion 1.5 has a very important consequence: The variables with which we represent physical objects depend on the variables with which we represent spacetime and the abstract space of various other properties of this physical object. Thereby, for example, the wave-function of an electron is a function of spacetime coordinates and also of numbers that determine its mass, charge, spin, chirality… Recall that every function is simply a rule that assigns to every value of its arguments – a point in the domain space – a value of its own. This value is represented as a point in the target space, i.e., the range of the function. Since spacetime coordinates of any physical system cannot be specified *within* the event horizon of that system but only up to the *surface* of this horizon, it follows that the domain spaces of the functions with which we describe physical systems are not volumes that permeate the open sets of spacetime, but the *surfaces* that enclose such volumes. This insight may be argued to engender the *holography principle*, which was introduced into the fundamental physics of elementary particles by Gerardus 't Hooft and which was in the 1990s gradually built into the (super)string theory and its *M*-theoretical extension [☞ Chapter 11] (first by Leonard Susskind, and then in his collaboration with Tom Banks, Willy Fischler, and Stephen Shenker).

1.3.4 Unification: the smaller, the more similar

Technically, the incompatibility between the general theory of relativity and quantum theory for point-like elementary particles introduces unavoidable divergences: When computing physically measurable quantities, prediction results are obtained in the form of hopelessly divergent (indefinite) mathematical expressions. By contrast, in the amalgamation of the *special* theory of relativity and quantum theory, in "relativistic field theory," formally divergent results may be removed through the process of "renormalization." The quantum theories of electromagnetic, weak, and strong nuclear interactions (together with all known matter) really do include the special theory of relativity and form a logically consistent structure.[25] Developed akin to quantum electrodynamics (the predictions of which are confirmed to an amazing 12 significant figures [293, 1]), the quantum field theory of electroweak interaction describes the observed electromagnetic and weak

[25] In fact, the existence of the "top" quark (which was only recently convincingly confirmed in experiments) was predicted using so-called anomaly cancellations. That is, without the "top" quark, the Standard Model of elementary particles and interactions would be self-contradictory!

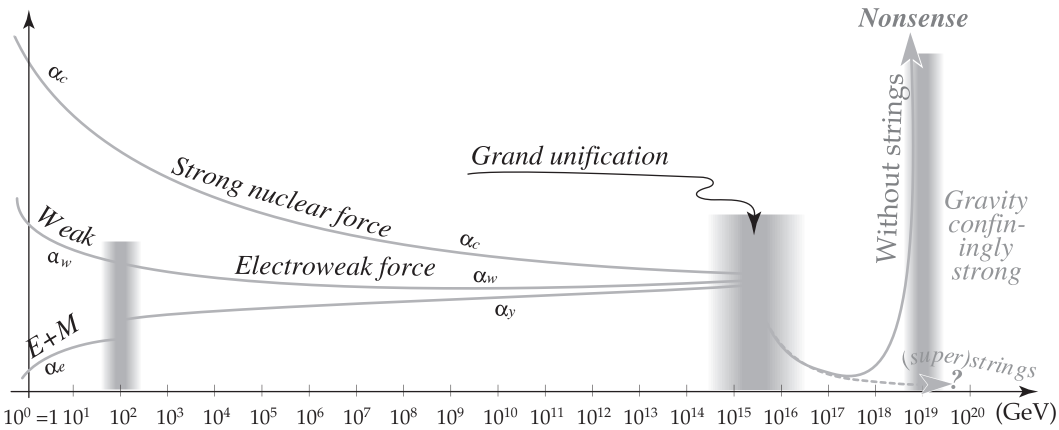

Figure 1.6 A logarithmic plot of coupling parameters, from 1 GeV (≈ proton's rest energy, $\approx 10^5$ times bigger than the hydrogen atom ionization energy), to 10^{19} GeV, where gravity becomes confiningly strong and point-like theories become nonsense, while string theories "pass" through a phase transition of sorts, albeit insufficiently known so far.

nuclear interactions as well as their unification very precisely. Both the theoretical and the experimental precision in both the electroweak and the strong interaction model are considerably more humble, but the results do agree. Several models of quantum field theory have been developed that describe the unification of the electroweak and the strong interaction [☞ Figure 1.6], but only further experiments, starting with the LHC facility at CERN, can decide which of these unifying models – if any – describes the "real World."

In all of these cases, the transition regimes (the shaded areas in Figure 1.6) where a unification happens indicate a phase transition of sorts, in the sense that the qualitative properties of the theory are drastically different on one and the other "side" of the transition region. While these may well be related to (World-scale) bulk-material phase transitions that have presumably happened in the early universe, the subject of particle physics probes the related physics phenomena in individual particle collision experiments performed at high energies – where no actual bulk-material phase transition occurs. For example, below about 10^2 GeV, there is a clear distinction between electromagnetic and weak nuclear processes, each of which can happen without the other. At energies above about 10^2 GeV, however, these processes mix inseparably. This situation is very similar to the fact that electric and magnetic phenomena are well distinguished in stationary systems and often occur one without the other, but become inseparably united and involve electromagnetic waves when the electric charges move and create non-stationary currents.

In the case of electro-magnetic unification, whether or not electric and magnetic fields are distinguishable depends on the speeds, taken in ratio with the speed of light in vacuum ($c \approx 299{,}792$ km/s). For example, it is well known that an electrical current (flow of electric charges) creates a magnetic field around it. The speed of the individual charged particles is typically small as compared to the speed of light in vacuum. However, the speed of *momentum transfer* within that current is very close to the speed of light in vacuum. In turn, the electromagnetic field itself adapts to the motion of electric charges – of course – at the speed of light. When the electric charges either do not move at all or form stationary currents, the ratio of the speed of changes in this current and the speed of light in vacuum (which is the reference parameter here) is zero, and the electric and the magnetic fields are well distinguishable. When the electric charges move so they do not form stationary currents, this ratio is (near or equal to) one, and the electromagnetic field can no longer be separated into an independent electric and an independent

https://doi.org/10.1017/9781009291507.003 Published online by Cambridge University Press

magnetic field. Furthermore, non-stationary currents cause a variable magnetic field, the changes in which create an additional electric field, the changes in which modify the magnetic field, etc. This feedback between the electric and the magnetic field for non-stationary currents produces the new phenomenon: electromagnetic waves, which carry away some of the energy carried by the current.

For the electro-weak unification ("electro-magneto-weak" would be more accurate, but is too much of a mouthful), the unifying parameter is the ratio of energies of the processes as compared to the $W^{\pm}$ and Z^0 particle masses. (Just a few years after their discovery at CERN, these particles were routinely observed and studied; their masses are close to 10^2 GeV/c^2.) By contrast, the mass of the particles of light is zero; the total energy of light is entirely of kinetic nature and there is no coordinate system in which light is at rest. Clearly, when an experiment is conducted at energies much below about 10^2 GeV, real $W^{\pm}$ and Z^0 particles cannot be produced and cannot contribute to the processes we study. Weak nuclear processes then happen by exchanging *virtual*[26] $W^{\pm}$ and Z^0 particles [☞ Definition 3.5 on p. 104]. This happens owing to the Heisenberg principle of indeterminacy and the fact (Pauli, 1933 [29, p. 334]) that the *energy* and the *characteristic duration of time* of every process also satisfy the indeterminacy relations:

$$\triangle E \,\triangle\tau \;\geqslant\; \tfrac{1}{2}\hbar\,. \tag{1.47}$$

Roughly, during the time $\triangle\tau \leqslant \frac{\hbar}{2\triangle E}$ for $\triangle E \sim 10^2$ GeV, a $W^{\pm}$ or a Z^0 particle may be freely produced – if it also decays within this time. The necessity that *two* consecutive processes (creation and decay) must happen in such a short time decreases the probability of the *joint* process mediated by an intermediate $W^{\pm}$- or Z^0-boson, and this permits an unambiguous identification of the said process as weak nuclear, rather than electromagnetic. However, when the energies in the experiment become much bigger than 10^2 GeV, *real* $W^{\pm}$ and Z^0 particles are produced with the same probability as the electromagnetic waves (photons). Owing to charge conservation, W^+- and W^--radiation do not mix with the others, but the Z^0-radiation and electromagnetic radiation do mix, inextricably, and form new kinds of phenomena – very similar to the unification of (variable and mutually inducing) electric and magnetic fields [☞ comparative Table 8.1 on p. 299].

A similar novel phenomenon is expected around $\geqslant 10^{15}$ GeV, where the electroweak and the strong interactions tend towards having the same strength. The entire graph in Figure 1.6 on p. 32, is, however, based on experimental data at currently available energies, $\leqslant 10^2$ GeV, and so is necessarily an extrapolation. The assumption that neither new phenomena nor new particles will be found between $\sim 10^2$ GeV and $\sim 10^{15}$ GeV is often called the "grand desert hypothesis." This follows from Ockham's principle, whereby novelties are introduced only if necessary. The subsequent ideas and arguments rely on this at least in part, and must be re-examined as soon as there is compelling evidence that the "grand desert" turns out to be populated. Several so-called Grand-Unified Theory (GUT) models have been developed attempting to sort through the possible phenomena that could occur in this region, but in ways that leave the physics below $\sim 10^2$ GeV unchanged, in agreement with the experiments performed so far. Such (and all other) models are expected to predict at least some event that could soon be experimentally verified (or refuted).

— ❦ —

A few quick and perhaps overdue remarks: Just *what* exactly do these weak and strong nuclear interactions in fact *do*? Besides esoteric phenomena of particle physics, these interactions are in fact responsible for our own existence! 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 that produces the immense energy of our Sun is *nuclear fusion*, in which the nuclei of

[26] In distinction from real particles, the virtual ones – by definition – cannot be directly observed.

https://doi.org/10.1017/9781009291507.003 Published online by Cambridge University Press

deuterium and *tritium* (two heavier isotopes of hydrogen) fuse into helium and release a neutron and energy. The reason that there is surplus energy is due to the details of strong nuclear interactions. Finally, note that nuclei of pure hydrogen would not fuse; instead, deuterium and tritium are needed. The hydrogen nucleus is a single proton; deuterium and tritium nuclei consist of a proton and respectively one and two neutrons – all held together by strong nuclear forces. These much needed neutrons are all being produced in weak nuclear interactions such as the inverse β-decay, $p^+ + e^- \to n^0 + \nu$. These can – and indeed do – occur within stars [☞ equations (7.118)], and also occurred in the young universe, long before the stars were formed. In addition to providing the required fuel for (strong nuclear) fusion, weak nuclear interactions also moderate this process, thereby preventing our Sun from burning out in one brilliant explosion.

> ***Conclusion 1.6*** *Thus, by making the Sun burn in the first place, by 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, respectively, bring about the conditions on the Earth that sustain 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 stable orbit near the Sun. Were it not for these four interactions, dear Reader, you would not exist to read this book.*

1.3.5 A shift in understanding

The relativity of Nature prevents us from thinking of space and time as two disparate "things," and forces us to join them into a single, undivided spacetime. The concept of simultaneity is recognized to be relative, which then disperses the so-called paradoxes of twins/clocks, of the ladder and the barn, of the ruler and the hole in the table, etc.

The quantumness of Nature disillusions us from thinking of "things" around us as unchanged, and clearly separable from their environment, and forces us to think of them as determined by imposed circumstances. Thus, the electron may behave *both* as a point-like particle and as a wave – and is in fact neither, but "something" that in appropriate circumstances *may look* like a particle or like a wave. Similarly, instead of talking about entangled states of two separate objects in EPR (Einstein–Podolsky–Rosen)-type experiments, it would be wiser to talk about a single, undivided state of a single system, which in certain circumstances may be interpreted as two spatially separated sub-systems.

Besides, the quantumness of Nature indicates the importance of the Hilbert space as a very real space in which processes occur – although neither the Hilbert space nor the unfolding of events in it can ever be seen by unaided human eye or mind. By contrast, although the very spacetime itself is just as "invisible," we do see the unfolding of events in spacetime. This makes thinking about events unfolding in spacetime intuitively easier, while thinking of events unfolding in the Hilbert space appears to be much less natural and very counter-intuitive. That makes the quantumness of Nature baffling.

The combination of (special-)relativistic and quantum physics is then *doubly* baffling and counter-intuitive, but is no less rigorous as a scientific discipline than the familiar and intuitive classical mechanics. In fact, the most precise agreement between theoretical and experimental physics occurs exactly in the realm of quantum field theory. For some of the characteristic and observable quantities such as the fine structure constant and the so-called magnetic moment anomaly, the comparison of various measurements and theoretical computations in quantum electrodynamics agrees with experimental data to an all-time record of *12 significant figures* [1, 293]!

The fundamental physics and its description of Nature – which must include both quantum physics and general (not only special) relativity – are then even more baffling as compared to commonplace experiences. Onward then, into this mutliply baffling journey!

1.3.6 Exercises for Section 1.3

✎ **1.3.1** *Express the value of Newton's gravitational constant,* $G_N = 6.6742 \times 10^{-11} \frac{m^2}{kg\,s}$*, in "particle units,"* $\hbar^x c^y (\text{MeV})^z$*, for some* (x, y, z)*.*

✎ **1.3.2** *Using the definitions in Table 1.1 on p. 24, compute the value in Planck units (suitable powers of* ℓ_P*,* M_P *and* t_P*) of the SI units: (**a**) 1 m/s (speed), (**b**) 1 m/s*^{2} *(acceleration), (**c**) 1 N s (linear momentum), (**d**) 1 J s (angular momentum), (**e**) 1 J (work and energy), and (f) 1 W (power).*

✎ **1.3.3** *Using the data in Table 1.2 on p. 25, compute the* MeV ↔ SI *conversion factors for (**a**) speed, (**b**) angular speed, (**c**) acceleration, (**d**) angular acceleration, (**e**) linear momentum, (f) angular momentum, (**g**) work, and (**h**) power.*

✎ **1.3.4** *Using the definitions in Table 1.1 on p. 24, compute the value in Planck units (suitable powers of* ℓ_P*,* M_P *and* t_P*) of the SI units: (**a**) 1 C (charge), (**b**) 1 N/C (electric field), (**c**) 1 T (magnetic field), (**d**) 1 A (electric current), (**e**) 1 V (voltage), and (f) 1 K (temperature).*

✎ **1.3.5** *Using Table 1.2 on p. 25, and the result (1.12),*[27] *compute the* MeV ↔ SI *conversion factors for (**a**) charge, (**b**) electric field (from* $\vec{F}_C = q\vec{E}$*), (**c**) magnetic field (*$\vec{F}_M = q\,\vec{v} \times \vec{B}$*), (**d**) electric current (*$I := dq/dt$*), (**e**) voltage, a.k.a. potential (from* $P = VI$*), and (f) temperature (= average kinetic energy/*k_B*).*

✎ **1.3.6** *Using that* $M_P = \sqrt{\hbar c / G_N}$*, verify that the leading term in the expansion of the result (1.46b) when* $m_t \ll M_P$ *agrees with the solution of the leading term in the expansion of the equation (1.46a) when* $m_t c^2 \ll E_p$, i.e.*, with the result (1.46c).*

✎ **1.3.7** *Considering that* $M_P = 2.18 \times 10^{-8}$ *kg does not seem very large in everyday terms, obtain the leading term in the expansion of the result (1.46b) when* $m_t \gg M_P$*, and compute the corresponding range of values for* E_p*. How feasible is it to provide an elementary particle probe with the lowest such energy?*

[27] Recall that the speed of light $c = 1/\sqrt{\epsilon_0 \mu_0}$, and that the Boltzmann constant $k_B = 1.38 \times 10^{-23}$ J/K is a conversion factor between units of energy and units of temperature, which statistical physics defines as average translational kinetic energy of molecules in some large ensemble, while other forms of energy of the molecules (vibrational, rotational, binding. . .) contribute to the so-called internal energy.