

# Counterfactual thinking in physics

Miko Elwenspoek

FRIAS, University of Freiburg, Albertstrasse 19, 79104 Freiburg, Germany  
And MESA+ Institute for Nanotechnology, University of Twente, P.O. Box 217,  
7500 AE Enschede, The Netherlands

Email: [m.elwenspoek@el.utwente.nl](mailto:m.elwenspoek@el.utwente.nl)

Counterfactual thinking plays a key role in research in physics and, I believe, in research in all natural sciences. In this contribution I will describe a few examples of counterfactual thinking, how it is used, the power of this method of inquiry, and the types of results that can be achieved.

A brief account of the way physicists carry out research will be given and three main types of questions will be identified. Two of them will be used to illustrate the value of counterfactual thinking, one example regarding astronomy, and the other example dealing with electromagnetic forces. The latter might be quite tough for non-physicists. The last, most extended, part of this paper gives an analysis of counterfactual situations: what would the world look like if the constants of nature had different values. This discussion leads to the conclusion that slight changes of these numbers leads to uninhabitable worlds.

## *How physicists work*

Before I discuss counterfactual thinking in physics, I want to describe what it is that physicists actually do. Physicists have a method of inquiry that puts them, in a sense, into a most comfortable position. They have equations that can be used to examine a situation and predict what would happen in this situation. The results of these examinations are constantly compared to what can be seen in the real world. This is most often done in experiments, but there are many cases when the experiments are impossible using current technology, impossible in principle, or undesirable. The general method in physics and in other, so-called, exact sciences is to develop a *quantitative* theory (that's why these sciences are called "exact") and to compare the predictions of the theory with what can be seen in the real world. The power of this method lies in the possibility of quantitative comparison: the theory does not only predict that – as an example – there is light, but how much, its color, its polarization, its speed, the interrelation between speed, wavelength and frequency, its interaction with matter and so on. Comparisons can be done very precisely, and if the results of the theory do not match the observation then something is wrong: The theory, the observation, or both. Theory and observation go hand in hand. Observations give rise to changes of a theory, to further refinements of it, or to its abandonment. Moreover, the theory helps to design experiments, to interpret the observations, and it calls for new observations.

There are three exciting situations for physicists.

- i. When a critical test of a theory is possible. The theory grows from a large number of experiments and observations and makes predictions about previously unknown facts (if the theory does not predict new facts it just describes what has

been observed and is not very interesting). In the next section there is an example from a comparison of Kepler's empirical laws and Newton's theory of gravitation.

- ii. When the critical tests give a result that shows that the theory is wrong, that is, when all possible errors from the tests have been examined and eliminated but a discrepancy remains between the theoretical prediction and the observational results. This situation is encountered constantly in physics. The real world is so complex that in nearly all situations the theory cannot be used directly because the complexity makes exact calculations impossible, so one has to rely on approximations. In most situations it is not clear beforehand what the best route and the most appropriate approximation is.
- iii. When one detects internal inconsistencies within a theory. Below I'll describe one of the most famous examples in the history of physics.

The interesting point here is that counterfactual thinking is a key method to tackle the problems in these situations. In fact, it is usually not really known in physics what is counterfactual and what is not.

#### *Counterfactual thinking as a method of inquiry*

Generally, scientists try to make sense of the world and use everything to be as precise and critical as possible. They collect all available data and try to find rules that fit *all* data. To do this they work with experiments, measurements, mathematical theories, computer simulations. Counterfactual thinking is an indispensable tool for the analysis. Basically it consists of confronting consequences of our ideas with observations, extending our ideas over the domain where observations technically are possible with current technology, and searching for inconsistencies of various aspects in the whole set of our ideas. To set the stage, counterfactual thinking of several types can be defined here as pondering a problem in physics by asking the "what if" question in the sense that the "if" refers to a statement believed to be incorrect, outside the range where ideas are thought to be applicable or outside the range accessible for observation.

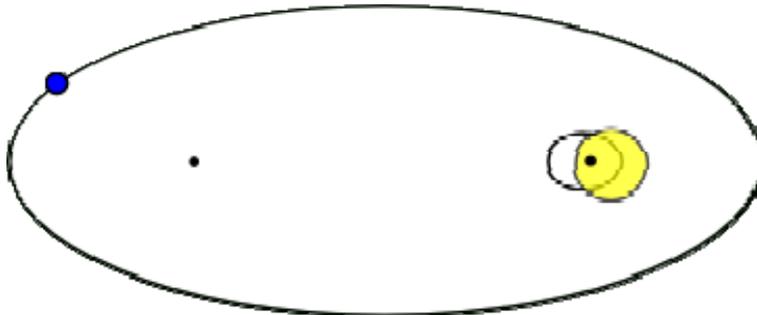
A meteorologist once said to me that he likes to stop the revolution of the earth just to see how the climate would change. This is physically impossible, but a simulation on a computer, or the examination of the equations that describe the climate is possible, although quite challenging. Of course meteorologists do the simulations, which refer to a counterfactual situation - the earth *is* spinning - and they are very curious as to whether their theories (e.g. no cyclones on a non-spinning world) give reasonable results and they want to test the predictions. In this example we see that counterfactual thinking can, in principle, be used to test theories. The theory is thought to be valid generally. For a certain situation, when a certain aspect of the situation is changed, the theory accurately predicts the results of the change: one can test this prediction in an experiment.

In a sense, the experiment itself is counterfactual as in an experiment one prepares a world that does not naturally exist. When Galileo performed his experiments on free fall, he tried to eliminate all forces (except the force we now call gravity) acting on a mass which was dropped from some height. He was lucky because the drag of air becomes appreciable only at rather high speeds and for small masses. Galileo's experiments studied the free fall of heavy bodies from a small height. He inferred that *all* masses at *any* speed are accelerated by gravity in the same way: when masses fall from a given height, they all attain exactly the same speed and the fall takes exactly the same time for all masses. His claim was that this would happen if only gravity were at work on

the mass in question. To really do the experiment one has to remove the air around the mass to eliminate friction. Galileo could not do this, but now we can: we can prepare an artificial world to perform the experiment. We also see in this example that one can discover a rule (all masses *tested* attain the same acceleration) and extend the rule to situations where the test cannot be performed. The formulation of the rule extrapolates beyond what can be seen in practice – the “factual”, the rule itself in his strict formulation is literally counterfactual.

### *Kepler's laws and Newton's theory of gravitation*

A famous example for the first situation in the above list is Newton's theory of gravitation<sup>ii</sup>, shown in Figure 1. Newton was able to reproduce Kepler's observations of the motion of the planets. Kepler found that planets move along ellipses around the sun, with the sun located in one of the two foci of the ellipse. Newton's theory made a similar prediction: planets do revolve along an ellipse but so does the sun. A well-defined mathematical point, the center of mass - not the sun - is at the focus. Because of the enormous difference in masses between the sun and the planets, the sun is very close to the center of mass. It was quite difficult to confirm this particular prediction, but we know now that it – and many other predictions which neither Kepler or Newton thought of - is correct. The precision of the measurement of the position of stars is now such that the slight wobble of those stars having planets can be detected. This way several hundred so-called exoplanets have been found. The aspect of counterfactual thinking lies in this example in the extension of the theory of gravitation over the range where observations were technologically possible in Newton's time.



**Figure 1.** The orbits of Earth (top left, blue/dark) and Sun (right, yellow/light) along ellipses (the large one is the earth's, the small one is the sun's orbit) with a common center of mass (the black dot inside the sun). The orbit of the Sun is greatly exaggerated in this graph. Ellipses have two focal points, shown as black dots, only one of them is the center of mass (right). An ellipse can be constructed by attaching the ends of a rope at the two focal points and stretching the rope to make an angle; the ellipse is at all points that can be reached by the tip of this angle. A circle results if the two focal points coincide into one single point.

### *Maxwell's theory of light*

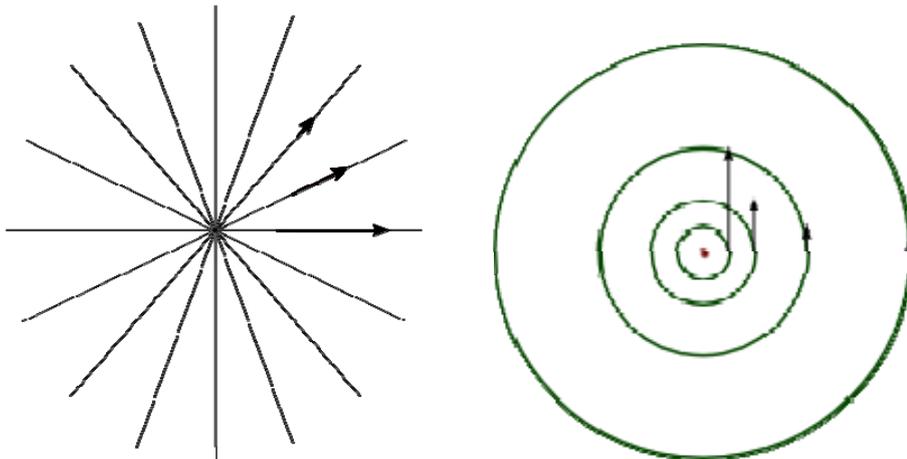
Sometimes counterfactual thinking leads to more than a test of a theory, namely to the discovery of elements of a new theory. We are held together by, so-called, electromagnetic forces. These forces are described by a set of equations which James Clerk Maxwell<sup>iii</sup> wrote down for the first time, which is why they are named after him. In fact, Maxwell added only one little piece to equations that had been found earlier, but this piece turned out to be essential. To get the flavor of the story some explanation of these equations is necessary although it is not necessary to actually write them down.

All matter is the source of gravitational forces and all matter feels these forces. Visible matter additionally is the source of a second fundamental force. Visible matter is composed of atoms which have a very compact nucleus surrounded by electrons. Electrons repel each other as do nuclei, but nuclei and electrons attract each other. We say, quite arbitrarily, that electrons carry a negative, and nuclei a positive charge<sup>iv</sup>.

Charges cause two types of forces to which they also react. There is a force between two charges that works in a straight line connecting the charges. When the charges are at rest this force is called the electrostatic force. A *moving* charge additionally causes magnetic forces, which in turn act only on *moving* charges. In general both forces are present, and what is seen as the magnetic and what is seen as the electric force depends on the observer. The phenomena of electric and magnetic forces are so interlinked that the whole subject area is called *electromagnetism*.

Mathematically these forces are described by an abstract concept called a field. In Figure 2 two types of fields are shown. One can envision a field as something which changes the properties of space everywhere: If there is an electromagnetic field at some point in space then a charge at this point would feel a force. A familiar field is the gravitational field at the surface of the earth. I expect that a mass would fall down whenever I let it go. To visualize a field one draws lines parallel to the direction of the field.

In principle there are two kinds of such lines: those that begin and end somewhere, and those that do not begin or end anywhere<sup>v</sup>. The latter ones form closed loops.



**Figure 2.** Representation of two types of fields. Left: a field with field lines starting at a source (a mass or a charge). Right: a field having lines without an end. In this case, a current through a straight wire coming out of the page (in the center) is surrounded by a magnetic field similar to the one shown. The arrows represent the field strength at the points where they start; their length represents the strength, and the orientation shows the direction of the field. The lines are everywhere parallel to the field and the density of the lines is related to the field strength.

Maxwell's equations describe these aspects of the electric and magnetic fields in relation to the charges and their velocities. There is one aspect of a force field in which a field starts at a charge and ends at a charge. This is quite analogous to a machine gun: bullets start at the gun and end where they hit. Loosely speaking, an electric field made by a charge (at rest or moving) starts at a positive charge and ends at a negative charge. (The field shown at the left hand side of Figure 2 is that of the positive spherical charge

in its vicinity, all other charges are very far away.) A magnetic field is very much different. It is generated by a moving charge and the field starts and ends nowhere; the field forms closed lines. A wire through which a current flows makes a magnetic field that winds in circles around the wire, see right hand side of Figure 2.

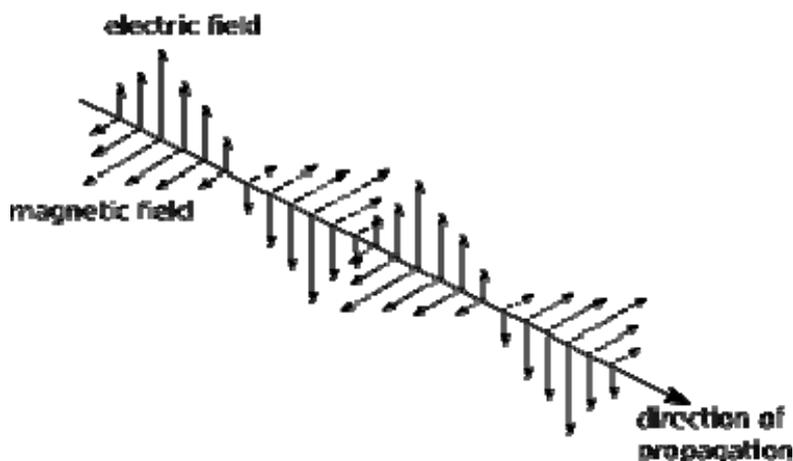
However, electric fields can also be generated by changing the value or direction of a magnetic field. This happens in generators, where a magnet is forced to rotate (e.g. by connecting the magnet to the wheel of a bicycle). The resulting electric fields are similar to magnetic fields: they have no beginning and no end. In some cases they form closed circular loops, so, in order to let the lamp of the bicycle glow, we wrap a wire around the rotating magnet, the charges in the wire are forced by the electric field which is generated by the changing magnetic field, a current flows and the lamp glows.

Maxwell discovered that, in order to get a consistent theory, a changing electric field must also generate a magnetic field in the same way that a changing magnetic field generates an electric field. Inconsistency pops up when the magnetic field is calculated in a situation where a current carrying wire is interrupted. Then there are two ways to calculate the magnetic field, which give different results.

So the equations published in 1864 look something like this:

- Ending electric field lines are connected to charges
- Not ending electric field lines are connected to changing magnetic fields
- Ending magnetic field lines do not exist
- Not ending magnetic field lines are connected to moving charges *and to changing electric fields*

Maxwell's addition is printed in italic. This little addition has enormous consequences. Most importantly, it predicts the existence of light, and it described all the properties of light that had been detected at the time that Maxwell was working. When combined with quantum mechanics, it predicts properties of light that had not yet been detected at the time that Maxwell was working. All modern inventions made long after Maxwell's death, including radio, TV, wireless mobile phones, optical telecommunication, are based on these four equations. In all cases where a precision experiment is possible, the predictions hold true. For example, theory says that the speed of light is independent of the speed of both the source and the receiver and, in vacuum, the speed of light is independent of its color. These predictions turn out to be true with an experimental uncertainty of 1 in one hundred billion billion ( $10^{20}$ )!<sup>vi</sup> In Figure 3 we show a graphic representation of an electromagnetic wave.



**Figure 3.** Representation of an electromagnetic wave. The arrows indicate the orientation and the magnitude of the fields. The electric field (solid lines) points in the vertical direction, the magnetic field

(dashed lines), in the horizontal direction. The propagation of the wave is normal to both fields. Electromagnetic waves can only exist in three (or more) dimensions.

Let me stress that there was no experimental evidence that forced Maxwell to add the extra term. On the contrary, the first reactions of his colleagues were skeptical and, at this time, it was experimentally very difficult to measure the effects of the extra term. It took until 1887 for Heinrich Hertz to show that Maxwell's theory was correct when it described electromagnetic waves. So, actually, Maxwell's addition was not based on anything known: it was a counterfactual theory at the time the theory was made. The theory is the result of the critical analysis of the ideas on electric and magnetic forces which turned out to be inconsistent.

There is an oddity in the equations in that they are asymmetric. There is no fundamental reason known to us why ending magnetic field lines do not exist. To put it differently, there are no magnetic charges. Magnetic poles only exist in pairs. A piece of iron has a magnetic north pole and a magnetic south pole. Cutting the iron into pieces does not deliver isolated poles instead all the pieces have north and south poles. Maxwell's equations give a neat explanation of this fact: the magnetic fields are the result of current loops, analogous to a current carrying wire in a closed loop. The poles are located on either side of the loop. In this analogy cutting the loop into pieces must be done *along* the loop (otherwise the loop would not be closed anymore): this delivers two thinner loops with two poles. The poles will always come in pairs.

If there were magnetic charges, the equations would look like this:

- Ending electric field lines are connected to charges
- Not ending electric field lines are connected *to moving magnetic charges and to changing magnetic fields*
- Ending magnetic field lines are *connected to magnetic charges*
- Not ending magnetic field lines are connected to moving charges and by changing electric fields.

This is a counterfactual theory. It is deliberately counterfactual. It explores the consequences of what would change in our world if the theory were different. One of the great physicists who investigated these (counterfactual [wrong]) equations was Nobel laureate Paul Dirac. He discovered that if there was one single magnetic charge somewhere in the universe, both electric and magnetic charges would occur only in lumps. This means that both types of charge would be composed of elementary charges that could not be divided. Without magnetic charges Maxwell's equations say nothing about lumped charges. BUT: Charges do come in lumps! Dirac assumed something counterfactual and got a factual result that had been previously unexplained! Physicists like magnetic charges because they explain something very basic.

No one has ever found a magnetic charge. Once, in a balloon experiment something was found with the right signature<sup>vii</sup>, but only once, so it is very probable that this event was an error.<sup>viii</sup> Very recently it was found that in a certain type of material, called spin ice, the magnetic properties are such that it is as if there are magnetic monopoles<sup>ix</sup>. However, these are not elementary particles but the result of the collective behavior of moving electrons.

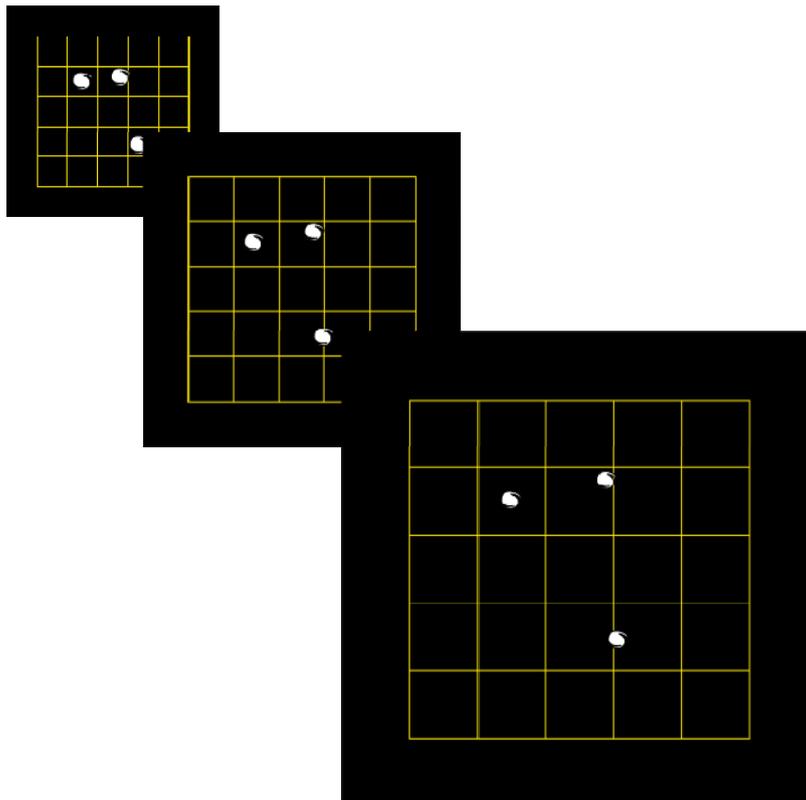
Magnetic charges pop up in the theory of elementary particle physics<sup>x</sup>, and theories beyond the, so-called, standard model. These predict them to have been

generated in the early universe in enormous numbers and with large masses (for elementary particles). It would have been impossible to miss them if they were around and yet, they do not exist now. Allan Guth<sup>7</sup> came up with a brilliant idea of a brief period of very fast expansion in the very early universe that diluted the density of the magnetic charges such that now there is, maybe, a single magnetic charge in the observable universe. Guth's idea has other attractive consequences for cosmology and is, therefore, taken quite seriously. In fact, all the data we have now supports the idea of a period of extremely fast expansion. This idea is now known by the name of *cosmic inflation*. This will be discussed further in the next section. So Dirac possibly was right: magnetic charges might exist, but are just too few to be found.

#### *The cosmological anthropic principle*

*A short history of the cosmos.* The properties of space, time, matter and interaction are combined in the standard model. There are two aspects of the standard model, one is related to the types of material particles and their interactions, the other relates to the beginning and evolution of the universe as a whole. This aspect states that the universe began  $13.8 \pm 0.1$  billion years ago with the so-called big bang<sup>xi</sup>. The picture we have of the big bang is that the universe started with infinitely large temperature and density, had a brief period of inflation during which it expanded by the enormous factor of  $10^{50}$  or so. When this period ended, after a tiny fraction of a second, our currently observable universe had a size of 1 m. The universe continued to expand at a much slower pace, closer to the rate that is currently observed. In doing so, it cooled down and the density of matter decreased. Once it was cool enough, protons (the positively charged nucleus of the lightest element, hydrogen) and neutrons (neutral particles a little heavier than the proton) formed. From these a certain amount of Helium (the second lightest element) and tiny amounts of heavier nuclei were synthesized within the first few minutes after the big bang. About 400,000 years after the big bang, the free electrons and nuclei combined to form atoms of hydrogen and helium. From this time on, the universe was transparent for light. This light can still be seen as a cosmic microwave background. It comes from all directions with the same intensity and uniformly pervades the universe. It has the signature of a body in perfect thermal equilibrium at a temperature of  $3.7^\circ\text{K}$ , with tiny fluctuations in intensity of 1 in 100,000.

A common misunderstanding regarding the big bang is that it was an explosion that happened at some location and from which material is flying away. Actually, the big bang theory describes a space filled with matter and radiation and it is the space itself that is expanding. A helpful analogy is the increase of the surface of a balloon or an expanding checkerboard, see fig. 4. The checkerboard's surface increases by homogeneous and isotropic stretching<sup>xii</sup>. The increase does not start from a spot on the surface, instead, the surface itself becomes larger. While the balloon is inflated at a constant rate, spots on the surface drift apart with some velocity. The greater the separation between the spots, the larger the velocity. Sitting at any one spot, one would see all other spots drift away and it would appear as if one were sitting in the center. This is not the case of course, this is just what is observed and one would have this impression from every spot! All but a few spots – notably the Andromeda nebula – move away from us with a velocity that increases with distance. We are not at the centre of the universe.



**Figure 4.** Checkerboard analogy of the expanding universe. The galaxies shown do not expand since their internal parts (stars, dust) are held together by gravitational forces

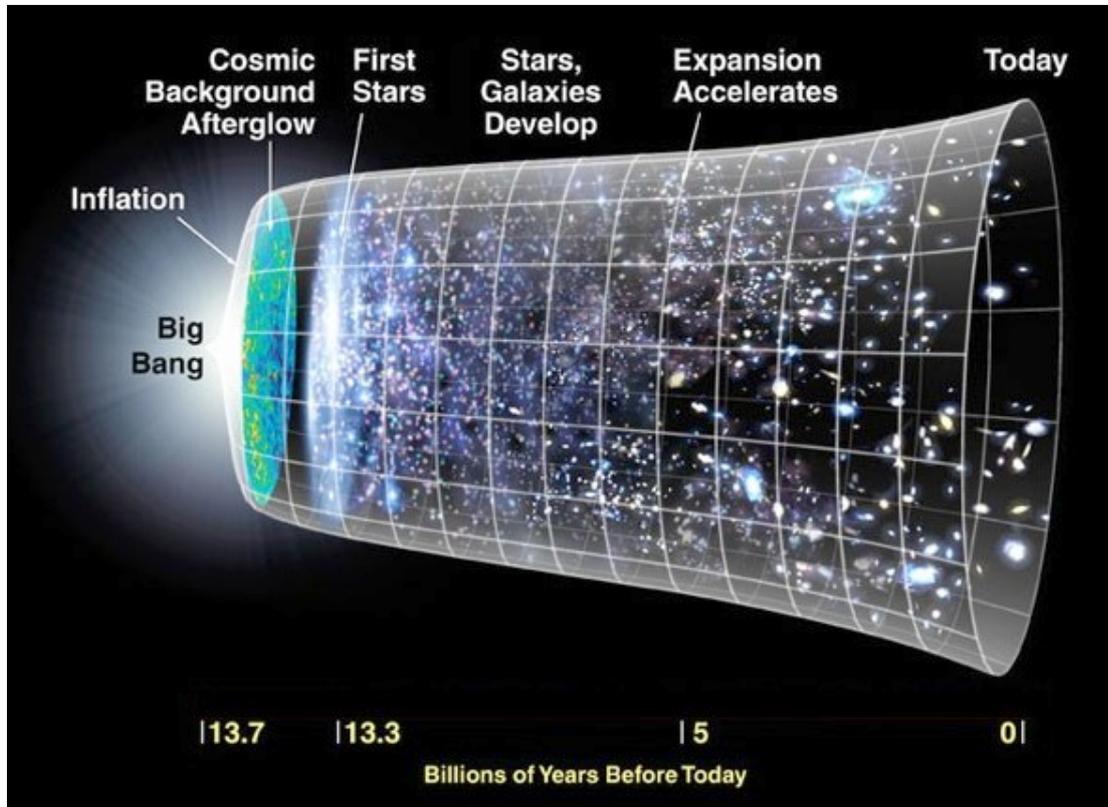
The cosmic background radiation contains a great amount of information. This information, together with a host of other measurements and observations, enables scientists to know, quite precisely, the age of the universe, its current expansion rate and, the global geometry (the universe is “flat”, meaning that it is infinitely large and that Euclidean geometry applies). Quite disturbingly, we also know that only 4% of all matter in the universe is made of “ordinary” visible matter, the rest is made up of, “dark matter” (ca. 30%) and “dark energy” (ca. 70%). Scientists have only vague ideas as to what dark matter could be, and no good ideas at all as to the nature of dark energy. All that is known is the gravitational effects of these types of matter. Dark matter is thought to be composed of elementary particles which interact, like all other matter, by ordinary gravitation alone. There is little doubt that dark matter really exists. Dark energy has the effect of pushing space apart. Einstein introduced into his theory of gravitation (the theory of general relativity, from 1916) a quantity that he called the cosmological constant. This constant pushes space apart in the same manner as dark energy.

For ca. 400 million years the cosmic background radiation was the only light in the universe. It was around this time that the first stars were born. According to current theory, these stars were much bigger than our sun and exploded in gigantic cataclysms a few million years after their birth (for comparison: our sun has a life span of ca. 10 billion years, some thousand times longer). At the same time, visible matter started to organize into small galaxies which grew, over time, into the gigantic galaxies with up to 100 billions stars that can be seen today.

All elements heavier than Hydrogen and Helium were formed in stars – we are literally made of star dust! It took approximately 7 billion years before enough matter made from heavy elements was available for the formation of planetary systems with rocky planets that were able to sustain life. In the centers of galaxies this happened earlier, however, galactic centers are not friendly for living organisms because of high

radiation levels and frequent supernovae explosions which sterilize their neighborhood. Only at a safe distance, such as our current location, organic life is able to develop and thrive over extended periods of time. It is an interesting fact that life on earth originated close to the earliest possible cosmic time.

The evolution of the universe is shown schematically in Figure 5.



**Figure 5.** Schematic representation of the development of the universe (credit: NASA). Note the increase in the expansion rate starting about 5 billion years ago.

*Dark energy.* A serious hint for the existence of dark energy was found only 10 years ago<sup>xiii,xiv</sup>. This result was supported a few years later by the measurements of the Wilkinson Microwave Anisotropy Probe (WMAP). This measured the tiny ripples in the cosmic background radiation mentioned above. All observational results seem to support the conclusion that, today, about 70% of all energy in the universe is dark energy. Dark energy pushes space apart. The expansion of the universe under the influence of dark matter is exponential<sup>xv</sup>, and there is no known mechanism that could halt this.

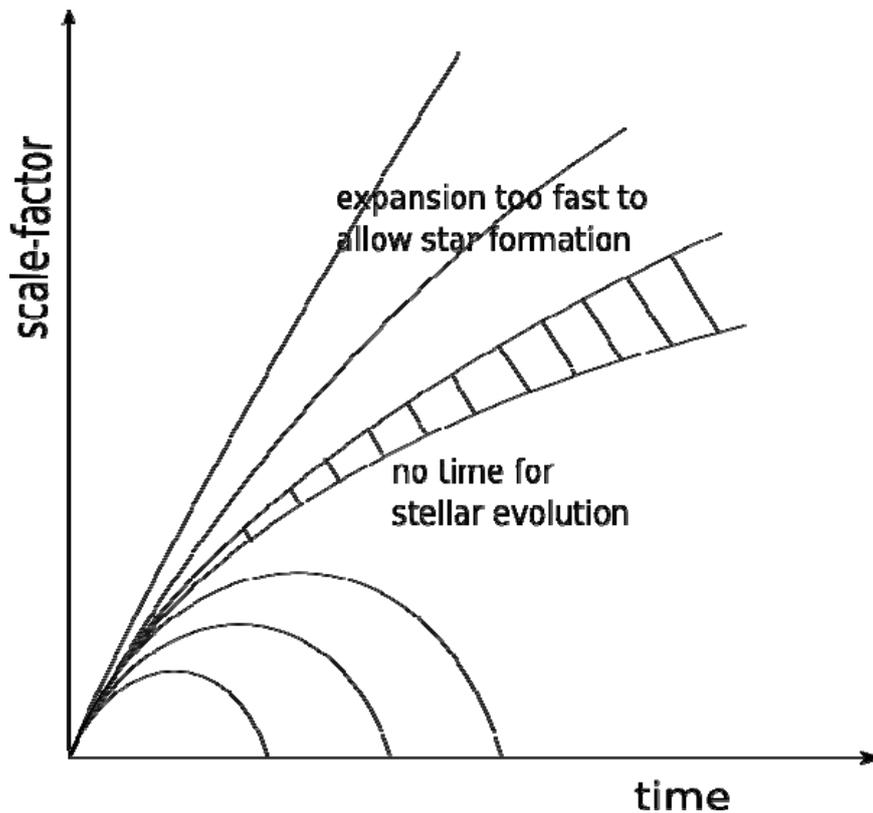
To fully appreciate the consequences of dark matter, an explanation of a special type of horizon is needed. On earth, the horizon is the line beyond which, one cannot see due to the curvature of the earth. In cosmology there is a horizon beyond which one cannot see due to a much more fundamental reason. The velocity of galaxies seen drifting away from us increases with the distance. There is no limit to this velocity, because the galaxies are not moving in space, but because space itself expands. Thus at a certain distance, called the Hubble distance, space and everything in it moves away from us with a velocity greater than that of light<sup>xvi</sup>. The value for this distance is given by the velocity of light divided by the expansion rate. A large expansion rate leads to a small Hubble distance. Only events within the Hubble distance can be causally related to each other. Beyond this distance, signals characterizing the event never reach that far. Exponential expansion means that the expansion rate will increase beyond all limits,

therefore the Hubble distance, which is the diameter of the causally connected space, in other words, the observable universe, will decrease beyond all limits.

*Cosmological anthropic principle.* Life as we know it would be impossible without hydrogen and stars to synthesize the other materials which we are made of, walk on, eat, etc. As we have seen, it took a certain amount of time to produce sufficient amounts of the required materials, and for the development of life. There is also a minimum amount of time required for the emergence of creatures that can admire their own existence. So one could pose the following question: what are the essential conditions the cosmos must meet for our existence? Or to put it differently: what conclusions can we reach about the properties of the cosmos from the fact that we exist? The latter question is related to the *cosmological anthropic principle*. It states that the fact that we exist tells us something about the world. There are several formulations. One of them, the so-called strong anthropic principle, states that the world is such that we (or life) *must* emerge. The method used to tackle this question is generally *counterfactual thinking*: how would the universe look like if this or that were different?

This inquiry gives some astonishing results and I want to discuss a few of them. I refer the interested reader to John Barrow's and Frank Tipler's rather technical book<sup>xvii</sup>; there is also a popular science book by Martin Rees, which is a much easier read<sup>xviii</sup>.

*Expansion rate.* Three effects control the expansion of the universe: the dark energy (which is of no importance for the following point), the global curvature and the energy content. These turn out to be balanced in a way such that the curvature in today's universe is very small. Astonishingly, it is as if the density of our universe is fine tuned so that the expansion is slow enough to allow sufficient time for the formation of stars and galaxies before they drift apart. This would not be the case if the density of the universe were a little bit smaller than it is. On the other hand, if the density were greater, the universe would have ceased expanding and instead it would have collapsed to an infinite density again a long time ago. A schematic representation of this situation is shown in fig. 6. The shaded region indicates the one in which star formation is possible. Note that the curves are very similar early on. This means that in the very beginning of the universe the density and expansion rate must have been extremely fine tuned in order to end up in the shaded region. Outside the shaded region, there are no stars in the universe and therefore no life comparable to our life.



**Figure 6.** The scale of the universe as a function of time. In the region in the lower right the expansion of the universe is so slow that it collapses; this situation is analogous to a rocket shot into the sky with a speed too small to escape earth. In the upper left region, the expansion is too fast to allow stars and galaxies to form: space is ripped apart so that the density of matter quickly becomes too diluted for the star formation process. It can be seen that all lines join to a single line at the lower left of the figure. Extremely small deviations from the actual evolution of the universe in the beginning would lead to either: a too fast collapse or a too fast expansion.

*Three dimensions.* Our world is three-dimensional. It seems to be silly to contemplate a world with four or two dimensions, but only at first sight. Why are there three dimensions instead of two or four? The anthropic principle says: because we can exist only in three dimensions. If the world had two or four dimensions we would not be here to pose silly questions. The equations of physics are all easily extended to greater or fewer dimensions.

The case of two dimensions is simple: there is no interesting world in two dimensions. Maxwell's equations work only in three dimensions, there is no light in two dimensions. Lightwaves come about because a changing magnetic field makes an electric field and vice versa. These fields are intimately connected with each other. There are no electric fields without magnetic fields and no magnetic fields without electric fields. In light waves these fields are oriented normal to each other and the propagation of the light is in the third direction. One needs three dimensions for light. The world in two dimensions would be dark. There would be no electromagnetic interaction possible and therefore no atoms, no molecules, no life and – darkness.

In four dimensions there are no stable atoms, no stable planetary systems, and no stable galaxies. Gravitational forces between masses would not depend on the distance between them like  $1/(\text{distance squared})$ , as it is in our actual world, but would decay faster. One of the consequences of this is, as a second year student of physics would be able to calculate, Kepler's laws would be very much different. The orbits would no longer form closed ellipses and they would be much more complicated; they would look rather similar to ellipses but such that the ellipses themselves would rotate in space. In

Kepler's laws the orientation of the ellipse is fixed in space. In four dimensions, the planets would constantly sweep through each other's orbits, giving rise to an unstable system. Similarly, the electrons in atoms would not be in fixed stable orbits. Having more than 4 dimensions does not improve this situation.

The equations have also been analyzed for the case that time would be two-dimensional instead of one-dimensional<sup>xix</sup>. Again, no stable systems would exist in such a world. So, we might say: I am, therefore the world has three spatial dimensions and one temporal dimension.

*Gravitational forces* are much weaker than the other forces. The repulsive electromagnetic force between two electrons is much, much larger than the attractive gravitational force<sup>xx</sup>. Yet, if gravitation were 1000 times weaker, the light that is created when matter condenses in the star forming process would blow the matter apart. If it were stronger by a similar amount, black holes would form instead of stars, Massive objects that release no light – gravitation is so strong that light cannot escape from the object – the universe would be dark.

*Stability of atomic nuclei.* Atomic nuclei are formed by protons and neutrons. There is a delicate balance of forces; there are attractive nuclear forces and repulsive electromagnetic forces. This balance leads to the stable elements we have in our world. One might ask what would happen if the balance of these two forces changes. The consequences would be disastrous. If the nuclear force were a tiny amount stronger two protons would stick together forming the element Helium but without the two neutrons that we have in our real world. Helium is stable because the two neutrons push the two protons a little bit farther away from each other. Therefore the repulsive electromagnetic force is insufficient to rip the nucleus apart, the neutrons make Helium stable. A nuclear force 1% stronger is sufficient to stabilize a Helium nucleus without neutrons. This (not existing) nucleus is called the *diproton*. Assuming a stronger nuclear force during the period in which Helium was formed, so during the first few minutes, all protons in the universe would have formed diprotons, some would have included neutrons. There would be no free protons left. There would be no hydrogen and, accordingly, no water, no organic chemistry, no life. Helium is an inert, chemically inactive element.

On the other hand, if the nuclear force were a little bit weaker, oxygen, carbon and all the heavier elements would be radioactive. Abundant Hydrogen but, again, no water, no organic chemistry, no life.

There are many more of these remarkable facts. For example, to form heavier elements in stars, there must be a collision between three Helium nuclei. This would be a very ineffective process were there not a, so-called, resonance, an energy state of the nucleus of Carbon, which makes the triple collision effective enough for the production of heavy elements. The resonance is a consequence of the value of the forces in the nucleus. The synthesis of heavy elements would have stopped if this resonance were not at this particular energy.

Water is so important for life because it is very special compared to other molecules. This is related to the position of the two protons that form a triangle together with the oxygen. This special form causes an unusual interaction of water molecules that makes many molecular processes, which are essential for life, possible. It is also the cause of the fact that solid water (ice) is less dense than liquid water, so ice drifts on the ocean's surface. Otherwise it would sink down to the bottom of the oceans, and most of the water in the oceans would be frozen.

So counterfactual thinking in physics leads to an amazing conclusion: if one maps possible worlds in a space spanned by the values of fundamental constants, our world

will be found in an isolated, tiny island in this space that supports complex structures. Deviate a little from the constants of nature, and no complex structures are possible. These universes would be mostly dark, dull places. The few where light existed would be without interesting chemistry. These universes would either be too small, or they would exist for too short a time to develop interesting structures, or matter would be too diluted to form stars leading to an interesting chemistry.

*Why does life exist at all?* If we can agree that life is necessarily complex, we arrive at the conclusion that life is impossible everywhere except in the close neighborhood of this small island<sup>xxi</sup>.

This property of our world begs for explanation. However, the explanations proposed are quite speculative, or even outside science. There is a line of argument<sup>xxii</sup> related to cosmologic inflation, the physical mechanism of which is also the subject of speculation. The laws of physics (to be precise: the standard model) give no hint as to why the constants of nature have these particular values. It could be that these values were the result of processes during inflation - what is known about physics does not exclude this possibility. It could very well be that the values emerged by an erratic, random process and that these are not the same everywhere in the universe. If this were the case, it is no coincidence that we happen to live in a region where the constants enable the emergence of life, because in other regions where life is impossible there simply are no observers. An alternative speculation (put forward by e.g. Robbert Dijkstra) is that some, as yet unknown physics, forces the value of the constants to the actual ones and that there is no other possibility for the values.

Some authors (e.g. Cees Dekker<sup>xxiii</sup>) conclude that the universe must have been made by some superior being. This must be a cruel kind of god having equipped the universe with a small density of dark energy, which causes exponential expansion of space. Due to dark energy the observable universe shrinks exponentially, making causally connected regions of space smaller and smaller, until galaxies, planetary systems, plants, continents, and finally molecules, atoms and nuclei are ripped apart. The prospects of life are bleak<sup>xxiv</sup>.

Freeman Dyson<sup>xxv</sup> has shown that some form of thinking will stay possible in an expanding universe without dark energy. The demiurge created a universe made fit for life only for some time. He condemned life to death even before it came into existence.

### *Conclusions*

Counterfactual thinking is one a physicist's core tools and, given its importance in physics, it would be astonishing if this were not the case for all other natural sciences. The "what if" question stands at the beginning of many scientific inquiries. Here, only some highlights in physics have been described, but much more mundane examples would have lead to the same conclusion.

In particular, we have seen that in cosmology the "what if" question leads to quite remarkable consequences: The laws of nature seem to be fine tuned to make possible the existence of life and the existence of observers. Slight changes in the laws and of the values of the basic constants would lead to inhabitable universes.

## References and Notes

---

- <sup>i</sup> This situation also has uncomfortable aspects, in particular, not everything which we find interesting can be cast into equations, and not everything can be quantified or measured in a sensible manner. So, physics deals only with those aspects of the world for which the quantitative method is adequate. It appears, however, that the realm of physics is expanding, and that the method can be applied to situations which had not been thought possible earlier.
- <sup>ii</sup> Feynman, Richard P., 1992, *The Character of Physical Law*, London: Penguin Books.
- <sup>iii</sup> Griffiths, David J., 1999, *Introduction to Electrodynamics*, Upper Saddle River: Prentice Hal.
- <sup>iv</sup> We could also have called the charges *white* and *black*, *high* and *low*, *a* and *b*. Incidentally, plus and minus turns out to be handy.
- <sup>v</sup> Fields may “end” at infinity, but infinity is not somewhere.
- <sup>vi</sup> See the discussion in Jackson, J.D., 1999, *Classical Electrodynamics*, Hoboken: John Wiley and Sons.
- <sup>vii</sup> Price, P.B., Shirk, E.K., Osborne, W.Z. and Pinsky, L.S., Evidence for detection of a moving magnetic monopole, *Phys. Rev. Lett.* 35, 1975, p. 487 – 490.
- <sup>viii</sup> A review regarding the present state of affairs is given by Milton: Milton, K.A., Theoretical and experimental status of magnetic monopoles, *Rep. Progress Phys.* 69, 2006, p. 1637-1711.
- <sup>ix</sup> Bramwell, I.S., Giblin, S.R., Calder, S., Aldus, R., Prabhakran, D. and Fennell, T., Measurement of the charge and current of magnetic monopoles in spin ice, *Nature* 461, 2009, p. 956 – 959.
- <sup>x</sup> Guth, Alan, 1997, *The Inflationary Universe: The Quest for a New Theory of Cosmic Origins*, New York: Perseus Books.
- <sup>xi</sup> The phrase “big bang” was coined by Fred Hoyle, who curiously never believed in this idea. Instead, he supported the steady state theory which states that the universe is infinitely old. The steady state idea has generally been abandoned by cosmologists because existing observations do not support it, but they do support the idea of a big bang.
- <sup>xii</sup> homogeneous refers to everywhere the same, isotropic refers to in all directions the same.
- <sup>xiii</sup> Hogan, Craig J., Kirshner, Robert P. and Suntzeff, Nicholas B., Surveying Space-Time with Supernovae, *Scientific American*, January 1999, p. 45 – 51.
- <sup>xiv</sup> Krauss, Lawrence M., Cosmological Antigravity, *Scientific American*, January 1999, p. 52 – 59.
- <sup>xv</sup> Exponential growth means that the scale of the universe doubles within a certain time interval and this will continue. As a consequence, the rate of expansion will also grow exponentially and double within the same time interval.
- <sup>xvi</sup> Special relativity forbids the motion of masses in space with a velocity higher than the speed of light, however expansion of space faster than light is not in conflict with the relativity theory.
- <sup>xvii</sup> Barrow, John D. and Tippler, Frank J., 1986, *The Anthropic Cosmological Principle*, New York: Oxford University Press.
- <sup>xviii</sup> Rees, Martin J., 2000, *Just Six Numbers, The Deep Forces that Shape the Universe*, New York: Basic Books.
- <sup>xix</sup> Tegmark, Max, Parallel Universes, *Scientific American*, May 2003, p. 40 – 51.
- <sup>xx</sup> The factor between these forces is  $10^{46}$ . No one is able to imagine such a large number. The ratio of the circumference of the earth to my length is  $2 \times 10^7$ ,  $10^{39}$  times smaller! The ratio of the diameter of the milky way (100,000 light-years) to my length is  $10^{21}$ , unimaginably large but still vastly smaller. The visible universe is believed to be as large as 13 billion light-years,  $10^{26}$  times larger than me, still not close to  $10^{46}$  – a factor of  $10^{20}$  to go...
- <sup>xxi</sup> Current state of research cannot strictly exclude the existence of other islands, see e.g. Jenkins, A. and Perez, G., Looking for life in the multiverse, *Scientific American*, January 2010, p. 42 - 49
- <sup>xxii</sup> Susskind, Leonard, 2006, *The cosmic landscape: String theory and the illusion of intelligent design*, New York: Little Brown.
- <sup>xxiii</sup> Dekker, Cees, Meester, Ronald and van Woudenberg, Rene (ed.), 2005, *Schitterend ongeluk of sporen van ontwerp? - Over toeval en doelgerichtheid in de evolutie*, Kampen: Uitgeverij Ten Have.
- <sup>xxiv</sup> Krauss, Lawrence M. and Starkman, Glenn D., The fate of life in the universe, *Scientific American*, November 1999, p. 58-66.
- <sup>xxv</sup> Dyson, Freeman, Physics and Biology in an Open Universe, *Reviews of modern physics* 51, 1974, p. 447-460.