Looking at the Universe through the Eyes of Physics, Cosmology and Philosophy

Looking at the Universe through the Eyes of Physics, Cosmology and Philosophy

By

Angelo Tartaglia

Cambridge Scholars Publishing

Looking at the Universe through the Eyes of Physics, Cosmology and Philosophy

By Angelo Tartaglia

This book first published 2023

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN (10): 1-5275-1956-2
ISBN (13): 978-1-5275-1956-5

TABLE OF CONTENTS

INTRODUCTION

Cosmology, in a broad sense, is an extremely fascinating branch of knowledge and, like most physics in the last century, has also made enormous progress. Our present capabilities of observation were unimaginable a hundred years ago and even our interpretative tools have greatly improved. We may now expect to be capable, in the not-too-distant future, of sewing together, in a single consistent framework, the knowledge we have acquired in different fields of physics, and with that, build a comprehensive description and interpretation of the vast phenomenology we see in the visible universe.

However, the further one proceeds, the more contradictory elements emerge in the provisional framework that is being built; indeed, it is precisely our great ability to collect information from observation and experimentation that brings inconsistencies to the fore.

The main objective of this text is not to summarize the state of the art in understanding the universe and the foundations of physics: for that there is a vast plethora of excellent literature. The goal is rather to look for contradictions, weaknesses and their structural characteristics, and the "clay feet" of a giant that displays strength and splendour in every other part, while continuing to grow.

Who is this book for? For anyone who already possesses the bases of scientific culture, but I wish I could just say for anyone. In fact, at least part of the open questions that will emerge border on philosophy and we must not forget that today's physics is the heir of natural philosophy of the past. However, it is true that over time and with the refinement of the methods of observation/measurement and of the interpretative tools, the technical complexity of the problems has also grown and a specialized language has been developed, indeed, a set of specialized "dialects" proper to each sub-branch of knowledge. All of this constitutes, for most, a barrier that is difficult to overcome to get to the essence of the knowledge we would like to discuss. It is well known of language that, the more it becomes initiatory, the more it creates a protective fence against the incursions of others that could potentially, under the mantle of awe-inspiring formulas, discover the weaknesses of this or that community of initiates.

On the other hand, it would have been unthinkable, assuming I would have even been capable of doing so, to begin drafting a sort of universal treatise starting from the basic rudiments of physics, to arrive only at the end, after a long (and boring) journey, to the interesting questions. This being the case, I tried to compromise, which, as often happens, risks being unsatisfactory for everyone. I tried to write with a reader in mind who has basic scientific knowledge, sufficient not to be uncomfortable with the (mathematical) language in which the physical theories are formulated; or who is able to find elsewhere an explanation of those tools he/she possibly lacks.

At the same time, I tried to re-establish the consolidated foundations of the two great paradigms of twentieth century physics: (general) relativity and quantum mechanics. Also because the contradiction between these two paradigms is one of the cognitive problems on which to reflect. Having said all this, I went looking for doubts rather than certainties, trusting in the fact that doubt is the engine of research and that results often come from humility rather than from triumphalism. Paraphrasing Socrates, as narrated by Plato, I can remember that the wisest man is the one who knows he does not know and the worst ignorant is he who believes he knows.

The greatest doubt we will arrive at, for those who have the patience to follow this path to the end, will be whether or not our reason is a sufficient and adequate tool to understand the universe, even in its physical aspects.

The reader will judge if and to what extent I managed to get closer to the stated objectives and if this work has been helpful.

CHAPTER I

EXCURSUS

The old times: myths and legends

Since the dawn of human thought, we can find traces of reflections on the world or on the universe, as we say today. These distant echoes reach us at the end of a long chain of transmission that attains the written, or at least depicted, form in the second millennium before Christ. What interests us here is the search for the traces of intuitions that, with extremely different languages and formalisms, have reached the contemporary age and are in fact incorporated into scientific knowledge or at least into the current scientific debate.

The thinking humans of the early days were struck by the observation of what was happening around them, in the seemingly unlimited variety and mutability of forms of the earth's environment and in the apparent regularity and stability of the sky above meteorological phenomena. The experience of oneself was that of an animated being and the diversity of behaviours arose from our being animated; therefore, most natural phenomena whose cause was not understood were read as someone's behaviour, and consequently, the whole world was dominated and determined by a large number of animated agents far above ordinary humans and, in this sense, divine. Wind, clouds, rains, springs, sea, mountains, etc., everything lives and is not understandable; at most it can be ingratiated through special rites. This approach can be dealt with by psychoanalysis, sociology, ethnology and other human sciences. Upstream, however, albeit not formulated in the language of later philosophical thought, a few questions hover: Where does all this world come from, be it understandable or not? Why and how did it take this form?

A common trait of all the main cosmogonies of ancient times is that of assuming, in an indefinite time in the past, the existence of a sort of primordial matter, whose connotation is that of being undifferentiated. For the Greek tradition, it was the primeval chaos; ancient Chinese culture called it *qi*; even the Egyptian tradition spoke of a primordial ocean; and in the Bible, we speak generically of *waters*. Mostly, in short, we start from something that exists and that is undifferentiated and has within itself the roots of everything that will come later. Beyond the imaginative language of the myth, this idea of a shapeless primordial ingredient has a curious resonance with at least one of the modern theories concerning the very first moments of the evolution of the universe: that of cosmic inflation. According to this theory, originally proposed by Alan Guth (1981) and Andrei Linde (1983), at a certain stage of its existence the universe was permeated by a scalar field called *inflaton*, initially in a state of *false vacuum*. These terms, which we will return to in due course, can turn out to be rather obscure to those who are not involved in theoretical physics. For the moment, let us be content to say that this ingredient, the name of which is certainly much funnier than those used in ancient cosmogonies, was in a condition of very high energy density in which all future fundamental interactions (gravity, weak force, electromagnetism, strong force) were indistinguishable and in which there were no excitations (i.e. particles) typical of the field: in short, precisely an undifferentiated state in which there is already everything that will be there later, but nothing is distinguishable due to the very high energy density. The expansion of space and quantum fluctuations will then give rise to an incredibly rapid transition (with a duration of the order of 10^{-34} seconds) which, expanding space in an inflationary way[1] (hence the name of the field), will separate the four fundamental forces and will cascade all the particles that populate the universe.

In the ancient cosmogonies, at a certain point, in an indefinite time, something happens and the first divine entities "are born" and in turn generate others. A separate case is represented by the Bible in which God is not part of the universe, but, with an act of will, creates everything from nothing. Although in the Genesis, echoing something of the Mesopotamian cosmogony, we speak of the spirit of God that "hovers over the waters", clearly God does not spring from something and what comes after is not indicated as drawn from the "waters". In short, the *nothing* from which creation takes place is precisely nothing; it is not the "vacuum" of quantum mechanics, which is actually something.

Returning to the myths of the origins of almost all ancient civilizations, after the first indistinct and unexplained events, the subsequent vicissitudes take a varied course, often with gory connotations (especially in Mesopotamian legends, but to some degree also in the Egyptian, Greek, Nordic, Cen-

---

[1] That is, with an exponential law over time.

tral American traditions, as well as others): divine parents cruel to their children, divine children who "kill" (also divine) fathers or mothers and perhaps dismember them, and so on. In fact, ordinary humans are projected into the sky with their contrasts and connotations. These stories have sometimes been analysed from the point of view of psychoanalysis; however, in some cases you can read something that refers to intuitions that today are found in the domain of physics, after having passed through that of philosophy.

Both in ancient Greek thought and in Indian thought, there is the idea of a progressive decay of the world. In Hesiod's narration, five ages have taken place: golden, silver, bronze, heroic, and finally the present iron age. Each age, in various respects, was worse than the previous one. Archaic Indian thought describes a cyclic universe, in which each cycle is composed of four ages or *yugas*, each worse than the previous one, up to the final one, the *kali yuga*, followed by the destruction of the world, before the start of a new cycle or *kalpa*. Other ancient traditions (of the peoples of the Americas and of Nordic traditions) also contemplate a progressive decay with some end of the world, with or without subsequent rebirth. Could we see in all this a representation of the “arrow of time” and perhaps of the second law of thermodynamics?

It is not necessary to insist, and vice versa, it is appropriate to get closer to what we know and think of the universe today. I shall not generally follow a historical path, but I will try to give precedence to facts over interpretations, even though, as we shall see, at this scale and on this matter it is not so easy to separate facts from interpretations. I will endeavour to make limited use of mathematical formalism, which in some cases is useful to synthesize concepts effectively, but in some others can also obscure them.

Looking at the sky, century after century

The material universe is everything that surrounds us or our body is made of, but usually, when we talk about the universe, it is understood that we are essentially referring to a large scale. This is what we will do here too, focusing our attention on dimensions that are certainly higher than those of the earth, but mostly also above those of the solar system: let's say that we will begin by looking up to the sky.

In the past, this is what humanity literally did. In times when life expectancy at birth was below forty years and writing was not yet available, there were people who, for both ritual and practical reasons, systematically observed the sky and passed on their observations orally to their apprentices.

In this way, when writing came, a fairly large body of knowledge about celestial phenomena emerged. While the terrestrial environment is characterized by mutability and irregularity, the sky, above the clouds, is a place of regular and repetitive phenomena, from the alternation of day/night to the cycle of seasons. If, in the most naive and superficial vision the earth appears to be flat (and generally surrounded by waters) and the sky is an upside down cup (in the Chinese tradition the earth is "square" and the sky is "round"), when they put all of their scientific knowledge together, and in particular their observations from long journeys by sea and by land about the configuration of the sky, humans realized that the earth must be spherical. This is what Greek philosophy affirmed, starting from the sixth century BC, and the science of the time not only developed deductive conjectures, but also included direct measurements. Eratosthenes, in the 4th century BC calculated the diameter of the earth, starting from the observation of the different lengths of the shadow of a vertical stick on the same day at the same solar time at two different latitudes:[2] the value he obtained was correct within 10−15%.

A spherical earth, and an equally spherical sky above, carried with them the idea that the earth was the centre of the universe, even though, once again, Greek thought, in the person of the astronomer Aristarchus of Samos, who lived at the turn of the 4th and the 3rd centuries BC and considered the hypothesis that the sun and the stars were still and that the earth revolved around the sun as well as on itself. Aristarchus' theory explained, in a simple way, the apparent motion of the planets on the celestial vault and, assuming that the earth's rotation axis was inclined with respect to that of the orbital rotation, explained the seasons; in order to have a coherent construction, it was also necessary to assume that the distance of the stars from us was much greater than that between the earth and the sun.

Aristarchus' heliocentrism was present in ancient thought parallel to the geocentric vision, but it was the latter that prevailed after being systematized by Claudius Ptolemy, of Alexandria, in the 2nd century AD. Ptolemy introduced a series of mechanisms that were able to justify the apparent motion of the planets, safeguarding the centrality and immobility of the earth. The Ptolemaic vision dominated the late ancient era and the whole of the Middle Ages, ending up being also incorporated into the philosophical-theological

[2] The day was the summer solstice; the time was noon. The two locations were Alexandria in Egypt, on the Mediterranean coast, and Siene, much further south along the Nile.

theories of the time. Heliocentrism reappeared in the 15th century with Copernicus, who only at the end of his life decided to publish his *De revolutionibus orbium coelestium*, so much so that by the time he received the first copy of his text he was on his deathbed and no longer aware of what was happening around him.

As we know, the end of the Middle Ages and the beginning of the Renaissance coincided with important transformations in European society and culture. So it happened that, even with respect to the transition from Ptolemaic geocentrism to Copernican heliocentrism, politics and power, as they were practiced, represented and justified at that time, got in the way. Galileo paid the price, without stopping the Copernican revolution why by this point in time it had become unrestrainable. It is also worth remembering that at the end of the 16th century, the former monk Giordano Bruno argued, not on the basis of observational data but for philosophical reasons, that the universe could have no centre (not even the sun) and that there were infinite worlds, more or less like ours, scattered in the immensity. The universe also had to be eternal. These ideas are expressed in particular in *De l'infinito, universo e mondi*, published in London in 1584, and are part of the reflections on religion and the divine. Giordano Bruno had it much worse than Galileo, because he was burned at the stake as a heretic.

Until the 17th century, the human eye was the essential tool for the observation of the sky, supported at most by quadrants, astrolabes, armillary spheres, capable of identifying and measuring angles between different directions of observation. With these technologies a total of a few thousand stars are visible and five planets can be distinguished, in addition to the major bodies; occasionally, transient phenomena can be observed in the form of comets, or perhaps *stellae novae*, which suddenly appear in a position that then remains fixed, disappearing after days or even weeks. To this category belonged the *nova* (today we qualify it as a *supernova*) which appeared on May 1st 1006 and was observed both in the East and in the West; the same goes for the two (super) novae of 1572 and 1604, observed by Tycho Brahé, in Denmark, and by Kepler, in Prague.

From the 17th century, however, an instrument based on the properties of lenses, known since ancient times, made its entry into observational astronomy: the telescope. The object was invented in Holland, but was perfected and made famous by Galileo, who made extensive and systematic use of it for his observations of celestial bodies. The astronomical telescope led to the revision of a number of ideas about the sky and celestial bodies. Galileo discovered that mountains and plains can be seen on the moon, from which it could be deduced that at least that celestial body was not so different from the earth. In the following centuries, some even went so far as to exaggerate the supposed similarity between earth and moon: mistaking the plains for seas and imagining possible inhabitants of our satellite. Galileo also discovered that four "moons" (baptized *Cosmica sidera* in honour of Cosimo II de' Medici)[3] revolved around the planet Jupiter, which once again generalized the idea of a universe made up of major bodies around which minor bodies rotate. Another discovery made by Galileo using his telescope was the sun's own rotation, detected thanks to the observation of "spots" (a fair surprise at the time) on the surface of our star.

A fundamental step forward for the interpretation of the new celestial observations was taken by Newton with his theory of universal gravitation which, in one fell swoop, made it possible to describe, starting from a single formula, the connections of celestial bodies to each other, allowing the laws of motion of the planets formulated by Kepler on the basis of observations to be derived.

Leaving aside the details of the subsequent advances both in the description of the sky and in the formalization of physical laws, it is convenient to arrive at the idea of the universe prevailing in the second part of the 19th century, at a time when electromagnetism had also found its unitary formulation thanks to the synthesis carried out by Maxwell in 1865.[4] The belief was that the universe was an unlimited (infinite?) expanse of stars scattered everywhere; each star probably with its retinue of planets. No centre and no privileged time, whence probably an eternal duration. There were a series of open problems that ultimately (physical) science did not care too much about, being all focused on the technological dimension and on the local scale.

A problem was the supposed uniformity in the distribution of stars in the universe, consistent with the assumption of not having to have privileged positions. Observation seemed to contradict this hypothesis, with a significant part of the sky occupied by the Milky Way, which telescopes had clearly indicated to be made up of stars, probably further away from us than the others observed individually in the foreground.

---

[3] Today the known satellites of Jupiter are 67.

[4] This is at least the date on which the book "A Dynamical Theory of the Electromagnetic Field" was printed even though Maxwell had already published his equations in writings of 1861/62.

Olbers' paradox

Another problem was condensed in the so-called *Olbers' paradox*.[5] This paradox stems from the apparently silly question: why is the sky dark at night? The naive answer is that the sun is not visible at night, while the stars are scattered and far away. However, if we imagine that the universe is infinite and full of stars, in whatever direction we look, sooner or later, our line of sight will end on the surface of a star, though far away and consequently seen as a very small dot. That dot would in any case have the typical brightness of a point on the surface of a star and if there are stars everywhere, more or less close or more or less distant, the celestial vault should always be as bright as the surface of the sun. A way out could be to forgo the eternity of the universe (but not its infinite extension in space): if we imagine a beginning from which the stars "light up", light from the most distant sources may simply not have had time to reach us yet. If so, the brightness of the sky should progressively increase, asymptotically tending to equal that of a stellar surface.

Another explanation of the paradox could be based on the consideration that the heavens cannot actually be completely transparent. In addition to stars, the universe also contains non-luminous bodies, such as planets, satellites, asteroids and so on down to meteorites and then dust. The farther you go, the greater the number of these bodies interposed along the line of sight: the incoming light from the most distant sources is progressively absorbed on the way. However, this explanation again conflicts with the hypothesis of the eternity of the universe through thermodynamics. Absorbing light implies an increase in the temperature of the absorbing body, until the latter becomes able to re-emit in the form of radiation all the energy it is receiving. In short, the interstellar dust,[6] in an infinite time, would reach thermal equilibrium, which means that it would have the same temperature as the absorbed radiation, i.e. the same surface temperature of the emitting stars: once again the sky would be as bright as the sun.

Thermal death

In the 19th century, classical thermodynamics was also formalized and completed (and at the end of the century it was the turn of statistical me-

[5] Heinrich Olbers (who was a doctor with a passion for astronomy) formulated it in 1823, but the idea had been around in one way or another for a couple of centuries.
[6] *Dust* here includes all small or large opaque bodies.

chanics). Among the general laws of thermodynamics, there is the fundamental *second principle*, the simplest and most intuitive formulation of which says that heat passes spontaneously from hotter to colder bodies and not vice versa.

This law also conflicts with the idea of an eternal and essentially stationary universe because it is clear that over time stars can only cool down by heating the opaque bodies that are scattered around in the surrounding space and these in turn radiate energy into the "empty" space. Asymptotically, the universe evolves towards a state of equilibrium in which each body is at the same temperature as the black body radiation that fills space. In the absence of residual temperature differences, everything stops and we arrive at what has been called *thermal death*.

On the other hand, by going backwards over time we find conditions of increasing imbalance with very hot emitters and very cold receptors. Again we stumble upon some "beginning" which is followed, sooner or later, by an "end". Of course, the final state can be reached asymptotically in an infinite time and, without specifying the physical mechanisms or discussing whether or not they are possible, we can also think that the beginning is placed in an infinitely distant time in the past and at an infinite temperature. The fact remains, though, that the universe cannot always be more or less the same and that subsequent eras are progressively different from each other.

From a mathematical point of view, a conformal transformation of the time coordinate is sufficient to convert the $+\infty$ and $-\infty$ values into two finite values. We can schematize what happens using the difference in temperature between warmer and colder bodies, ΔT,[7] as a benchmark: it should be a decreasing monotone function of time t. We may also assume that it is:

$$\begin{cases} lim_{t\to-\infty}\Delta T = \infty \\ lim_{t\to+\infty}\Delta T = 0 \end{cases} \qquad \text{Eq. (1.1)}$$

At this point, we may change the "evolution parameter" (time t) by moving to the new variable τ, with a transformation like:

$$t = \frac{1}{b-\tau} - \frac{1}{\tau-a} \qquad \text{Eq. (1.2)}$$

[7] An average is implied between the temperatures of all the emitters (the stars) and another between those of the receptors (the opaque bodies).

The evolution of the system begins in $\tau = a$ and ends in $\tau = b$.

Instant propagation of the gravitational interaction

A problem that had already bothered Newton concerns the law of universal gravitation, recognized as the large-scale binder of the cosmos. The acceleration of gravity at any point, $\vec{g}$, depends on the distribution of the masses all around according to the law:

$$\vec{g} = -G \int_{Sources} \rho \frac{\hat{u}_r}{r^2} dV \qquad \text{Eq. (1.3)}$$

G is Newton's constant.[8] The mass density (as a function of r) is represented by ρ; the product of the mass density by the volume element dV gives a mass element contributing to the gravitational acceleration in the origin of the reference frame being used. The distance between a given element of mass and the origin is r and $\hat{u}_r$ is a unit vector oriented from the origin towards the mass element.

The sum in Eq. (1.3) is in principle extended to the whole universe. According to the formula, if somewhere the mass distribution undergoes a change over time, this immediately affects the place where the observer is. In other words, Newtonian universal gravitation propagates at infinite speed and without any intermediary. In short, gravity would give rise to the typical action at a distance, which in itself recalls more magic than physics. Newton's own words, written in a letter to Bentley, dated 1692/3 are significant in this regard:

> *It is inconceivable that inanimate Matter should, without the Mediation of something else, which is not material, operate upon, and affect other matter without mutual Contact ... That Gravity should be innate, inherent and essential to Matter, so that one body may act upon another at a distance thro' a Vacuum, without the Mediation of any thing else, by and through which their Action and Force may be conveyed from one to another, is to me so great an Absurdity that I believe no Man who has in philosophical Matters a competent Faculty of thinking can ever fall into it. Gravity must be caused by an Agent acting constantly according to certain laws; but whether this Agent be material or immaterial, I have left to the Consideration of my readers.*

---

[8] Its value is $(6.67384 \pm 0.00080) \times 10^{-11}$ $m^3/(kg \times s^2)$.

The luminiferous aether

The problem of the intermediary actually arose also in the case of electromagnetism, the other fundamental interaction known at the end of the 19th century. The experiments had shown that light, unlike what Newton thought, had to be a wave and in fact the complete formalization of electrical and magnetic phenomena, obtained by Maxwell, had given substance to the concept of an electromagnetic field and had shown that waves propagate in this field and that light is precisely an electromagnetic wave.

The concept of a wave is familiar to everyone, now as it was at that time, but a wave must be of *something*: it needs a medium in which vibrations can propagate. At first glance, the light that comes from celestial bodies also travels in a vacuum and therefore it follows that the *vacuum* cannot really be empty.

In the absence of possible known intermediaries, in the 19th century[9] it was decided to call the medium that was to permeate the skies in order to allow the propagation of light, (luminiferous) *aether*. The name was borrowed from Aristotle who used it to designate the fifth fundamental element of nature, which represented the substance of which heavens were made.

Name aside, the problems began when trying to identify the physical properties of this aether. Electromagnetic waves turn out to be *transverse*[10] and only solids are capable of propagating waves of that type. The speed of light calculated from Maxwell's equations and from the value of the typical constants of electromagnetism is very high and experiments confirm this. Now, it is known (and it was known in the 18th century) that the speed of the waves in a medium is higher the more rigid the medium, from which it would follow that aether is solid and very rigid. On the other hand, all celestial bodies are immersed in the aether as well as all objects that surround us and between which light travels: apparently this extremely rigid aether does not place any obstacles to the relative motion of the bodies.

As if that weren’t enough, the experiments attempted to measure the speed of the earth’s motion with respect to the aether (in particular by Albert

---

[9] Actually, the idea had been considered even earlier, for example by René Descartes in the 17th century.

[10] The electric field and the magnetic field of a wave are perpendicular to each other and oscillate perpendicular to the direction of propagation.

Michelson in 1881 and again, together with Edward Morley, in 1887) did not yield any results.

The times are ripe

Concluding this chapter, we can say that, at the end of the 19th century, the time was ripe for a new paradigm in which to insert the known phenomena in order to overcome the uncertainties and inconsistencies accumulated up to then.

CHAPTER II

A NEW PARADIGM

The initial situation

The open questions in the physics of the late nineteenth and early twentieth centuries were, at the time, the subject of a debate and of various attempts at a solution, but the main lines of experimental and, above all, technological research did not seem to care too much. The debate seemed to be limited to the restricted world of theoretical physicists, albeit with some connection with measurement: the experiments of Michelson and Morley (later repeated by others as well), for example.

Actually, a real paradigm shift was necessary to deal with the inconsistencies. The new paradigm was developed and proposed by a young man who graduated from the *Eidgenössische Technische Hochschule* (ETH) in Zurich and, at the time of his first revolutionary studies, was employed in the patent office in Bern. I am of course talking about Albert Einstein, who published, in 1905, the article *Zur Elektrodynamik bewegter Körper*,[11] in which the foundations of the new theory of *special relativity*, as it was later called, were expounded.

The purpose of this book is not to explain the foundations of relativity or of modern physics in general, so I will not follow the historical stages of the theory's evolution nor reproduce the traditional explanatory path. Our interest is directed to the cosmological implications, so I will focus on the new entry of Einstein's theory: space-time.

Space and time before Einstein

The description of the physical world built over the centuries up to the early 20th century was based on two conceptual pillars, assumed, in their abstract properties, as obvious: space, within which to place the objects to

---

[11] On the electrodynamics of moving bodies.

be described, and time, to mark and describe, in a formal and mathematical way, change and in particular movement.

That these two entities, both abstract and terribly concrete, did not correspond to trivial concepts is demonstrated by the entire history of philosophy and in particular of natural philosophy, the antecedent of modern physics.

Space

Space has always been considered as a kind of empty container featuring mathematical and geometric properties. In philosophical terms, we can say that an essential feature of space is extension. In this sense, space is, from a logical point of view, a set of contiguous "positions", each of which can be associated with a set of three real numbers[12] chosen so as to be unique for each distinct position. The "positions" in geometry correspond to *points* without extension. The continuity of space is expressed by the fact that two arbitrarily close points are characterized by triplets of numbers which in turn differ by arbitrarily small quantities.

Without prejudice to what has been written, the choice of the specific triplet of numbers to be associated with a given point remains arbitrary. Having made a choice and identified a criterion for extending it, without contradictions and with continuity, to the whole space (or to the particular domain to be described), a *coordinate system* has in fact been built. Without prejudice to the logical constraints, the choice of coordinates remains arbitrary and will be dictated by criteria of practicality and convenience in relation to the problem faced: in any case, we start from a point to be associated with the triad (0,0,0) and which will be the origin O of the reference frame.[13]

A very common and very simple choice is that of orthogonal Cartesian coordinates: we consider three one-dimensional subspaces that intersect at the origin. Each subspace has as its own (unique) coordinate a real number

---

[12] Three, because the dimensions perceived in ordinary space are precisely three; but geometry knows generalizations in which the dimensions necessary to locate a position in some abstract space can be $N > 3$, in which case the logic can be the same, but we are dealing not with triples but with N-tuples. Similarly, it is possible to think and describe spaces which in addition to having N dimensions use N-tuples of complex numbers (or their generalizations).

[13] Strictly speaking, "reference frame" and "coordinate system" are not the same thing, in the sense that in general the former relies on some physical benchmark, while the latter can be totally disjoint from real objects.

that varies from $-\infty$ to $+\infty$, assuming the zero value at the origin. In three dimensions, the points of each subset have one coordinate coinciding with the one just defined and the other two zero. The sequence of coordinates is also chosen once and for all, so that in each subset the non-zero coordinate is in a different position from that in the other two.

Coming to a more familiar language, let's say that the “one-dimensional subspaces” are three mutually perpendicular lines[14] representing the axes on each of which we read a different coordinate; the symbols traditionally used for orthogonal Cartesian coordinates are x, y and z. Given a point, its projection on the axes identifies the corresponding coordinates.

Everything written above is valid regardless of the presence or absence of material objects and it is this autonomy with respect to matter that creates the problem. Space is empty in itself; but what does *empty* mean? Aristotle vigorously rejected the idea that void could "exist", considering it a contradiction in terms, and this in controversy with the atomist school of his time (or of the immediately preceding centuries), for which everything that exists is formed by *atoms* immersed in the void, within which they move here and there. For Aristotle, if we say that between two objects there is "nothing" then it means that the two objects are in contact; in short, extension is a property of bodies, not of nothing, and the void cannot therefore be *nothing*. Hence the aether as a *fifth essence*[15] that permeates every space.

More than 1900 years later, that is still essentially Descartes' position. However, in formalizing the laws of mechanics and introducing its universal gravitation, Newton returns to think of space as something existing for itself, regardless of matter, and endowed with *a priori* geometric properties, independent of the observer's choices. The idea is well expressed in point two of the *scholium* at the beginning (page 5 of the first edition) of the *Philosophiae naturalis principia mathematica* (Isaac Newton, 1687):[16]

---

[14] In reality, they do not need to be orthogonal to each other. It is sufficient that the three lines are not coplanar.

[15] After the four traditional elements: air, water, earth and fire.

[16] «*Spatium absolutum natura sua absque relatione ad externum quodvis semper manet similare et immobile; relativum est spatii hujus mensura seu dimensio quaelibet mobilis, quae a sensibus nostris per situm suum ad corpora definitur, et a vulgo pro spatio immobili usurpatur: uti dimensio spatii subterranei, aerei vel coelesti definita per situm suum ad Terram. Idem sunt spatium absolutum et relativum, specie et magnitudine, sed non permanent idem semper numero. Nam si Terra, verbi gratia, movetur, spatium Aeris nostri quod relative et respectu Terrae*

> Absolute space, by its nature and regardless of the relationship with any external thing, always remains similar and immobile; the relative one is a measure or any mobile dimension of this, which is defined by our senses in relation to its arrangement with respect to bodies, and is arbitrarily exchanged by the vulgar with immobile space: as the dimension of an underground, air or celestial space defined by its location with respect to the Earth. Regarding species and size, absolute and relative space are the same, but they do not remain the same in number. In fact, if, for example, the Earth moves, the space of our Air that always remains the same with respect to the Earth, will now be a part of the absolute space in which the air passes, now another, and so it will change continuously.

Time

With regard to time, the situation is even more complicated than for space. Time has always been considered a measure of change and, in particular, of movement. However, its nature remains somewhat obscure and at a glance paradoxical. Distinguishing past, present and future, we can say that time measures the transition from a state that does not exist (any longer) to one that does not exist (yet), passing through one of zero duration: since what does not exist is certainly not an object in scientific research, nor in philosophy, what can time possibly be?

The Greek sophists extensively applied themselves to the problems of time as duration and its measurability, even managing to “demonstrate” that motion and change do not exist. A famous example is that of Zeno's paradox which shows how, in a foot race, the swift-footed Achilles will never reach a tortoise that has started before him!

The confusing situation is well expressed by a sentence written at the end of the 4th century AD in the Confessions of Saint Augustine (quoted in Buckner, 2006):

> "What then is time? If no one asks me, I know. If I wanted to explain it to someone who asks me, I don't know!"[17]

With all this, it is clear that the concept of time is fundamental to describe the events of the physical world. It is again Newton who tries to give

---

semper manet idem, nunc erit una pars spatii absoluti in quam Aer transit, nunc alia pars ejus, et sic absolute mutabitur perpetuo.»

[17] «*Quid est ergo tempus? Si nemo a me quaerat, scio. Si quaerenti explicare velim, nescio.*»

a definition of absolute time, at point one of the aforementioned scholium (Isaac Newton, 1687):

> Absolute time, true and mathematical, in itself and by its nature without relation to anything external, flows uniformly and, with another name, is called duration. The relative, apparent and vulgar is some sensitive and external measure (accurate or not) of duration, obtained through movement. The vulgar uses this in place of true time; like the hour, the day, the month, the year.

There is no answer to what time is, but it is something that flows. Strictly speaking, the definition is a vicious circle because the idea of flow implies time: how can we therefore say that time "flows" since flow itself is in time?

Be that as it may, Newtonian definitions of space and time remained the basis, implicit or explicit, of all physical theories until 1905.

Space-time

Physics deals with phenomena that must be observed and subjected to measurement by someone who is conventionally called an *observer*. Among all possible observers, a somewhat privileged class can be identified: these are *inertial* observers. By definition, they are not subject to the action of any force and therefore are not accelerated in any way. In short, the infinite possible inertial observers can at most be in uniform rectilinear motion with respect to each other. Underneath this there is Newton's absolute space: it is with respect to it that uniform rectilinear motion can be defined.

If we isolate two inertial observers in motion, one with respect to the other, at the uniform speed V, we can assume that each uses a reference system whose origin corresponds to its position. The two origins identify a line that both can use, for example, as the x axis. If both observers consider the same object, they can identify its position with respect to themselves by attributing to it coordinates in the reference frame at the origin of which they are located.

Limiting ourselves to the line joining the two, what for the observer O has abscissa x will have for O' abscissa $x' = x - Vt$, due to the relative motion; if then the object is in turn moving along the x axis with the velocity v, seen from O, for O' the velocity will be $v' = v - V$. We have described a *Galilean transformation* of coordinates with the corresponding law of composition of velocities. The implication, of course, is that time is the same (to put it in

Newton's words: it flows the same way) for all observers; in short, time is "obviously" absolute.

When we analyse dynamic phenomena that imply the presence of forces, we will find ourselves dealing with accelerations, that is, the second derivatives of space with respect to time. Since Galilean transformations are linear in time, we find that all inertial observers "see" the same accelerations, therefore the same forces, despite their relative motion. In other words, we expect that physical laws are the same for all inertial observers, that is, that they are invariant under Galilean transformations of the coordinates: this is the so-called *Galilean relativity*.

Alas, however, Maxwell's equations, which describe electromagnetism so well, are not invariant under Galilean transformations. Faced with this problem, young Einstein hypothesized that the speed of light was not like the others, which depend on the observer (according to the paradigm of that time: on the absolute motion of the observer with respect to the aether), but was exactly the same for all the inertial reference frames, regardless of their relative motion with respect to each other. In fact, the speed of light c was transformed, from a quantity to be measured from time to time, into a universal constant, as it is now formally defined by assigning it the exact value c=299,792,458 m/s; from the value of c we then proceed to define the units of measurement of times and lengths.

The assumption made by Einstein obviously solves the problem of the null result of Michelson's and Morley's experiments, but it also requires that the coordinate transformations upon passing from one inertial reference frame to another are no longer the Galilean ones but those of Lorentz, which furthermore also have the virtue of leaving Maxwell's equations unchanged.

Limiting ourselves to the case of relative motion of two inertial observers along the common x axis, the Lorentz transformations are written as:

$$\begin{cases} t' = \gamma\left(t - \frac{V}{c^2}x\right) \\ x' = \gamma(x - Vt) \\ y' = y \\ z' = z \end{cases} \tag{2.1}$$

where

$$\gamma = \frac{1}{\sqrt{1-\frac{V^2}{c^2}}} \tag{2.2}$$

is the Lorentz factor.

Hendrik Lorentz had already published the transformations that go under his name in 1904, noting that they leave Maxwell's equations unchanged. Indeed, even earlier, in 1897, Joseph Larmor had found the same transformations when studying the problem of Maxwell's equations.[18] However, in the absence of a physical interpretative framework, Eqs. (2.1) appeared as a kind of mathematical curiosity without any not purely instrumental justification. All the more so since Lorentz transformations presented a series of consequences that were rather perplexing in terms of their physical meaning.

To begin with, time also appeared to depend on the observer, and then a series of strange effects emerged, such as the contraction of moving objects in the direction of motion regardless of the material they were made of.

Einstein's approach was to assume that the invariance of the speed of light was a physical fact and that, consequently, Lorentz transformations were natural and Galilean ones were inaccurate. In fact, the transformation law of velocities that can be deduced from (2.1) is:

$$v' = \frac{v-V}{1-\frac{vV}{c^2}} \tag{2.3}$$

It is immediately verifiable that if $v = c$, v' also equals c, regardless of the value of V.

At this point, everything was fine in electromagnetism, but on the other hand, it was necessary to modify the equations of classical mechanics. In particular, it was necessary to admit that the ability of a body to resist the action of a driving force, that is, its inert mass, had to grow with the speed of the body with respect to the observer, according to the law

$$m = \gamma m_0 = \frac{m_0}{\sqrt{1-\frac{V^2}{c^2}}} \tag{2.4}$$

[18] It was Enry Poincaré who introduced the name *Lorentz transformations,* universally used since.

in which m_0 is the *rest mass* of the object, i.e. the mass that it presents in a reference frame in which it is at rest.

Then there was still an equivalence between energy, W, and mass, according to the famous formula:

$$W = mc^2 \tag{2.5}$$

The reason why, generally, the need to take these new elements into account was not felt was because the V^2/c^2 ratio in the Lorentz factor is usually of the order of 10^{-16}. It does therefore not produce perceptible corrections, except when we're dealing with relative speeds of the order of that of light.

We will not retrace the stages of the affirmation of the theory of relativity, but I want to recall some of the consequences seen as paradoxical and certainly counterintuitive, in addition to the contraction of the lengths of moving objects.

According to the new theory, an observer who keeps an eye on what happens in a moving system will find that time is "dilated" there: everything happens more slowly. Even more surprisingly, the simultaneity between two events depends on the observer: what happens for one at the same instant, for others happens in succession and the sequence can occur in a certain order or the reverse, for different observers.

All these quirks, while continuing to disrupt our day-to-day intuition, find a consistent representation (and, if you like, an explanation) if we resort to a geometric approach regarding space and time, as did Herman Minkowski, former professor of Einstein at the ETH.

Until then, the description of physical phenomena had required, separately, space and time, both being "absolute". In the new relativistic paradigm, instead, the background, or the "container", of physical phenomena was a single four-dimensional continuum: *space-time*, which, as the name implies, brought together space and time. We could also say that now space and time, separately, become relative, while space-time is absolute.

The new four-dimensional continuum, like traditional space, is characterized by its geometric properties. Newtonian space was associated with Euclidean geometry; space-time, with an extra dimension, is associated with a geometry that is like that of Euclid except for the invariance constraint of the speed of light: it is Minkowski's geometry.

In practical terms, once a Cartesian coordinate system in four dimensions has been chosen, the distance *ds* between two arbitrarily close events[19] can be defined so that it is:[20]

$$ds^2 = c^2dt^2 - dx^2 - dy^2 - dz^2 \qquad (2.6)$$

It looks like Pythagoras' theorem, except for the signs. In fact, we see that, if the interval[21] is the one between two events along a light ray, we find $ds^2 = 0$; if vice versa the two events are along the trajectory of an object (which travels at a speed lower than c) then it is $ds^2 > 0$ and we speak of a *time-like* interval; if the two points are along a line that cannot be followed by any physical object, since it would involve a speed greater than c, then, despite the square, we have $ds^2 < 0$ and we speak of a *space-like* interval. The quantity expressed by (2.6) is also called the square of the line element of Minkowski's space-time.

Like distances in ordinary space, space-time intervals are what they are regardless of the choice of the coordinate system and of the observer. However, if, in ordinary space, we project a given segment onto the axes, the projections (the *components*, if we use vector language) depend on the choice of the axes of the reference system. We can say that the distances (or the moduli of vectors) are absolute (they do not depend on us and our choices), while the components are relative. The same happens in Minkowskian space. In fact, the Lorentz transformations, which allow the conversion from the reference frame of an inertial observer to that of another, are "rotations" in space-time; however, due to the signature of (2.6), the rotation angle turns out to be imaginary and equal to $\alpha = i\chi$ with $\chi = \text{arcosh}\ (1/\gamma)$.

The intervals are therefore objective (i.e. independent from us), but their breaking down into spatial and temporal components changes with the observer: therefore, the contraction of lengths and dilation of times find their explanation as corresponding to the result of the projection on the axes in use, without in any way affecting the physical quantity from which they come. The relativity of simultaneity is also reduced to a problem of projec-

---

19 This is the name given to the points in space-time.

20 The choice of the *signature* is conventional, so much so that in the scientific literature one finds both the formula in the text and the alternative form $ds^2 = -c^2dt^2 + dx^2 + dy^2 + dz^2$. Provided all formal steps are made correctly, nothing changes in the final physical conclusions.

21 This is the term used in space-time instead of *distance*.