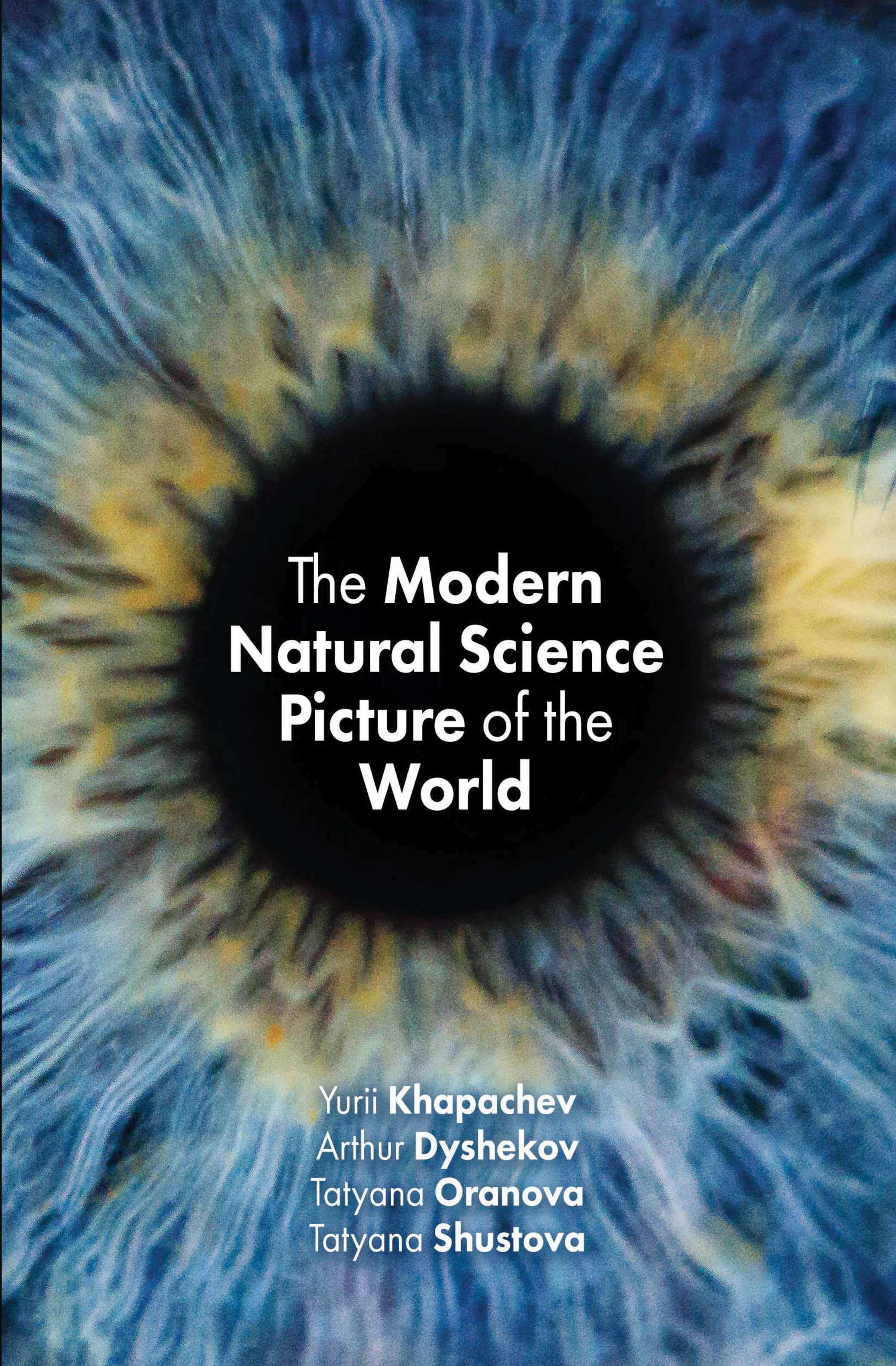


The Modern Natural Science Picture of the World

Yurii Khapachev
Arthur Dyshekov
Tatyana Oranova
Tatyana Shustova

The Modern Natural Science Picture of the World

The Modern Natural Science Picture of the World

By

Yurii Khapachev, Arthur Dyshekov,
Tatyana Oranova and Tatyana Shustova

Cambridge
Scholars
Publishing

The Modern Natural Science Picture of the World

By Yurii Khapachev, Arthur Dyshekov, Tatyana Oranova
and Tatyana Shustova

This book first published 2020

Cambridge Scholars Publishing

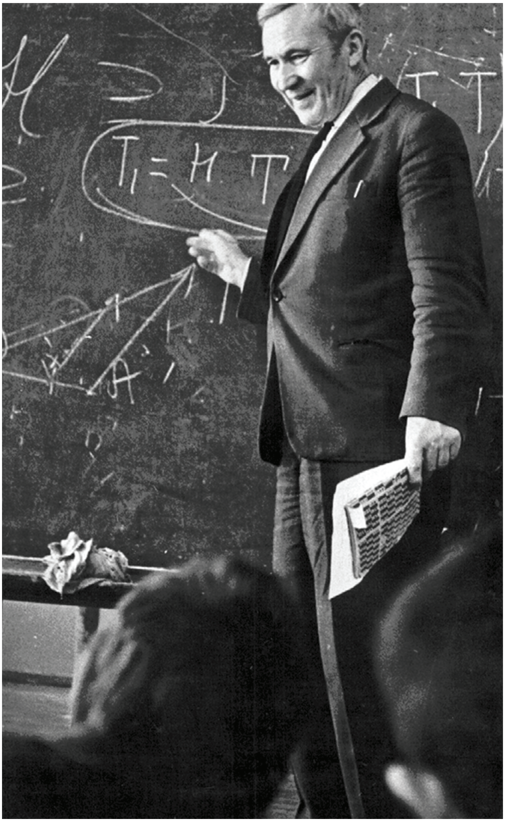
Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

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

ISBN (10): 1-5275-4611-X
ISBN (13): 978-1-5275-4611-0

This book is dedicated to the memory

of the Great Scientist and Teacher
Andrei Nikolaevich Kolmogorov
(25.04.1903–20.10.1987)



Giants thanks to which we see the scientific picture of the World “from the height of bird flight”.

Kolmogorov – Poincaré – Gauss – Euler – Newton: only five such lives separate us from the sources of our science.

Academician V.I. Arnold

TABLE OF CONTENTS

Part III: Living Systems

Preamble

We are not given to predict
How our word will respond ...
Fyodor Tyutchev

The achievements of the natural sciences over the past decades are so great that the very attempt to realize them today has become a difficult cognitive problem. Increasing differentiation is an obvious fact of modern science. Modern science has set itself the task of finding and developing common fundamental principles of scientific knowledge. Solving such a problem is by no means trivial. The search for a single foundation of natural science has long ceased to be the lot of philosophers alone. Today scientists from various fields such as physics, chemistry, biology, mathematics, etc., have been attached to its resolution.

Hence the desire of the authors of this course to lead the reader to the breadth of views on the problems of natural science corresponding to today. Unfortunately, traditional standards of education, especially in the humanitarian fields, do not allow the comprehension of the modern achievements of the science of nature.

Michael Faraday, who first organized public scientific readings for unprepared listeners, came to the conclusion that a truly instructive lecture can never be popular, but a really popular lecture will never achieve real instruction. In this book we will try to partly refute this point of view of the great scientist.

Separate ideas of the course are taken from the works of Vladimir Arnold, Paul Davies, Pavel Florensky, Alexander Markov, Boris Mednikov, Yuval Ne'eman, and Ilya Prigogine. We have included fragments of these works in the book, only slightly *mutatis mutandis*, not changing their meaning.

The authors remember the famous Gödel theorem on the incompleteness of an axiomatic description. Applied to our case, it boils down to the following statement. There is no such finite axiom system within which all problems can be described. Nevertheless, we consciously, where possible, axiomatized the presentation of the material. This approach, according to the authors, is in keeping with the idea of the book. In addition, it is possible to say *multa paucis*.

One or another concept becomes understandable to a person or even obvious a priori, if the ratio of the parameters characterizing it, K_i to the values of L_i corresponding to life experience becomes of the order of or less than unity (K_i/L_i <1). In some cases, this ratio can be satisfied by constant repetition, leading to habit. The incomprehensible becomes clear when it becomes habitual. However, one should not forget that, although *Repetitio est mater studiorum* – repetition is the mother of learning, it is the enemy of creativity. Here it is appropriate to quote Ovid: "Repetition is the mother of learning and the shelter of donkeys (the consolation of fools)".

The inclusion of art in the book seemed to us important for the following reason. This is due to the concept of the law of nature. The fact is that this concept was formed only within the framework of the European civilization. In the pictures of the world of some other civilizations, the concept of the laws of nature was simply absent. Consequently, this concept requires justification, which philosophy historically claimed, and at present it is becoming the subject of science itself. However, man's knowledge of the world occurs not only through science or philosophy, but also through art. Therefore, the idea of the laws of nature is reflected in art.

It is necessary to know, respect and appreciate philosophy and art. In the 18th century, the head of the Prussian Department of Education, Baron von Tsedlitz (Immanuel Kant dedicated his "Critique of Pure Reason" to him) inspired students: "After completing the science course, you will have to be a doctor, judge, lawyer, etc., only a few hours a day, but you will be a man for the whole day".

In conclusion of the preamble, we consider it necessary to explain why science is popularized. Currently, there is an increased separation between fundamental science and the mass consciousness. This gap is particularly noticeable in the natural sciences. The deeper scientists penetrate the secrets of nature, the more distorted their discoveries are in the media. As a result, an inadequate scientific picture of the world is being formed in the public consciousness. This trend may lead to the fact that society finally ceases to understand what scientists are doing and why they are needed.

Many people's lack of modern scientific knowledge at least at the elementary level is not at all harmless. Many of the most important decisions in society are endorsed through democratic procedures. In this case, everyone has the right to vote, regardless of their level of education.

People have a natural need to understand natural processes. In this regard, scientists are waiting for answers to key questions about the structure of the universe. It is assumed that these answers should be simple and straightforward. Of course, people would like the understanding of these answers not to require excessive intellectual effort. However, the world is much more complex than many would like to think. Therefore, in order for the modern scientific picture of the world to penetrate the mass consciousness, targeted systematic efforts are needed.

Recently, an outstanding philologist Andrey Zaliznyak at the award ceremony of the Solzhenitsyn Literary Prize drew attention to the following important circumstance. Nowadays, unfortunately, two old, banal ideas have gone out of fashion. First, truth exists, and the goal of science is to find it. Secondly, in any matter under discussion, a professional is in the normal case more correct than an amateur. These are confronted today by new, much more fashionable positions. The first point is that there is no truth, but a multitude of opinions. The second clause states that no one's opinion weighs more than someone else's. Therefore, popularization in the modern world is a public duty and a necessary means of civilization's self-preservation.

The idea of this book came from Prof. Yu.P. Khapachev as a result of the creation of an original training course for the humanitarian specialties of Kabardino-Balkarian State University, first read by the authors in the mid-1990s. Since then, a number of fundamental results have been obtained in the field of natural sciences. The most interesting are given in the book offered to the reader. This book generally follows our publications (Khapachev at al. 1997; Khapachev, Dyshekov, and Oranova 2013; Khapachev at al. 2017; Khapachev at al. 2018).

The translation into English and the general editing of the book were carried out by Prof. Yu.P. Khapachev and Prof. A.A. Dyshekov. Statements by Russian authors used as epigraphs were also translated by them. This translation, of course, is not poetic, but most importantly, it does not lose the meaning of the original.

Thanks.

The authors are grateful to Prof. B. Karamurzov and PhD. A. Tashilov for encouraging the publication of this book.

LIST OF ABBREVIATIONS

ALS	artificial living system
ANS	autonomic nervous system
BFB	biological feedback
CAR	compensatory-adaptive reaction
CMB	cosmic microwave background
CNS	central neural system
CR	conditioned reflex
DNA	deoxyribonucleic acid
fMRI	functional magnetic resonance imaging
GTR	general theory of relativity
GUT	grand unified theories
HNA	higher nervous activity
KAM	Kolmogorov – Arnold – Moser theory
mOFC	medial orbito-frontal cortex
mRNA	messenger RNA
mtDNA	mitochondrial DNA
mtEve	mitochondrial Eve
PNS	peripherical nervous system
RNA	ribonucleic acid
SFWU	structural functional working units
STR	special theory of relativity
TD	thermodynamic
TS	transition state
UCR	unconditioned reflex
WB	“Weisman barrier”

INTRODUCTION

Oh, how many wonderful discoveries to us
Prepares an enlightened spirit
And experience, the son of difficult mistakes,
And genius, paradoxes friend,
And the chance, God is the inventor.
Alexander Pushkin

Natural and humanitarian culture. Beauty

Let's face what is at first glance a trivial question. What is culture? Despite the seemingly obvious answer to this question, there are many interpretations of this concept. We confine ourselves to the most obvious sign of culture. Culture refers to everything that relates to the results of human activity. These results can have a direct material character (material culture), as well as be fixed on various material carriers (books, notes, etc.) without having a material embodiment (non-material culture).

In this book we mainly consider conceptual issues of fundamental science which naturally relate to the sphere of non-material culture. Ideas permeate not only the field of scientific knowledge, but also art as the most important part of humanitarian culture. One of the main concepts that unite the two types of culture is beauty. We will not try to give a definition of this concept but simply give a few examples from various fields of art and science. At the same time, we hope that the reader understands that art is fundamentally subjective in its essence, and the examples we cite cannot be an absolute example of beauty for all people. On the other hand, science claims to be objective knowledge, and therefore the beauty in it is absolute.

Beauty in Humanitarian Culture

Painting and Sculpture

Sandro Botticelli: Spring, The Birth of Venus. Uffizi, Florence, Italy.

Leonardo da Vinci: The Last Supper, Convent of Sta. Maria delle Grazie, Milan, Italy.

Raphael: The School of Athens, in the Apostolic Palace in the Vatican. Madonna Sixtina. Dresden Gallery, Dresden, Germany.

Michelangelo: Pietà, in St. Peter's Basilica in the Vatican. The Creation of Adam, Last Judgement, in the Apostolic Palace in the Vatican.

Andrei Rublev: Trinity, Christ the Redeemer. Tretyakov Gallery, Moscow, Russia.

Reproductions of these works are available on the Internet.

The task of painting is not to duplicate reality, but to give the most profound comprehension of its material and its meaning. The artist comprehends this meaning directly contacting reality. He gets used to this reality. Therefore, painting is, or wants to be the truth of life. This truth of life does not replace life, but only symbolically reflects it in the deepest reality.

In 1919 Florensky wrote an article “Reverse perspective” (Florensky 1999). This article is devoted to understanding the phenomenon of the way in which images take up space on a plane. Florensky examines the iconographic canon and compares it with examples of world art. He points to the regularity of the periodic return of artists to the reverse perspective and abandonment of it. This was due to the spirit of the times and historical circumstances which influenced one or another worldview of the artist.

People who first examine Orthodox icons from the XIV and XV centuries, and partly from the XVI century, are struck by unexpected promising relationships. These paradoxical relations contradict the rules of linear perspective from the point of view of which they are the gross illiteracy of the drawing. However, after a while, icons with a large perspective violation for artistic perception turn out to be more attractive than those that more closely match the perspective textbook. These violations of the rules of perspective are so persistent and so systematic that involuntarily the thought arises about the non-randomness of these violations. Thus, there is a special system of image and perception of reality, on the icons depicted.

Is perspective really, always and everywhere, to be seen as an absolute prerequisite of artistic truthfulness? Or is it one of the possible schemes of visualization, corresponding not to the worldview as a whole, but only to one of the possible interpretations of the world. And this is connected with quite a certain life feeling and life understanding.

Florensky reports referring to Vitruvius. Around about 470 BC, Aeschylus staged his tragedies in Athens, and Agafarh made the scenery and wrote a treatise about it, “Commentarius”. It is on this occasion that Anaxagoras and Democritus had the opportunity to consider this very subject (writing scenery) scientifically.

Florensky analyzes the work of the geniuses of the Renaissance and classic perspectivism. He notes their deliberate violation of the direct perspective rules of The Last Supper as Leonardo has the task of removing the spatial distinction between that world, the gospel, and this, the everyday. The fresco presents a stage setting. In this scene the laws of Kant, space and Newtonian mechanics reign. But if it were only thus, there would be no supper. And Leonardo commemorates the special value of the committed violation of the uniqueness of scale. A simple measurement will show that the room barely has a double human height, with a triple width. Thus, the room does not correspond to the number of people in it, or the greatness of the event. However, the ceiling does not seem oppressive, and the smallness of the room gives the picture a dramatic richness and fullness. Unnoticed, but surely, the master resorted to a perspective violation and applied different units of measurement to the actors and to the room. He reduced the size of the room, though differently in different directions. Thus, he glorified people and gave the humble farewell dinner the significance of a world-historical event at, moreover, the center of history. The unity of perspective is broken, the duality of the Renaissance soul has manifested itself, but on the other hand, the picture has become aesthetically convincing.

It is known what a magnificent impression the architecture makes in the "Athenian School" of Raphael. If you characterize these arches, then you want to compare them, for example, with the Cathedral of St. Peter in Rome or the Moscow Cathedral of Christ the Savior. It seems that the arches are equal in height to the church. However, measurement shows that the height of the columns is slightly more than twice the height of the figures. So the whole building, apparently so magnificent, would be very insignificant if it were actually built. Raphael's method in this case is also very simple. He accepted two points of view, located on two horizons from the top point of view, the floor is drawn and the whole group of faces, from the bottom drawn arches and in general the entire upper part of the picture. If the figures of people had a common vanishing point with the lines of the ceiling, then the heads of people in the depths of the picture would have dropped below and would have been closed off by the people standing in front. This would damage the picture. The vanishing point of the ceiling lines is in the right hand of the central figure (Aristotle) who holds a book in his left hand and points to the right with his right hand. To conceal this promising violation, Raphael placed the actors in the depths of the picture and disguised the lines of the floor to the horizon.

As in many other paintings Raphael balances two principles, perspective and reverse. This corresponds to the peaceful coexistence of two worlds, two spaces. It is just as if the veil of another world has unfolded silently before us, and our eyes see not a scene, not an illusion in this world, but a real, though not invading, different reality. Raphael gives a hint of such a spatial property in the Madonna Sixtina with its separated curtains.

There is a balance of the two principles of spatiality in “Apostle Paul's Conversion” by Michelangelo. But there is a completely different spatiality in his “Last Judgment”. The fresco represents a certain slope: the higher a point in the picture, the further from the viewer is the point that it depicts. Consequently, as the gaze is raised, the eye will see smaller and smaller figures, due to the perspective reduction. This is evident from the fact that the lower figures block the upper ones. But, the size of the figures increases as they rise on the fresco, i.e., as they move away from the viewer.

Such is the property of that spiritual space. Within it, the farther away something is, the greater it is, and the closer, the smaller it is. This is the reverse perspective. Having understood it, and, moreover, so consistently carried it out, we begin to feel its complete incommensurability with the space of the fresco. We are not drawn into this space. On the contrary, it pushes us. Although visible, it is transcendent to us, thinking according to Kant and Euclid. Michelangelo who lived in the Baroque was, however, not in the past, not in the future of the Middle Ages. He was a contemporary and not at all contemporary to Leonardo.

Architecture

Church of the Protection of the Theotokos on the Nerl. Vladimir Region, Russia. Architect unknown.

The Tempietto within a narrow courtyard. Rome, Italy. Architect Donato Bramante.

Taj Mahal. Agra, Uttar Pradesh, India. Ustad Ahmad Lahauri. Persian architect, student of Sinan, he is often called the main creator of the architectural image of the monument.

Music

Johann Sebastian Bach: Toccata and Fugue in d minor.
Wolfgang Amadeus Mozart: Symphony No. 40.
Sergey Rachmaninoff: Prelude in g minor, Op. 23.

Poetry

William Shakespeare
Sonnet 117
Accuse me thus: that I have scanted all
Wherein I should your great deserts repay,
Forgot upon your dearest love to call,
Whereto all bonds do tie me day by day;
That I have frequent been with unknown minds
And given to time your own dear-purchased right
That I have hoisted sail to all the winds
Which should transport me farthest from your sight.
Book both my wilfulness and errors down
And on just proof surmise accumulate;
Bring me within the level of your frown,
But shoot not at me in your waken'd hate;
Since my appeal says I did strive to prove
The constancy and virtue of your love.

John Donne
The Anniversary
All kings, and all their favourites,
All glory of honours, beauties, wits,
The sun itself, which makes time, as they pass,
Is elder by a year now than it was
When thou and I first one another saw.
All other things to their destruction draw,
Only our love hath no decay;
This no to-morrow hath, nor yesterday;
Running it never runs from us away,
But truly keeps his first, last, everlasting day.

Two graves must hide thine and my corse;
If one might, death were no divorce.
Alas ! as well as other princes, we
– Who prince enough in one another be –
Must leave at last in death these eyes and ears,
Oft fed with true oaths, and with sweet salt tears;
But souls where nothing dwells but love
– All other thoughts being inmates – then shall prove
This or a love increasèd there above,
When bodies to their graves, souls from their graves remove.

And then we shall be thoroughly blest;
But now no more than all the rest.
Here upon earth we're kings, and none but we

Can be such kings, nor of such subjects be.
Who is so safe as we? where none can do
Treason to us, except one of us two.
True and false fears let us refrain,
Let us love nobly, and live, and add again
Years and years unto years, till we attain
To write threescore; this is the second of our reign.

Rudyard Kipling
If
If you can keep your head when all about you
Are losing theirs and blaming it on you,
If you can trust yourself when all men doubt you,
But make allowance for their doubting too;
If you can wait and not be tired by waiting,
Or being lied about, don't deal in lies,
Or being hated, don't give way to hating,
And yet don't look too good, nor talk too wise:

If you can dream – and not make dreams your master;
If you can think – and not make thoughts your aim;
If you can meet with Triumph and Disaster
And treat those two impostors just the same;
If you can bear to hear the truth you've spoken
Twisted by knaves to make a trap for fools,
Or watch the things you gave your life to, broken,
And stoop and build'em up with worn-out tools:

If you can make one heap of all your winnings
And risk it on one turn of pitch-and-toss,
And lose, and start again at your beginnings
And never breathe a word about your loss;
If you can force your heart and nerve and sinew
To serve your turn long after they are gone,
And so hold on when there is nothing in you
Except the Will which says to them: "Hold on!"

If you can talk with crowds and keep your virtue,
Or walk with Kings – nor lose the common touch,
If neither foes nor loving friends can hurt you,
If all men count with you, but none too much;
If you can fill the unforgiving minute
With sixty seconds' worth of distance run,
Yours is the Earth and everything that's in it,
And – which is more – you'll be a Man, my son!

Some British people say that one who knows these lines of Kipling is capable of bringing the future closer. Who knows, maybe they are right?

The Servant When He Reigneth

For three things the earth is disquieted,
and for four which it cannot bear.
For a servant when he reigneth,
and a fool when he is filled with meat;
for an odious woman when she is married,
and an handmaid that is heir to her mistress.
PROV. XXX. 21-22-23.

Three things make earth unquiet
And four she cannot brook
The godly Agur counted them
And put them in a book –
Those Four Tremendous Curses
With which mankind is cursed;
But a Servant when He Reigneth
Old Agur entered first.
An Handmaid that is Mistress
We need not call upon.
A Fool when he is full of Meat
Will fall asleep anon.
An Odious Woman Married
May bear a babe and mend;
But a Servant when He Reigneth
Is Confusion to the end.

His feet are swift to tumult,
His hands are slow to toil,
His ears are deaf to reason,
His lips are loud in broil.
He knows no use for power
Except to show his might.
He gives no heed to judgment
Unless it prove him right.

Because he served a master
Before his Kingship came,
And hid in all disaster
Behind his master's name,
So, when his Folly opens
The unnecessary hells,
A Servant when He Reigneth

Throws the blame on someone else.

His vows are lightly spoken,
His faith is hard to bind,
His trust is easy broken,
He fears his fellow-kind.
The nearest mob will move him
To break the pledge he gave –
Oh, a Servant when he Reigneth
Is more than ever slave!

Johann Wolfgang Goethe
O'er all the hill-tops
Is quiet now,
In all the tree-tops
Hearest thou
Hardly a breath;
The birds are asleep in the trees:
Wait; soon like these
Thou too shalt rest.

Translated by Henry Wadsworth Longfellow

This poem is a diamond of German poetry. The translation into Russian was by the great Russian poet Mikhail Lermontov. In turn, this translation is considered a brilliant example of Russian poetry.

Beauty in mathematics

> The mathematician plays a game in which he himself invents the rules while the physicist plays a game in which the rules are provided by nature, but as time goes on it becomes increasingly evident that the rules which the mathematician finds interesting are the same as those which nature has chosen.

So wrote one of the founders of quantum mechanics, the Nobel laureate Paul Dirac in 1939. The beauty of mathematics is the ability to see the true essence of things. Perhaps this refers to any beauty. First of all, it should be noted that beauty is associated with taste. It is better not to argue about tastes, but we believe that taste can sometimes be developed through education. Secondly, beauty can be external (form) and internal (meaning). There is a lot of beauty of both types in mathematics. The second type of beauty is deeper and inaccessible not only to people with a liberal arts education but also to representatives of other sciences. Let us give a few examples that, we think, are available to everyone.

1. Geometers of Ancient Greece achieved much, but they could not calculate the volume of the ball. Brilliant Archimedes derived a formula, applying the idea of weighing!

2. In Euclid's book "Elements", a method for constructing regular figures is described: a triangle, a square, a pentagon, a fifteen-square and all polygons obtained by doubling the number of sides (using only a divider and a ruler). After a thousand years, it was proved that it is impossible to construct correct 7- and 9-gons (with the same tools). There was no consideration about building 11- and 13-gons. Mathematicians believed that other regular polygons could not be built. However, the 19-year-old Gauss, using imaginary numbers, found a way to construct a regular 17-gon.

3. Take an arbitrary triangle; draw trisectrices of angles (i.e. rays that divide the angles into three equal parts). Mark the points of intersection of three pairs of trisectrices leaning toward the sides. It turns out that these three points form an equilateral triangle (Morley's theorem). The proof of this theorem by means of elementary mathematics is not easy. The French mathematician Alain Connes found a very short and beautiful proof of Morley's theorem using imaginary numbers.

The sum of the corners of the triangle

However, the very concept of beauty is not limited to the visual aspect. There are beautiful poems, beautiful music, beautiful relationships, beautiful reasoning, and beautiful mathematical constructions.

Beauty in mathematics is a fine line between simplicity and complexity, naturalness and unusualness, a riddle and its solution. Beauty is what allows us to see more than we saw a moment ago. Beauty surprises us.

Apparently, the category of beauty in mathematics first appeared in ancient Greece with the advent of geometry.

Of course, intellectual pleasure in solving geometric problems is valuable in itself. But we believe that the aesthetic side also matters. In fact, what else can explain the desire to study abstract drawings made up of straight lines, line segments and circles?

Put yourself in the place of ancient geometers. The practical significance of most of your research will become clear only after many centuries. And now you just draw triangles in the sand and discover an amazing regularity: whatever triangle you build, the sum of its internal corners always makes a straight angle.

You feel that this cannot be an accident. There must be some reason, some explanation. But in the picture, there is not. This fact does not give you peace; you constantly think about it. Finally, perhaps almost randomly, intuitively, you add a new stroke to a drawing with a triangle. You draw a line passing through one of its vertices parallel to the opposite side.

Look at the picture (Fig. a), reason (logic turns on) and understand. The three angles are equal to the angles of your triangle, which together form a straight angle. Here they are, in front of you!

Now it is clear that nothing else could be. The fact, that a few minutes ago still seemed an unsolved mystery, has become a strictly proven fact. As a result, an amazing and wonderful meaning was revealed to your mental gaze due to intuition and logic.

You look at your drawing, which consists of only a few line segments, and you understand that this is one of the most beautiful pictures in your life.

This is what mathematical beauty looks like.

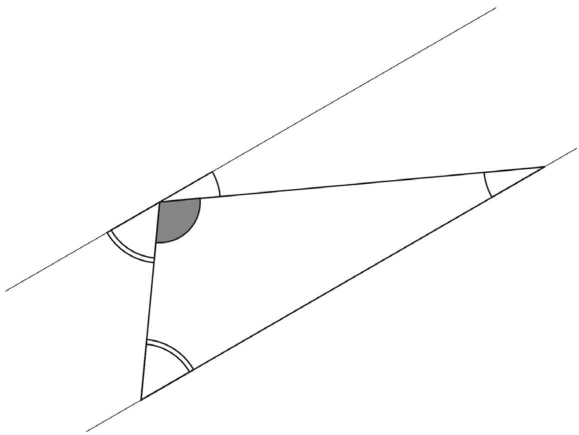


Fig. a.
With an additional straight line, it becomes clear that the sum of the angles of the triangle is equal to two right angles

Part of not less than a whole. The set of a power continuum

On line segments of different lengths are the same number of points. To verify this, it is enough to arrange the points in pairs (that is, in a pair of one point from one line segment, and the second from the other). Here is an illustration of how to do it (Fig. b). Each dotted line intersects each line segment, here are the intersection points and formed pairs.

Similarly, there are as many points on a line segment as on a straight line, and even on the whole plane. In a more general form, it turns out that a part is not always smaller than a whole. Moreover, the part (in a certain sense) is equal to the whole!

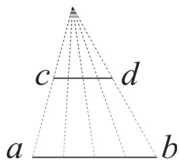


Fig. b.
The figure shows that there are as many points on the line segment as on the straight line

That which does not seem obvious or even impossible is realized. The impossible is thus possible. Here is an example of how the "Impossible is possible." Take a round target with infinitely many points on it, so the probability of hitting any of them is zero. Now let's shoot an arrow, it will hit some point. Let me remind you that this point was no better and no worse than all the others, with a zero probability of hitting it. That event happened, the probability of which was zero.

Fractals

The word beautiful is associated primarily with something visually pleasing, pleasing to our eyes, like a painting in a museum. There is such beauty in mathematics. Some mathematical objects allow a representation in the form of images, sometimes very pleasing to the eye.

The beauty of this example is understood by the humanities. How it turns out mathematically is not so easy to understand even for many techies. It is noteworthy in fractals that from a small mathematical formula it turns out to be an incredibly complex and sophisticated piece, the like of which we would rather get used to seeing in nature or in art.

The most beautiful formula

Two numbers: π and *e*. These numbers are everywhere; you can find and read a bunch of different facts about these numbers. It is simply impossible for an unprepared person to comprehend which of the facts is the cause, and which the result and where it all comes from. The most surprising is the connection of these numbers. If, nevertheless, you understand where π and e come from, when you see the Euler identity, in which there are five fundamental constants of mathematics π, *e*, 0, 1 and *i*, a sense of admiration encompasses you, if not your own mind, then the minds of other people.

Here it is:

$$\exp(i\pi)+1=0.$$

In this formula the number *e* = 2.718281828..., or the base of the natural logarithm, *i* is the imaginary unit, the number π = 3.141592653... the ratio of the circumference to the length of its diameter, 1 is the unit, the neutral element in the multiplication operation, 0 – zero, the neutral element of the operation of addition. Truly brevity is the sister of talent, if there is one, of course.

Unfortunately, or fortunately in mathematics and theoretical physics, not everything is always simple. Therefore, we must remember the following words of Paul Dirac:

> A theory with mathematical beauty is more likely to be correct than an ugly one that fits some experimental data. God is a mathematician of a very high order, and He used very advanced mathematics in constructing the universe.

Part I:

The World as a Physical System

Chapter 1

Discourse and intuition. Criterion of evidence. The problem of scientific axiomatics

Vita brevis, ars longa, occasio praeceps,
experientia fallax, iudicium difficile.
Hippocrates

Let's start with the concretization of our topic. To do this, we should at least briefly dwell on the terminology that we will use. This is extremely important, because it is the terminology that specifies in some cases the subject of the research itself.

Recall that in the introduction we talked about natural and humanitarian cultures. In our book, we will talk further mainly about the first one. The language of science is created on a mathematical basis. However, our book is intended primarily for people who do not professionally own the mathematical apparatus. Therefore, we will deliberately avoid the presentation of excessive mathematization giving only a minimum of publicly available formulas.

In natural science the way of thinking, first of all, is logical, rational, and discursive. However, and it is very important to clarify from the very beginning, the construction of science is impossible without a kind of irrational thinking – intuition.

It is intuition that makes it possible to express a previously unknown statement as a hypothesis, which can then either be confirmed or disproved. What is intuition? For our purposes, it suffices to mention the most obvious feature of this concept. Intuition is a direct guessing of the result. Note here that the result may be false.

For example, at some stage of the knowledge of mankind, it seemed intuitively obvious that the Sun revolves around the Earth. In fact, even our distant ancestors observed the visible movement of the Sun across the sky. Now we know that the true picture is just the opposite.

The second example will seem to you less believable. Nevertheless, we will give it right now, postponing the explanation to the corresponding chapter. Consider a "global" task. Suppose we are standing on the platform of a train station in London and are looking at the roof of a train going to Paris. On the roof of the "London-Paris" train runs Professor Moriarty (Fig. 1-1).

Fig. 1-1.
Professor Moriarty runs on the roof of the train

The London-Paris train is moving at a speed of v_T. Prof. Moriarty's speed on the train is v_M. Calculate the speed of Prof. Moriarty relative to the London railway station. You will say that this is a primitive task from the course of school physics, and it all depends on which way Prof. Moriarty is running. If in the direction of the train, then $V = v_T + v_M$, if against the movement of the train, then $V = v_T - v_M$. Simple, but absolutely wrong!

Now we will not explain why these simple formulas of school physics are wrong. And this is despite the fact that they give in our everyday life a result that perfectly matches the experiment. The point here, of course, is not in the personality of the great criminal, Professor Moriarty. It turns out that a good, even very good coincidence with an experiment does not mean more truth.

It is reasonable to remind ourselves why this or that concept becomes understandable to a person or even intuitively obvious as it were, *a priori*. This happens if the ratio of the parameter characterizing the concept K to the value L corresponding to life experience becomes of the order of or less than unity K/L <1. Otherwise, the concept seems absurd to us or at least incomprehensible. The following examples can explain what has been said.

While a person was thinking of usual distances, i.e. about meters or kilometers, the idea of the sphericity of the Earth caused considerable difficulties. Recall that the radius of the Earth is about 6400 km. Meanwhile, it is known that even at the turn of the III-II centuries BC in Egypt, the Alexandrian scientist Eratosthenes of Cyrene (276–194 BC) rather accurately measured the Earth's radius from the difference on a shadow deviation in Alexandria and Luxor on the day of the summer solstice. Interestingly, Christopher Columbus had a significantly underestimated value of the Earth's radius. That is why he hoped to go around the globe and sail to India so quickly. As we see, sometimes a mistake leads to discoveries.

Another example is the incorrectness of the above discussed speed formula. The problem is that our usual speed does not exceed 10^3 m/s even for a rocket. This is significantly less than the speed of light $c = 3 \cdot 10^8$ m/s.

The last characteristic example is connected with the seeming paradox of the laws of the microworld. The problem here is that our natural pace of life is determined by a pulse rate of about 60 beats per minute, i.e. 1 Hz. This value is at least 16 orders of magnitude less than the "world" of atomic frequencies (10^{16} Hz for optical radiation and 10^{19} Hz for *X*-rays and gamma radiations).

We now turn to the axioms of science. We consider the axioms of science to be the criteria of the scientific worldview that separate it from art and religion.

Axiom 1. *Sine ira et studio*; which means: without anger and addiction. In a broader sense, this means the following. To comprehend the scientific truth, do not have a preconceived opinion and question everything.

It is clear that this axiom separates scientific thinking from religious, reconstructive-prophetic thinking. For example, the doubt that the sum of the angles of a triangle is always 180°, led to the creation of a fundamentally new branch of mathematics – non-Euclidean geometry (Figs. 1-2 and 1-3). It is remarkable that one of these geometries has found an application in the description of the world in the general theory of relativity, which we will discuss later. This is Riemann's geometry.

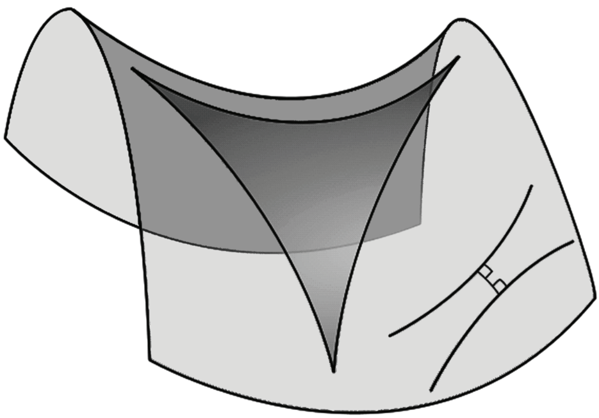


Fig. 1-2.
Hyperbolic geometry. The sum of the angles of the triangle is less than 180°

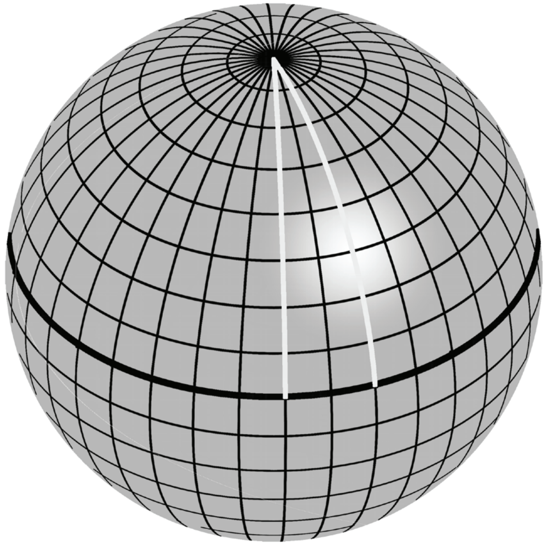


Fig. 1-3.
Elliptical geometry. The sum of the angles of a triangle is greater than 180°

Axiom 2. A hypothesis formulated on the basis of intuition-guesswork should be tested experimentally.

Note that, most important and fundamental for science is the intuitive judgment of the sufficiency of the experimental test.

In science, it is believed that experience has evidential value, although it is always limited in itself.

This statement is called intuition-judgment. It does not boil down to any axioms, since it has itself the character of an axiom. In this way, intuition-judgment differs from intuition-guesswork, which is a product of hypotheses. Both of these different intuitions are nothing but two kinds of human superconsciousness.

The known statement "practice is the criterion of truth" has been adopted by mankind as an axiom since the times of ancient Greece. It was then, for the first time in European civilization, that the idea of the law of nature arose. This view had far-reaching consequences in the development of science and technology. It should be noted that for other civilizations such a view did not exist. Therefore, empirically found discoveries (for example, gunpowder, the compass, etc.) were not considered from the point of view of the consequences of the laws of nature. They were handed down from generation to generation as a secret.

Axiom 3. Experimental facts must be reliable.

The reliability of the experiment is a necessary condition for the principle of the objectivity of scientific knowledge. Reliability in our case means the following. If the same experimental conditions are fulfilled, then no matter who conducted it, and when, similar results should be obtained. In this case, the similarity of the results does not mean in the general case their identity. First, this is due to inevitable measurement errors. Secondly, with the very nature of the investigated phenomena, which may be fundamentally unpredictable, it is probabilistic in nature. Such a character can manifest itself in physical, chemical, and biological systems.

Axiom 4. The theory should be based only on reliable facts.

Justus von Liebig believed that philosophers are partly right that facts give rise to ideas. But he also focused our attention on the fact that in order to understand the facts, “it is necessary to have certain ideas in mind”. If these ideas, in the interpretation of facts, are correct, then, according to Liebig, “you cannot see with your eyes what the mind sees.”

The theory should thus be built on such reliable facts that are correctly understood. Otherwise, reliable but misunderstood facts may confirm an erroneous theory. To discuss such a highly indicative episode in science, we will stop here. Briefly the background is as follows.

In 1940 Kolmogorov published an article on the new confirmation of Mendel's laws (Kolmogorov 1940). The reason for the publication of this work was the following circumstance. In 1939 an article appeared in which reliable experimental data were obtained. They showed that the frequency of manifestation of the dominant trait in second-generation hybrids fluctuates quite strongly, not corresponding very well to the value of ¾ (as it should be according to Mendel's law). From this it was concluded that the laws of Gregor Mendel are not fulfilled. In the work of Kolmogorov, it is proved that with the actual number of observations of the order of hundreds the coincidence with Mendel's theory should be considered very good.

To analyze the results of experiments, the method proposed by Kolmogorov in the article on the empirical definition of the distribution law (Kolmogorov 1933) was used. In this article a criterion was formulated which is now universally recognized in mathematical statistics and is called the Kolmogorov criterion. The use of this criterion leads to the following, at first glance, paradoxical conclusion. The experimental results of the work of 1939 contrary to the opinion of its author turn out to be another brilliant confirmation of the laws of Mendel.

With regard to axioms 3 and 4, it should be noted that a direct experiment is not always possible.

Axiom 5. *Subsantia non sunt multiplicanda praeter nesessitatem*. The so-called "principle of razor" of William of Ockham. Do not multiply entities unnecessarily, i.e. explain the facts in the simplest way. In fact, this means that when choosing between two theories, preference should be given to that based on fewer axioms, principles, propositions or assumptions.

In the future, we will show how this axiom works on important conceptual principles. Here is just one example.

In the 6th century BC, Pythagoras of Samos expressed the idea of a spherical Earth located at the center of a spherical Universe. For the satisfactory experimental confirmation of the geocentric hypothesis by Claudius Ptolemy in the 2nd century AD it took a lot of ingenuity. In order to preserve the circular motion of celestial bodies, which corresponds to the maximum symmetry and the ancient concept of harmony and aesthetic perfection, it was necessary to introduce so-called epicycles.

In the Ptolemy model, all the planets, as well as the Sun and the Moon move uniformly in circular orbits around the Earth. The center of orbit of each celestial body, in turn, moves around the Earth uniformly and also in a circular orbit. According to this model, the planet moves uniformly along a small circle called an epicycle the center of which, in turn, moves along a large circle which is called a deferent. Thus, the Ptolemy Universe was a set of mutually intersecting rotating spheres.

As a result, for a satisfactory coincidence with the experiment, Ptolemy took 77 epicycles and deferents. Despite the fact that in ancient times there were supporters of the heliocentric system, such as Aristarchus of Samos and Archimedes of Syracuse, the system of Ptolemy, consecrated by the Catholic Church, existed for one and a half thousand years.

The transition to the geocentric system, made by Nicolaus Copernicus in the 16th century, was also based not on elliptical, but on circular orbits of the planets. Therefore, for a satisfactory coincidence with the experiment, Copernicus needed to leave epicycles and deferents, but only 34. Such a reduction of entities immediately showed that the heliocentric system is better because it is simpler. For this reason, the Copernicus system relatively quickly gained a number of supporters.

The elimination of epicycles and deferents from the picture of the world became possible after observing Tycho Brahe and processing them with Johannes Kepler.

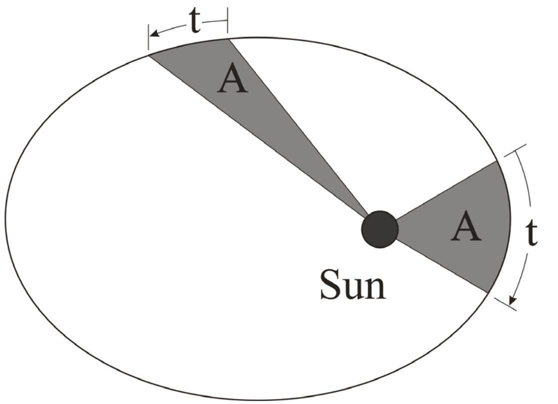


Fig. 1-4.
Kepler's second law. When the planet moves along an ellipse, the area of the sectors at equal intervals of time remains the same

The above axiomatics is certainly applicable in the natural sciences. In the humanities, it is not so categorical.

Chapter 2

Discretion and continuity in nature. Fundamental world constants and levels of the organization of matter

Philosophers believe that facts give rise
to ideas, and in a sense it is true.
But I find in the history of science the following:
in order to understand the facts you need
to have certain ideas in mind and
that the eyes cannot see what the mind sees.
Justus von Liebig

Even in ancient times, two mutually exclusive hypotheses about the internal structure of bodies were formulated. According to the first, a substance continuously consists of one or several primary elements. The second hypothesis stated that all substances consist of particles that are indivisible further – atoms. In fact, nature is more complicated than we used to think. We will talk about this at the end of this chapter.

This discrepancy was of fundamental importance both for the theory of knowledge and for science as a whole. If matter is continuous, then research tasks are significantly narrowed. It is not necessary to divide anything into elementary parts; we still get the same substance with the same properties. If the second hypothesis is true, then the task of the researchers is to study the properties of these atoms. It is also necessary to answer the question of how they are held together during the formation of various substances. Along the way, another problem arises. If we start to divide the substance into parts, until what moment does it retain its properties? If this problem is reversed, then in fact it means solving the next problem. How many atoms (or molecules) of a substance must be taken in order for it to exhibit properties known to us. This problem is in some sense similar to the ancient philosophical problem of the "heap" (grain and grain are two grains, another grain is three grains, etc. ..., and when is a heap?).

Having lasted more than 2500 years the dispute between the hypotheses was finally resolved only at the beginning of the XX century by the recognition of the atomistic concept. This became possible after the discovery of *X*-rays in 1896 by Wilhelm Röntgen. In the 1950s, Max von Laue, as well as the father and son, William H. Bragg and William L. Bragg, discovered diffraction on the atomic-crystalline structure.

The problem of atomism may seem quite trivial to the modern educated person. In fact, students already at school will learn about atoms and molecules, as obviously real objects. However, the idea of atomism is deeper than it is usually talked about at the popular level. It not only boils down to the problem of the discrete description of substance, but also the resultant property of matter. And it is unlikely that such high minds, since antiquity, dealt with such a trivial problem. We give a brief historical background. The atomists were Anaxagoras, Leucippus, Democritus, and Epicurus. They were opposed by Socrates, Plato, and Aristotle. In the Middle Ages, under the influence of the actually canonized teachings of Aristotle, the term "atom" disappears from use. In the new time, for the first time the corpuscular theory of the structure of matter was developed by Robert Boyle, having introduced the concept of a "chemical element as a simple body, not consisting of others". Then Antoine Lavoisier, John Dalton, and Amedeo Avogadro contributed. On the great hypothesis of Avogadro, we now stop.

The fact is that in 1808 Joseph Gay-Lussac found the law of simple volumetric relations. For example, two liters of hydrogen and one liter of oxygen give two liters of water vapor. This fact ($2 + 1 = 2$?) did not find an explanation in the atomic theory proposed in 1803 by Dalton. To save the atomic theory of Avogadro in 1811, he put forward a hypothesis that resolved this contradiction. To do this, he needed to introduce a new concept – molecules as compounds of atoms. Pay attention, all this was proposed at the time when the very existence of atoms was a hypothesis. He further suggested that the number of these new entities – the molecules – is always the same in the same volumes of any gases and always proportional to the volume. From here he concluded the law of Avogadro. At the same pressure and temperature, equal volumes of any gases contain the same number of structural elements. These are either atoms if the gas is monatomic, or molecules if the gas is polyatomic. This number of structural elements is the Avogadro number $N_A = 6.0227\ 10^{23}$ mol^{-1}. In the future, it is the order of this quantity that is significant for us, and not its dimension.

The Avogadro hypothesis consolidated the idea of the discrete structure of a substance in science although the official recognition of the hypothesis itself came only at the I International Congress of Chemists in 1860 four years after the author's death. The Avogadro number itself was calculated for the first time by Johann Loschmidt in 1865. It is also important that in the XIX century, a new concept arises – the amount of substance, on the basis of the twentieth-century Avogadro number is given the meaning of a sufficient condition for the realization of the macrostate.

Note that the necessary conditions in the general case do not exist. For some objects it can be a million structural elements, and for others it can be just a thousand. However, we certainly know that if an object consists of N_A structural elements, then it is always a macrosystem (i.e., a "heap").

Summing up, it makes sense to quote the Nobel Prize laureate Richard Feynman who believed that the atomistic hypothesis is exactly what you should take if in the future humanity has to forget all other knowledge.

Modern science gives the following idea of the hierarchical structure of matter. In the microworld, protons and neutrons are made up of quarks. They, in turn, form the nuclei of atoms. Atoms are made up of a nucleus and electrons. Atoms can be combined into molecules. Macroscopic bodies familiar to us consist of atoms and molecules. This is the macroworld. If we move up according to the hierarchy of scales, then we will move from the macroworld to the megaworld. The boundary of the megaworld conventionally begins with the planets. Further, the main objects of the megaworld in increasing order of size are stars, galaxies, and metagalaxies. Finally, the whole Universe is as a single physical object. Micro, macro and mega sized objects have a ratio to each other like this:

$$\frac{macro}{micro} \approx \frac{mega}{macro}.$$

In our book we must characterize not only each of these levels but also how such a hierarchy is carried out. What kind of "glue" makes these levels possible? It turns out that the most important characteristics of this "glue" are the fundamental world constants. As we shall see, there are not so many of these fundamental constants in physics, chemistry, and biology. At the present time we understand that even a very small change in these constants should lead to the formation of a qualitatively different world. In this world, the formation of existing micro-, macro- and megastructures would be impossible. In addition, the existence of highly organized forms of living matter would also be impossible. The problem of fundamental constants thus acquires, in conceptual terms, global worldview significance.

Chapter 3

Conservation laws as a consequence of symmetry. Properties of space and time

An unruffled build in everything
Consonant complete in nature,
Only in our ghostly freedom
Discord we are aware of it.
Fyodor Tyutchev

In this and the next two chapters we will demonstrate how the fifth axiom works on a number of very important conceptual examples.

We first dwell on two concepts: homogeneity and isotropy. We define them in relation to space.

1. A space is called homogeneous if its properties do not change with any parallel translation.

2. A space is called isotropic if its properties do not change with any rotation around a given axis.

These definitions reflect significantly different observable properties of space. In the first case we are dealing with a parallel translation operation. In the second case an operation of turn around the set axis is carried out. We note here that homogeneous and isotropic space has the maximum symmetry.

As a model of space, we use the desert, consisting exclusively of sand. Imagine now that we are launching into the desert two agents studying the properties of this space – the ant and the camel. What information about homogeneity and isotropy will our two agents give us?

The ant studies the desert for the property of homogeneity. For this, he moves parallel to himself for some arbitrary distance. Since the size of an ant is comparable to the size of grains of sand, he concludes that the desert is heterogeneous.

Next, the ant studies the isotropy of the desert. For this, it makes a turn at an arbitrary angle around its own axis. At the same time, he sees at a distance comparable to his size. Making such turns, he notes that the number of grains of sand in each arbitrary direction is on average the same. This means that the desert space for an ant is isotropic.

Consider the same situation for a camel. The camel does not feel separate grains of sand; therefore, the desert is homogeneous for it. On the other hand, he sees at a great distance and notices that there is a dune in one direction, but not in the other. Hence the desert for a camel turns out to be non-isotropic.

Thus, our two agents (each of whom is absolutely truthful under the terms of the problem) provided us with completely opposite information.

We give a physical example. If you light a crystal (an analogue of a desert) with visible light (an analogue of a camel), then it will not "feel" the structure of the crystal, and the environment will be homogeneous in relation to it. However, the light spreading in different directions "feels" the non-isotropy of the crystal. If the same crystal is illuminated with X-rays the wavelength of which is comparable to the size of atoms (an analogue of an ant), then it will "feel" the heterogeneity of the medium. But at large distances, the X-ray crystal is isotropic.

These examples lead us to the following statement. We receive any information with the accuracy of the agent. This principle is valid for any agents. However, agents can be subjective and objective. Since science claims objective knowledge, it is clear that we need objective agents. The question arises, are there any objective agents? If so, who or what is it? Such agents exist, and these are the laws of nature.

This statement is easiest to find out as follows.

Let's remember how many initial positions there are (we take them as axioms), i.e. physical laws, we need to know in order to solve school tasks in mechanics. First, these are the three Newtonian laws. Second, three laws of conservation: the law of conservation of mechanical energy, the law of conservation of momentum, and the law of conservation of angular momentum.

The law of conservation of energy states that the sum of potential and kinetic energy is a constant i.e. $E = mgh + mv^2/2$ is preserved. The law of conservation of momentum states that the value $\mathbf{P} = mv = const$ is conserved. The law of conservation of angular momentum states that the value $\mathbf{L} = [\mathbf{rP}] = const$ is conserved. ($\mathbf{L}$ is a vector product of two vectors $\mathbf{r}$ and $\mathbf{P}$.)

Thus, with this approach, in the Newtonian formalism there are only six initial positions or axioms. There are other formalism mechanics, for example, the formalism of Joseph-Louis Lagrange. There are only two axioms in this. Here we do not dwell on their formulation of these axioms. We also do not give the Lagrange equations and the derivation of the

Newton laws and conservation laws from them. All of this would require us to excessively mathematize the presentation.

Taking as a basis only two Lagrange axioms (instead of six axioms in Newtonian formalism), according to axiom 5 we should get either a fundamentally more correct idea, or something completely new as a reward for using fewer “entities”. Indeed, in the Lagrange formalism by means of appropriate mathematical transformations it is possible to obtain not only Newton's laws, but (and this is what is important for us) all three laws of conservation. Moreover, each of the conservation laws is now not an axiom (as in the Newton formalism) but a consequence of certain properties of time or space. More precisely, the conservation laws are a consequence of one or another symmetry of time and space. Specifically. The law of conservation of energy is a consequence of the homogeneity of time. The law of conservation of momentum is a consequence of the homogeneity of space. The law of conservation of angular momentum is a consequence of the isotropy of space. These three conservation laws are precisely those objective “agents” that answer questions about the homogeneity of time, homogeneity, and isotropy of space. That is, where the law of conservation of mechanical energy is executed, time flows uniformly. Similarly, with respect to the homogeneity and isotropy of space, the preservation of impulse is a guarantee of homogeneity of space, and the preservation of angular momentum is a guarantee of isotropy of space.

Thus, using the example of the transition from the formalism of the mechanics of Newton to the Lagrange formalism, we have seen that the use of fewer entities has led us to new knowledge. In addition, even with such a simple example, we have seen that the properties of symmetry are extremely important for the implementation of the laws of nature and, in particular, for the conservation of certain physical quantities. In the future, each time we will specifically focus on what happens to symmetry in the implementation of a particular law, because symmetry is also a peculiar language of nature.

It is necessary to clarify at least qualitatively the concept of symmetry. In the case when the state of the system (it can be a material object, process or equation) does not change as a result of any transformation to which it can be subjected, it is said that the system has symmetry about this transformation. In our book, we cannot describe in more detail the various types of symmetry. We will give just a few examples that are important for further discussion.

First example. We intuitively understand that inhomogeneous space has a lower symmetry than homogeneous space. Similarly, the transition from isotropic to non-isotropic space is also accompanied by a decrease in symmetry.

The second example. The chaotic state has a minimal order; therefore, it has a higher symmetry than an ordered state. Indeed, imagine a vessel divided by a movable partition. There is some gas in one part of the vessel. We take out the partition. At the first moment of time, our system is streamlined. In one part there is gas, in the other there is not. Over time, the gas spreads to the entire vessel. This second state is completely disordered, as chaotic as possible, and possesses a higher symmetry than the original.

The third example. There are two systems. The first one has an equal number of screws with left and right threads. In the second system are screws, for example, with only a left thread. Which system has a higher symmetry? It is easy to answer, if we imagine that we have a third system in which there are screws with only a right thread. Then it is clear that the first system, with an equal number of screws with right and left threads, is more symmetric than each of the other two.

We will need these examples in the future to explain extremely important patterns.

The above results on conservation laws and their connection with the symmetry of space and time are generalized in fundamental Emmy Noether theorems.

Noether theorems. The essence of Noether's results in a simplified form is as follows. The existence of characteristics of physical systems that retain their properties is directly related to the symmetries of action. Action is a fundamental physical quantity that determines the dynamics of a system. Thus, conservation laws are a direct consequence of the existence of certain symmetries of action. This result is a universal tool for finding such laws in various fields of physics. In addition, it is perhaps one of the most beautiful theoretical insights in the entire history of science.

In the first theorem we consider the consequences of the invariance of the action with respect to symmetries to which certain transformations correspond. They can be interpreted as laws of conservation of physical quantities that satisfy these symmetries. That is, these conservation laws are direct consequences of certain symmetries.

The second theorem describes situations when the action is invariant with respect to symmetry transformations which depend not on numerical parameters but on some arbitrary functions. The symmetries of the second Noether theorem are applied in fundamental physics. They make it possible to establish correspondences between the properties of particles and the fields with which these particles can interact.

CHAPTER 4

THE SPACE-TIME CONTINUUM AS A CONSEQUENCE OF A FUNDAMENTAL CONSTANT – THE SPEED OF LIGHT

There is no movement, said the wise man.
The other was silent and began to walk before him.
Stronger would could not he argue;
Praised all the answer intricate.
But, gentlemen, the curious case of this
The other on memory me leads:
After all, every day before us the sun goes,
However, well rights are stubborn Galileo.
Alexander Pushkin

We now turn to the not so obvious facts. Let us first give a brief historical background. In the 50s of the XIX century of James Clerk Maxwell on the basis of an incorrect model obtained four famous equations named after him. Maxwell's equations perfectly describe all the laws of electromagnetism and optics, previously obtained on the basis of numerous experimental data. Thus, all the laws of electromagnetism previously known and regarded as axioms can be obtained by appropriate mathematical transformations from just four equations. If a certain environment is considered, then another so-called material equation is added to them. What does the reduction of "entities" in this case lead to? To answer this question, we will continue our historical reference. Very soon physicists noticed that Maxwell's equations did not remain unchanged under the so-called Galilean transformations:

$$\begin{aligned} x &= x' + v_0 t \\ t &= t' \end{aligned} \quad . \tag{4.1}$$

Here x is the coordinate of the body in a fixed coordinate system, for example, the coordinate of Prof. Moriarty (who runs along a moving train) relative to a fixed observer who stands on the platform of the London station. The x' is the coordinate of the body in the moving coordinate system. That is, the coordinate of Prof. Moriarty is "attached" to the train; in other words, the train car number and the position on the train car itself. v_0 is the speed of the moving reference system. In our problem, this is the speed of the train in the direction of the x axis. Accordingly, the time t in a fixed system coincides with the time t' in a moving one. From the above formulas it is easy to get the wrong formula that we used in solving the problem of the speed of Prof. Moriarty running on the London-Paris train. Indeed, dividing the first equation by t, we get:

$$V = v' + v_0 . \tag{4.2}$$

For our task, v' is equal to the speed of Prof. Moriarty v_M, and v_o – to the speed of the train v_T.

If Maxwell's equations change under the "intuitively obvious" Galilean transformations (since formula (4.2) was considered obvious by us), then the question arises, What should be the other "non-obvious" transformations in order for Maxwell's equations to remain unchanged? Accordingly, what non-obvious formula for the speed addition is obtained from these non-Galilean transformations?

Henri Poincaré received these transformations and named them in honor of Hendrik Lorentz, the Lorentz transformations:

$$x = \frac{x' + v_0 t}{\sqrt{1 - v_0^2 / c^2}},$$

$$t = \frac{t' + v_0 x / c^2}{\sqrt{1 - v_0^2 / c^2}}. \tag{4.3}$$

In these new transformations, the value $c = 3 \cdot 10^8$ m/c is the speed of light in vacuum. From the Lorentz transformations follows the non-obvious formula for the addition of velocities:

$$V = \frac{v' + v_0}{1 + v' v_0 / c^2} . \tag{4.4}$$

It is absolutely clear that if $v_0 << c$ then the Lorentz transformations turn into Galileo transformations and the non-obvious formula (4.4) turns into the obvious formula (4.2). So what are still true, obvious (4.1) and (4.2) or non-obvious (4.3) and (4.4)? It turned out that the Lorentz transformations leave Maxwell's equations unchanged, that is they (and not the Galilean transformations) are correct from the point of view of electrodynamics and optics.

At first glance, it seems that the Lorentz transformations and the next law of speed addition are absurd. In fact, if we are interested in the *x*-coordinate, Prof. Moriarty running, then it is not clear why the speed of light is present in the first formula (4.3). It gets worse over time. According to the second formula (4.3), time not only depends on whether the observer moves or rests, but it turns out that it also depends on its location x and again on the speed of light. This is completely incomprehensible. The formula (4.4) is not clear either. Why, when solving such a simple kinematic problem as Prof. Moriarty, running on the train roof, should one take into account the speed of light?

In fact, the list of “incomprehensibility” is not yet exhausted. From transformations (4.3) one can get two more at first glance completely “absurd” results. It turns out that the linear dimensions of the body along the direction of motion are reduced compared with those for the fixed body, and time in the moving system slows down:

$$\begin{aligned} \Delta x' &< \Delta x, \\ \Delta t' &< \Delta t. \end{aligned} \tag{4.5}$$

These Lorentzian results (reduction of distance and deceleration of time) were a glaring contradiction to those ideas about the properties of space and time that had developed in science by the beginning of the XX century. However initially no further conceptual conclusions were made. The addiction to the paradigm of Galileo and Newton was too strong – space and time are absolute categories; they exist by themselves and do not depend on external circumstances.

Now it is useful to mention one anecdotal fact. At the end of the 19th century the then young student Max Planck, came to one of his professors for advice. What field of science should he choose in the future? The venerable professor did not advise Planck to engage in theoretical physics, since he believed that almost all of its fundamental problems had already been solved. However, this venerable professor noted that there are really two small clouds in the pure sky of theoretical physics. One of them is not being entirely clear as to what is happening with the measurement of the speed of light. The other is being not quite clear about the problem with the emission of an absolutely black body.

Very little time passed, and two small clouds gave rise to hurricanes. One of them is a special and general theory of relativity. Planck himself played a fundamental role in the creation of the second, which is quantum theory. Well, let's start in order.

Incomprehensibility over the measurement of the velocity of light was as follows. Since the experiments of Augustin Jean Fresnel and Thomas Young when such wave phenomena as interference and diffraction were first
established for light, it has been clear that light has a wave nature. According to the ideas of the XIX century each wave process should be distributed in the appropriate environment. For light waves, a certain mythical ether was considered such a medium. But that was not clear. Unlike other media whose properties are understandable and relatively constant this ether behaved very strangely.

To clarify the properties of this mysterious ether first Albert Michelson and then Michelson together with Edward Morley in the period 1881-1887 carried out a series of high-precision experiments on a specially designed instrument – the Michelson interferometer.

Schematically and in a very simplified form (for our purpose such a simplification is quite acceptable) this device is shown in Fig. 4-1. It consists of four mirrors, two of which are parallel to each other in pairs. The essence of the experiment was that two parallel mirrors were installed strictly along the earth meridian, and the other two were installed strictly along the earth parallel. A light beam was launched between the two pairs of mirrors to measure the speed of light. The rotation of the Earth does not affect the movement of light along the meridian, so its speed is equal to c. When moving along the parallel, the rotation of the Earth should seem to be felt. When light moves from west to east the direction of the speed of light c coincides with the speed of rotation of the Earth v_E therefore the measurement should seem to give the value $c + v_E$. For speed in the opposite direction, following this logic, we must get the value $c - v_E$ in full accordance with formula (4.2). But, it was precisely this that was "incomprehensible". In that, and in the other direction as well as along the meridian, the experiment gave the same value of the final speed equal to c.

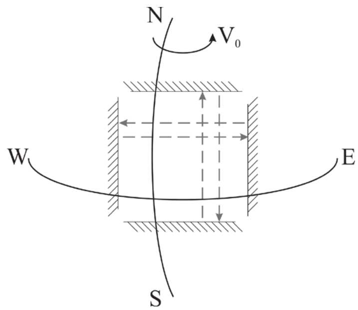


Fig. 4-1.
Michelson interferometer schematically depicted on the surface of the earth. At the speed of light spreading along the meridian, Earth's rotation has no effect. At the speed of light spreading along the parallel, Earth's rotation again has no effect, despite the obvious result of the task about the “London-Paris” train

In order to preserve the Galilean transformations and explain the experimentally observed properties of the speed of light, physicists invented a number of unique properties for the ether (“entities”). This example can serve as an illustration of ignoring axioms 1 and 5. Traditional ideas about the properties of space and time were so strongly rooted that initially other ideas were not considered. The decisive change of concepts about the properties of space and time was carried out by Poincaré in 1898 and then by Albert Einstein in 1905. It is characteristic that in this work Einstein clearly did not deny the existence of the ether. He built a new conceptual theory, and the ether with all its “entities” was simply thrown away by default.

In the new theory only two new axioms were required (again, the transition from many “entities” to a smaller number). The first axiom is the principle of relativity. All physical phenomena under the same initial conditions occur equally in all inertial reference systems. This principle was first formulated by Poincaré in 1889. The second axiom is the principle of invariance of the speed of light. The speed of light is finite and the same in all inertial reference systems. It does not depend on the speed of the source and is the maximum speed of propagation of any signal (interaction). These axioms were laid by Einstein as the basis of the special theory of relativity (STR). This theory led to a profound rethinking of the concepts of space and time.

It is important to note that the existence of the maximum propagation velocity of interactions is a necessary consequence of the fundamental principle of causality.

In 1908 Hermann Minkowski proposed a fundamentally new geometric interpretation of the STR results. At the same time, he naturally relied on the results of the work of Poincaré and Lorentz on the study of the symmetry of Maxwell's equations. According to the idea first proposed by Poincaré, a four-dimensional space-time interval was introduced into the STR:

$$R^2 = c^2t^2 - x^2 - y^2 - z^2 . \tag{4.6}$$

Expression (4.6) is an analogue of the Pythagorean theorem in four-dimensional space. It is of fundamental importance that the "temporal addend" c^2t^2 has a different sign than the spatial coordinates. Physically this means that the speed of light is the maximum possible speed of movement. Indeed, if we exceed the speed of light then the value of R^2 will become negative, i.e. the "distance" R in this space will be imaginary. We note here that Einstein didn't like this geometric interpretation at first. However, several years later he gladly accepted it to achieve other even more interesting results. Further, the STR is constructed on the basis of the following requirement. It is necessary that the interval (4.6) (in accordance with the second axiom) for all transformations of coordinates and time remains constant. Such transformations can be described as rotations of a four-dimensional coordinate system. This is the Lorentz-Poincaré symmetry. As a result, as we know, Lorentz's transformations are obtained where one can clearly see (the second relation of formula (4.3)) that time t and space x are not independent. If the speed of light were infinite, space and time would exist independently of each other. It took the mathematical genius of Poincaré and the physical understanding of his ideas by Einstein in order to fully realize this connection and to understand that space and time do not exist independently of each other. They are inextricably linked through a certain symmetry. This Lorentz-Poincaré symmetry is not just abstract mathematics, it occurs in the real world realized through motion. It is now clear that the existence of a four-dimensional space-time continuum is a consequence of the finiteness of the speed of any interaction which is bounded above by the speed of light.

Now it is clear that formula (4.2) is fundamentally wrong since it does not take into account the space-time relationship. In addition, from this formula it is impossible to get the wonderful effects of the STR which at first glance are contrary to common sense, such as for example, reducing the distance and slowing down time. One of the fundamental achievements of the STR was the famous formula linking mass and energy:

$$E = mc^2 . \tag{4.7}$$

Surprisingly this formula, regardless of Poincaré and 15 years before Einstein, was obtained by Oliver Heaviside. However, this is not the only result of Heaviside, though it was much ahead of its time and was obtained by him from considerations that are unknown to us.

Pay special attention to the fact that the lesson presented by Lorenz and Poincaré is the following. Mathematical research in this case based on the analysis of symmetry can be a source of outstanding achievements in science. Even if mathematical symmetry cannot be visualized it can point the way to revealing new fundamental principles of nature. In the presentation of the material below we will each time specifically focus on the value of one or another symmetry which determines the fundamental laws in inanimate and living nature.

CHAPTER 5

GEOMETRY OF SPACE-TIME. GRAVITY AS A CONSEQUENCE OF GEOMETRY IN THE EINSTEIN PARADIGM

Blessed is who visited this world
In his fatal minutes!
He was called the all-good
As an interlocutor at the feast.
He is their high spectator,
He was admitted to their advice –
And alive as a celestial being,
From the cup of their immortality drank!
Fyodor Tyutchev

The most important consequence of the STR is the replacement of absolute space and time by a new physical entity – the unified space-time of Minkowski (r, t). However, this space is essentially an extrapolation of the classical three-dimensional space into four dimensions and it therefore has a passive character. That is, it does not have the opposite effect on the physical processes occurring in it.

It is characteristic that the Minkowski space is Euclidean, i.e., flat (it has zero curvature). And this is understandable. In the STR, only inertial reference systems (moving in a straight line and uniformly relative to each other), which are not affected by gravitational forces, are considered. This is why the space of Minkowski is a definite physical abstraction, because no screen can protect you from gravity.

Now we turn to a less well-known story; the creation by Einstein of the general theory of relativity (GTR). The most striking fact here, from the point of view of the theory of knowledge, is perhaps that Einstein was looking for one thing, but found something completely different. And if several researchers almost simultaneously participated in the creation of the STR, only two participated in the creation of the new theory of gravity: David Hilbert and Einstein.

For many researchers, Einstein's creativity remained a mystery for a long time, how he passed from the STR to the GTR between 1905 and 1916. This riddle was clarified by Alfred Kastler at the conference dedicated to the 100th anniversary of Einstein's birth (Jerusalem 1979). It turned out to be an intermediate stage associated with a happy event. A little-known and hard-to-reach work has been published in German in the *Quarterly Journal of Forensic Medicine and Health*. The article reflects the search phase of the study and was dedicated to the anniversary of Einstein's friend, a doctor by profession.

In this article Einstein analyzes the behavior of light in a gravitational field. At the same time, he still uses (which is natural in 1909) the Newtonian theory of gravity. Since earlier, in the STR, he found that mass is a new component of energy (formula (4.7)), he believed that it was this energy that was associated with gravity, i.e. it serves as a gravitational charge. Further, Einstein comes to the conclusion that although the beam of light, which carries only momentum and angular momentum, has no mass, it nevertheless carries kinetic energy. Therefore, it must fall in the gravitational field, that is, attract and deviate (Fig. 5-1). This is only part of the result that he will receive in the new theory of gravity in 1916. Deviation, he argued further, implies a change in the speed of light, which should acquire the side component. Therefore, the light must accelerate in its movement to the source of gravity and slow down after it passes. And here we have a clear case of unusual luck.

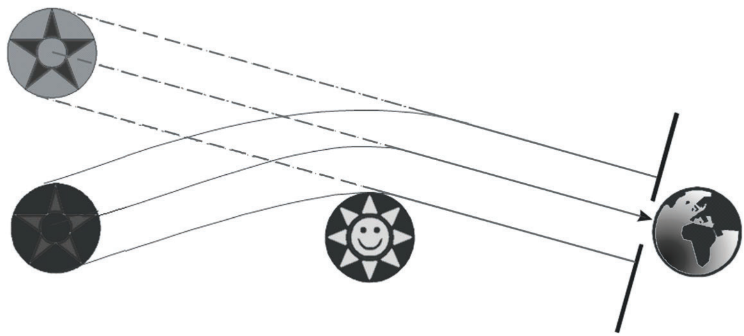


Fig. 5-1.
Passing near the Sun, the beam of light from the star deviates markedly due to the sun-induced curvature of space. As a result, observed on Earth the position of the star in the sky is offset from its real position

However, what about the constant speed of light? In connection with this problem Einstein turns to Minkowski's geometric interpretation because it can be the starting point for solving the problem of gravity. At the same time the paradigm about the limiting rate of any interaction is preserved. Einstein turns to his former classmate at the University Marcel Grossmann for advice. He is interested in the next question. Do four-dimensional geometries exist in which the analogue of the Pythagorean theorem holds only in small domains? It is clear that if such geometries exist, they will be non-Euclidean because he understands that the bending of a ray of light is connected with the curvature of space-time. Marcel Grossmann answered in the affirmative, pointing out the geometries of the curved spaces of Carl Gauss and Georg Riemann. From this point on Einstein focused his efforts on creating a new geometric theory of gravity – that is, not at all on the goal that he had originally set for himself. That's actually the whole little-known story about the intermediate work between the STR and the GTR.

In order to describe the gravitational forces, it is necessary to abandon the idea of the flat Euclidean space, and go to the geometry of the curved space. It is necessary that this geometry is determined by something. Therefore, we must abandon the independence of the properties of space and time from gravity.

Summarizing the classical concepts of space and time, as well as the nature of gravity Einstein declares a new paradigm. The gravitational field is a change in the geometric properties of space-time which, in turn, is determined by the distribution of masses. Moreover, all laws are manifested equally for any observer. Speaking in simplified mathematical language this means that the equations describing the laws of nature must be covariant. This means that the values included in these equations must be consistently transformed when moving to another frame of reference. Since the GTR is a more general theory than the STR it should include the STR as a special case when there is no gravity. This statement is a special case of the general principle of the development of science – the principle of conformity when a more general theory includes a particular theory as a certain limiting case.

Since the geometry of space-time in the presence of gravity should be non-Euclidean, then from the two variants of non-Euclidean geometry – hyperbolic and elliptic, you need to choose one. For the first geometry the sum of the angles of the triangle is $<180°$, for the second $>180°$. For the first, the ratio of the circumference of a circle to the diameter is $>\pi$, for the second, $<\pi$. Let us explain this with the simplest example of Riemann's elliptic geometry.

Consider the surface of a sphere (analogous to a plane in Euclidean geometry), see Fig. 5-2. Here, the “straight lines”, i.e. the shortest distance between two points, are arcs. Lines *A* and *B* (they are perpendicular to the equator) intersect at the pole *N*. Thus, the sum of the angles of the spherical triangle *ABN* is >180°. In the equatorial plane the ratio of the circumference to the diameter is $L/D = \pi$. On a sphere for the same circle with a diameter (the shortest distance between opposite points) will be the arc *CND* which is naturally larger than the diameter of the equatorial circle *CD*. Thus, for spherical geometry the ratio of the circumference of a circle to a diameter is $L/D < \pi$.

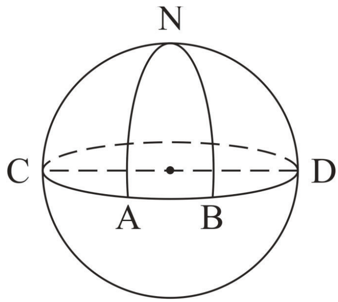


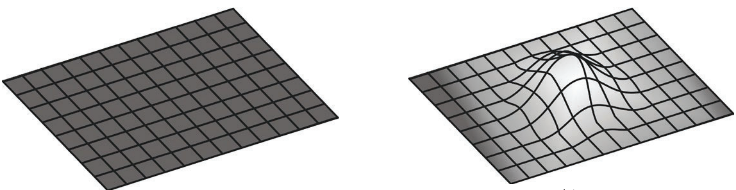
Fig. 5-2.
Illustration of the metric relations geometry of Riemann (sphere) and Euclid (plane)

We now consider the violation of the Euclidean metric in a non-inertial frame of reference. Let the circle rotate uniformly with respect to its center. We divide it mentally into small rectilinear elements. During rotation all rectilinear elements of the circumference of a circle experience a Lorentz contraction but the diameter does not change. Thus the total length of a rotating circle is less than a fixed one. Consequently, for a rotating circle (this is a non-inertial reference system) the ratio of the circumference of a circle to a diameter is $l/d < \pi$, and therefore the geometry of such a space is elliptic. Similarly, over time, it slows down.

In the GTR these effects are due to the distribution of masses in space which determine its geometry. Both effects have been confirmed experimentally. The first is when the trajectory of a ray of light coming from a star and passing near the Sun is curved. In the new paradigm the ray moves along its natural trajectory – the geodesic line which is the shortest distance in this space. The second is when propagating in space, it will change the frequency, i.e. the number of oscillations per second near the mass. So, as you move away from the massive body, the frequency will decrease, and as you approach it, it will increase. Consequently, space is bent near the gravitating mass, and time slows down.

Let's go back to Fig. 5-1 and imagine that there is a “tube” of light from the star. Since both beams of light in this tube (external and internal) arrive at the Earth simultaneously, and the path for the external beam is longer than that of the internal one, it becomes clear that the speed of light for the external beam is greater than that of the internal one. Thus Einstein came to the conclusion (only, at first glance, contrary to the postulate of “constancy” of the speed of light) that near the gravitating masses the speed of light is less than away from them. In other words, where the space is curved more strongly, the speed of light is less. The maximum speed of light corresponds of course to a flat space with a Euclidean geometry.

The movement of masses in space also changes its geometry. Let us give a vivid illustrative example (Fig. 5-3). Imagine that you put a rubber cloth on the table, firmly fastening the edges, and drew on it a series of mutually perpendicular lines (Euclidean space). Then you took the cat and shoved it under the tablecloth. Where the cat is, the tablecloth is stretched, and instead of straight lines you see intersecting arcs. If there is also a mouse under the tablecloth you will notice that the stretching, and therefore the curvature of the original lines in the location of the cat, is greater (this is a large gravitating mass) than where the mouse (the smaller mass) is. Further change to the geometric properties of the space of the tablecloth in the process of moving the cat and mouse is easy to imagine.



a) b)

Fig. 5-3.
Flat and curvilinear surfaces:
a) There is neither a cat nor a mouse. This is Euclidean geometry.
b) The mouse ran away, only the cat remained (one gravitating mass). This is a non-Euclidean geometry

Unfortunately, not all of the GTR results can be presented so clearly. We present the most interesting and important conclusion from the GTR.

Above we talked about reducing the speed of light when space is curved near a gravitating mass. Imagine now that the mass becomes so large and the curvature is so strong that the speed of light in this area of space becomes zero. Light, which by its property always moves, suddenly stops moving! If this is possible, then the light that has flown into this area of space cannot leave it. This area of space emits nothing and becomes black for the observer. This forms a kind of object called a black hole. Let's look at it from a mathematical point of view.

According to the GTR Newton's law of gravity should be changed as follows:

$$F_E = \frac{F_N}{\sqrt{1-2GM/Rc^2}},$$

$$F_N = \frac{GMm}{R^2}. \tag{5.1}$$

where $G = 6.67 \cdot 10^{-11}\ m^3 \cdot s^{-2} \cdot kg^{-1}$ is the constant of gravitational interaction first introduced by Newton in “The Mathematical Principles of Natural Philosophy” in 1687.

This formula follows from the particular solution of the Einstein equations, obtained for the first time by Karl Schwarzschild for a single star. Note that F_E becomes infinitely large for the so-called Schwarzschild radius R_S:

$$R_S = \frac{2GM}{c^2}. \tag{5.2}$$

A black hole is formed at such a radius. In the area of a black hole the space-time continuum is so curved that not only can the signal or object that falls into it not go outside, but time relative to an external observer is stopped. For Earth the Schwarzschild radius is 0.4 cm, for the Sun, 3 km, while their usual radii are $6.4 \cdot 10^3$ km and $7.7 \cdot 10^6$ km, respectively.

Until recently the existence of black holes has been questioned since their direct observation is impossible. However, literally in our days a significant event has occurred. The “shadow” of a giant black hole in the center of one of the galaxies was experimentally recorded. Thus the GTR received another experimental confirmation.

In 1929 Edwin Hubble experimentally discovered the currently existing expansion of the Universe. The speed of galaxies receding from each other (according to Hubble's law) is proportional to the distance D between them:

$$V = HD \tag{5.3}$$

where $H \approx (3 \div 5) \cdot 10^{-18}$ s^{-1} is the Hubble constant.

This Hubble extension is quite remarkable. Despite the fact that the universe is expanding it has no center of expansion! You can understand this on a two-dimensional model. Imagine that you have inflated a little ordinary balloon. Then randomly with a felt-tip pen you put dots on its surface, after which you continue to inflate the balloon. What do you see? The surface of the ball is stretched (analogous to the expansion of space), and each of the points marked with a marker pen moves away from the others. Thus you can conditionally take any point as the center of the extension from which all the others run. Such an infinite number of centers of expansion indicates that there is no single center of expansion on the surface of the sphere. In addition, it can be seen that the points themselves "sprawl" as the surface expands. Thus during the Hubble expansion of the Universe, space itself expands and stretches.

Seven years before Hubble in 1922, Alexander Friedmann was solving the Einstein equations of the GTR and proceeding from the condition of homogeneity of the Universe, when he came to the conclusion that it is possible to change the boundaries of the Universe. They can either expand or compress depending on the ratio between the average density of the Universe ρ_{av} and a certain critical value of the density $\rho_{cr} = 3H^2/8\pi G$. If $\rho_{cr} \geq \rho_{av}$ then the Universe is open and will expand all the time. If $\rho_{cr} < \rho_{av}$ then the Universe is closed and at some point, the expansion will be replaced by compression. At present we cannot give an unequivocal answer as to which of the inequalities between the densities ρ_{cr} and ρ_{av} is realized since part of the substance of the Universe is apparently in the "non-radiated" state (black holes, neutron stars, and dark matter). Modern astronomical data indicate that the observable Universe is flat within the limits of measurement error. According to the Friedman model this means that $\rho_{cr} = \rho_{av}$ with an accuracy up to 1%. Today there thus is no unambiguous choice of a model, and therefore, a scenario for the development of the Universe. We note the role of the world constants *G* and *H* in the scenario of the evolution of the Universe through which the fundamental parameter ρ_{cr} of the theory is expressed.

In 1998 it was found that the Universe is currently expanding with acceleration. In 2011 Saul Perlmutter, Brian Schmidt and Adam Riess received the Nobel Prize for this discovery. From the conclusion of the accelerated expansion of the Universe, a refinement of its evolution follows. It is most interesting that with some restrictive assumptions, it is proved that it is impossible in principle to achieve thermodynamic equilibrium in a Universe expanding with acceleration (Ignat'ev 2013).

The existence of gravitational waves was predicted in 1916 by Einstein on the basis of the general theory of relativity. The discovery of gravitational waves was carried out by their direct detection in 2015 by the LIGO and VIRGO collaborations. For this discovery, Rainer Weiss, Barry Barrish, and Kip Thorn were awarded the Nobel Prize in Physics in 2017.

Chapter 6

We enter the microcosm. The Planck constant and the wave of de Broglie. Heisenberg's uncertainty principle. Spin

Life is like a wall painting, created by you,
But the picture of the absurdities of the strange is full ...
Omar Khayyam

It all began in 1900 when Max Planck first introduced a new fundamental physical constant. This constant determines the value of the minimum action and is numerically equal to:

$$h = 6.63 \cdot 10^{-34} \text{ J} \cdot \text{s}$$

In classical physics such quantities as momentum – p, energy – E, and action – S, change continuously, i.e. by an arbitrarily small amount. However as soon as we "go into the microcosm", i.e. we are interested in objects whose sizes are less than 10^{-7} cm, the situation changes radically. In the microcosm, these values can vary both continuously and discretely, i.e. as a multiple of Planck's constant. It thus turns out that not only is matter discrete but also a range of physical characteristics that describe it.

Planck's idea of the discreteness levels of energy E

$$E = nh\nu, \tag{6.1}$$

where n – natural number and ν – radiation frequency, led to a formula for the radiation of an absolutely black body which perfectly corresponded to the experiment. On the other hand, this contradicted classical ideas about the properties of physical quantities that had developed by the beginning of the XX century, and therefore required a detailed analysis. Poincaré in 1911, after conducting a mathematical study of this question, showed that the quanta hypothesis is the only fundamentally new hypothesis that leads to the Planck radiation law. If the discreteness is slightly broken, i.e. n is not exactly equal to the natural number then there will be no Planck formula, and in general a whole class of problems on radiation theory simply cannot be solved.

The following principle for understanding the essence of the microcosm was formulated by Louis de Broglie in 1924. Its essence is that any free particle with an impulse p can be associated with a certain wavelength λ:

$$\lambda = \frac{h}{p} \tag{6.2}$$

Since then the concept of the de Broglie wave has appeared.

Thus, moving microparticles (electrons, neutrons, protons, and even whole atoms) have a wave essence and can demonstrate such pure wave phenomena as diffraction and interference. Based on this fruitful de Broglie idea Erwin Schrödinger proposed the wave equation which is the foundation of all of quantum mechanics.

The solution of the Erwin Schrödinger equation is the wave ψ-function, which describes the state of a quantum system. It is interpreted as the probability density of the implementation of a particular state and is not observed explicitly. The square modulus $\psi\psi^*$ of the wave function ψ is experimentally measured (ψ^* is a complex conjugate to ψ). The value $\psi\psi^*$ gives the probability distribution of finding a particle in a certain region of space. Thus, in the microcosm the picture of the description of objects is fundamentally probabilistic. The most important property of the microcosm is that particles have no concept of a trajectory in the usual macroscopic sense. A reflection of this is the uncertainty principle of Werner Heisenberg formulated in 1927. According to this principle the impulse uncertainty Δp_x (along the x axis) and the coordinates in the same direction Δx are related by the relation

$$\Delta p_x \Delta x \geq h \tag{6.3}$$

This means that neither the coordinate nor the momentum can be accurately measured at the same time, but only up to the value of the Planck constant h. The effect of this principle extends to other physical quantities that cannot be measured simultaneously. As a result, speaking, for example, of the orbits of electrons in an atom, we must understand that this is just a tribute to history – the planetary model of the atom. In fact, electrons of course do not rotate in any orbits. They are in certain quantum states. The complete quantum state, which describes the wave function of all electrons in an atom as a single system, does not allow it to be divided into states of individual electrons. However, within the framework of an extremely simplified model it can be said that some electrons are closer to the nucleus while others are farther away. At the same time orbits, i.e. fixed trajectories simply do not exist.

Questions may arise, how can all this be explained, and why is there such a "ridiculous" picture in the microcosm? Maybe we do not fully understand here? In other words, these principles are the result of some other laws of nature that are unknown to us, or are they themselves peculiar axioms that do not need to be justified? At the present level of development of science, we must recognize that there are no other justifications for these principles. Paraphrasing the words of Titus Lucretius Carus, we must say that in the microcosm "That's the way of things" and it should be limited to this.

In the quantum world not only can the position of a particle in space not be determined precisely; the same applies to its orientation in space. In quantum physics, each particle is assigned a special mechanical moment of its own, which is not associated with its movement in space or with rotation. This own moment is called spin. It is spin that determines the orientation of a particle in space. Let us show how the electron spin manifests itself.

In the world habitual to us, when turning around an axis by 360° everything will look exactly the same as it was before the start of rotation, i.e. we find ourselves in the same condition. Based on everyday experience it is natural to expect that the electron, when rotated 360° will return to its original state. However, this is absolutely not the case! Since the electron spin is not zero, then to return to its original state, it must be rotated again by 360º i.e., eventually at 720º. Therefore, the same space exhibits different properties in relation to us and to the electron. Fig. 6-1 gives a simple illustration of this.

Fig. 6-1 shows a double loop with a bead strung on it. At a great distance, we cannot distinguish between two turns, and it seems to us that the wire is simply folded into a circle. Therefore, one turn of the bead is perceived by us as a return to its original state. But in fact, to return to its original state, the bead needs to make another turn. Spin is a fundamental property of nature. This fact is a consequence of the nontrivial symmetry properties of space-time.

Spin always takes discrete values that are multiples of $h/2$. In particular, the absolute value of the spin of an electron is $h/2$. Since spin is discrete and a vector, this leads to the fact that only two mutually opposite spin orientations are possible for an electron. From here extremely important consequences follow. There can only be one electron in the same quantum state. This statement is a special case of the general Wolfgang Pauli prohibition principle. This principle is valid for particles with half-integer spin. The quantum state is understood as a set of quantum numbers. For example, for a hydrogen atom this state is determined by three characteristic quantum numbers: principal, azimuthal, and magnetic, taking discrete values in fractions of the constant *h*. The next in the Dmitri Mendeleev periodic table behind the hydrogen atom is the helium atom. Accordingly, in the helium atom, the second electron in the same state is obliged to change the orientation of the spin to the opposite, i.e. the spin should be $-h/2$.

Fig. 6-1.
A double loop with a bead on it gives a schematic representation of the properties of the spin of an electron. To return to its original state requires another turn

The Pauli prohibition principle leads to the regularities of filling with electrons of quantum states in an atom. This explains the frequency of changes in the chemical properties of the elements in the Mendeleev periodic table. Note that if the electron spin were half-integer, but having a different meaning, for example, $3h/2$ or $5h/2$ then the periodic table would look completely different. Therefore, chemistry would be completely different. In this case, it is not obvious whether life could arise and exist.

On the other hand, if the electron spin were a multiple of an integer h, then any number of electrons would be in the same state. In this case, the atoms would not differ in their chemical properties. Such atoms could not form molecules, which means there would be no chemistry, and, as a result, no life.

CHAPTER 7

FOUR FUNDAMENTAL INTERACTIONS: GRAVITATIONAL, ELECTROMAGNETIC, WEAK AND STRONG

Nature – Sphinx. And so it is more true
He ruins a man with his skill
What could be, no from the century
There is no riddle and she did not have it.
Fyodor Tyutchev

In the microcosm completely new concepts appear that are absent in the macrocosm. All elementary particles of a substance are divided into leptons (light) and hadrons which in turn are divided into mesons (medium) and baryons (heavy). There are only six leptons, these are the electron, muon, tau and lepton and the corresponding neutrinos. Leptons have no internal structure; they are the most elementary particles.

Unlike leptons, hadrons are much larger, and there are several hundred of them. Hadrons have special characteristics that are absent in leptons. For example, baryons have a quantum number called the baryon charge. There is a law of conservation of baryon charge due to which an electron cannot interact with a proton so that their charges are destroyed.

To describe the properties of hadrons other quantum numbers are also introduced: strangeness, isotopic spin, and for quarks, flavor and color charge. The particles, which we now call quarks, were introduced by Murray Gell-Mann in 1963 and independently of him, by George Zweig in 1964. They solved the problem of the systematization of hadrons that was known by that time.

For this Gell-Mann invented three hypothetical particles with a fractional value of the electric charge (2/3 and 1/3 of the electron charge). He borrowed the name from James Joyce's novel “Finnegan’s Wake” where one of the characters has a fantastic dream in which gulls fly and shout: “Three quarks for Muster Mark!”. Later, three more quarks were introduced so now there are six. Quarks have six flavors: up, down, strange, charm, bottom, and top. Each quark with its flavor has three more quantum numbers – color charge. There are three colors: red, green and blue. Like any elementary particle each quark corresponds to an antiquark. When a particle and an antiparticle collide, they are mutually destroyed (so-called annihilation), and their disappeared total mass is released as radiation energy according to formula (4.7).

It is significant that in no experiment are the quarks themselves with their fractional electric charge directly recorded. Only the conclusions from the theory of quarks are experimentally confirmed, i.e., if they exist then there must be a certain result in a specific nuclear reaction. The existence of quarks is thus confirmed not directly, but only indirectly. This is what we must take into account when applying axiom 2 in order for science to exist and develop.

Imagine that you met with representatives of an extraterrestrial civilization. You need in the shortest possible time to show them that you are thinking beings. In addition, our earthly civilization has achieved certain successes in comprehending nature. It is clear that it is necessary to give such information in such a way that it is understandable to any thinking beings that are approximately at our level of development. To teach them our language is hopeless, for it would take too long. You can, of course, draw proof of the Pythagorean theorem but this result was already known two and a half thousand years ago. This is a good but very low level. You can teach them our decimal number system, and then write some fundamental constants showing the level of achievement of the Earth's civilization. This is a promising way but which constants to write? The speed of light, the gravitational constant, the Planck constant, and many others, such as, for example, the mass of a proton (the most stable particle in the Universe), have a dimension. We have no opportunity to explain to them about our kilogram, meter, etc. You can, of course, write the Avogadro number. This constant is much younger than our Universe, and its order actually reflects its meaning. It should be clear to any other civilization, but this is a result from the XIX century. No, this Avogadro constant is good as the Pythagorean theorem is just for starting a conversation. Well, what about the leading edge of the development of science, or almost the front? It is this frontier that we must now approach.

Gravitational interaction. We are familiar with the scientific description of gravity from 1687, the time of Newton. The first laboratory measurement of the gravitational constant G was carried out in 1774 by Henry Cavendish. He measured an extremely weak force of attraction between two metal balls. It is characteristic that gravity has an infinite radius of interaction and all bodies are subject to it. To protect yourself from gravity by any screen is impossible. Thanks to this interaction, our solar system and other planetary systems and galaxies exist. In short, thanks to him, we are able to observe the whole megaworld.

Electromagnetic interaction. We also know about this interaction from school. Apparently the existence of electricity was first established by Thales of Miletus when he rubbed a piece of amber (in Greek, an electron) with silk or fur. Magnetism was also experimentally discovered by the ancient Greeks. They knew the properties of magnetic iron ore as early as 600 BC. After about 500 years the Chinese discovered the ability of this material to orient themselves in space and they actually created a primitive compass. However, due to the lack of the concept of the “law of nature” in ancient China, its use was limited only to various mystical actions. Only a few centuries later the compass became a navigation device. In the XVIII–XIX centuries, the nature of electricity and magnetism was gradually cleared up. As you know, the apotheosis was the writing by Maxwell of his four equations which combined electricity and magnetism into a single theory. Electromagnetic interaction has an infinite radius of interaction. According to electromagnetic interaction a negatively charged electron is attracted to a positively charged nucleus consisting of protons and neutrons. As a result, the electron does not fly away from the nucleus, making possible the very existence of the atom. Thus this interaction like the gravitational one forms our atomic-molecular world (including ourselves).

Strong interaction. This concept evolved as the structure of the atomic nucleus became clearer. According to the law of Charles-Augustin de Coulomb protons such as like-charged particles would have to fly away from the nucleus as gravitational forces are not enough (they are extremely small compared to electric ones) to keep protons in the nucleus region of 10^{-13} cm. However, atomic nuclei exist. This means that something must keep protons in the nucleus, i.e., a strong interaction. This is significant only at distances of about 10^{-13} cm, i.e. at short-range. It is clear that it also defines the existing world, since it is responsible for the stability of the nuclei, and, as a result, of the atoms themselves. In addition, in the depths of the Sun and the stars, thermonuclear reactions are continuously proceeding, caused by strong interactions. Thanks to these reactions, energy is released that provides life on Earth.

Weak interaction. Apparently without realizing, mankind witnessed this in 1054 when Chinese astronomers saw the appearance of a bright blue star in an area of the sky where they had not observed anything before. This new star appeared for several weeks, and then began to slowly go out. This outbreak of 1054 is considered a supernova explosion, i.e. a gigantic explosion of an old star caused by the sudden collapse of its core accompanied by the short-term emission of a huge number of neutrinos Neutrino particles only participate in weak interaction. Thus, neutrinos are an attribute of weak interaction. Now the supernova of 1054 is observed in the form of a hazy bright spot in the constellation Taurus. Weak interaction is noticeable at even smaller distances than strong interaction. It ceases to act at distances of 10^{-16} cm from the source. However, without it, there would be no existing world, since it causes the transformation of some particles into others, often setting the reaction products into motion at high speeds. In addition, weak interaction is responsible for the relatively slow and steady burning of our Sun. To a certain extent this also provides the kind of life that has been realized on Earth.

Each interaction is characterized by its constant which has a corresponding dimension. However, if we are talking about the formulation of a unified scientific picture of the world from the micro to the mega level inclusive, it is necessary to find dimensionless constants. If these exist, then they should truly be called world constants. They are actually the world language of communication for "all times and peoples". The procedure for obtaining dimensionless constants for the four fundamental interactions is well known in physics. In addition to the gravitational constant G and the electron charge e, it is necessary to use other fundamental constants. These are the speed of light c, the proton mass m_p, the color charge q_s and the g_F Enrico Fermi energy. The result is the following dimensionless quantities which you should communicate to representatives of an extraterrestrial civilization:

the constant of gravitational interaction – $\alpha_g \approx 10^{-39}$;

the constant of weak interaction – $\alpha_w \approx 10^{-5}$;

the constant of electromagnetic interaction – $\alpha_e \approx 10^{-2}$;

the constant of strong interaction – $\alpha_s \approx 1$.

It is characteristic that the numerical values of these constants (despite the possibility of their change) cannot be changed without destroying the stability of one or several basic structural elements of the Universe. We can assume that these constants are stable starting from a time of 10^{-35} seconds from the moment of the birth of the Universe. Such an accurate and stable "fitting" of the numerical values of the world constants necessary for the existence of nuclei, atoms, stars and galaxies is completely unclear. But it is this "fitting" that determines the existence of not only complex inorganic, organic and living structures, but ultimately *homo sapiens*.

CHAPTER 8

SCENARIO OF THE EVOLUTION OF THE UNIVERSE

What is this universe?
From what does it arise?
Into what does it go?
In freedom, it rises,
in freedom it rests,
and into freedom
it melts away
The Upanishads
Translation by Jawaharlal Nehru

About 13.8 billion years ago, an event occurred that not only established the relationship between elementary particle physics and cosmology, but also determined the current desire for the unity of physics. This event is called the Big Bang. To understand how our Universe developed from that moment we need to take a brief excursion into the so-called unified field theories.

In 1967 Steven Weinberg, Sheldon Glashow and Abdus Salam showed that weak and electromagnetic interactions become a single electroweak interaction at energies above 100 GeV. (1 GeV = 10^9 eV, and 1 electron volt is the energy that an electron acquires by passing a potential difference of 1 volt.) Below this energy the interaction symmetry is spontaneously broken, and in everyday life we observe them as different interactions. The theory of electroweak interaction was experimentally confirmed in a particle accelerator that produces energy over 100 GeV, whose ring diameter is several kilometers.

In 1974 Glashow and Howard Georgi published their idea that at energies above 10^{14} GeV, electroweak interaction unites with strong interaction and some symmetry is also restored. But, more about symmetry a bit later. Theories that consider the union of these three interactions are called Grand Unified Theories (GUT). It is hardly possible to verify the conclusions of a GUT in the usual way in an accelerator since the diameter of such an accelerator (in the traditional experiment) should be much larger than the size of the Earth. Rather complicated mathematics is needed for characterizing the abstract symmetry of the GUT.

Perhaps the only thing that needs to be said is that the symmetry of the GUT is geometric symmetries associated with the additional 7th dimension of space rolled (compactified) into a 7-dimensional sphere. If the results of the GUT are actually correct then we live in an 11-dimensional space, in which 3 + 1 is a 4-dimensional space-time continuum, and 7 spatial dimensions are rolled up into a compact. Thus, one more fundamental quantity appears – the dimension of the Universe: $N_U = 1 + 3 + 7 = 11$.

Continuing to move further up the energy scale we come to the theory of supergravity or supersymmetry. The results of this theory can manifest themselves at energies above 10^{19} GeV. In this case, the internal symmetry which is associated with the quantum numbers of elementary particles is GUT symmetry. This symmetry and the spatial symmetry of the general theory of relativity are combined. Thus, the complete scheme that we can present to aliens in order to prove our awareness in science is presented in Fig. 8-1. You should not be confused that the characteristic energies of the Union are represented in dimensional units (GeV). The ratio between the orders of these quantities is so characteristic that thinking aliens will understand what we are talking about if they are at our level of development.

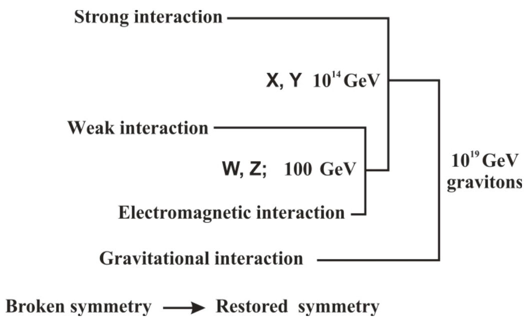


Fig. 8-1.
Energies of the unifying of fundamental interactions

The combination of electroweak and strong interactions is described by the Standard Model. This is one of the most successful physical theories, which has passed all tests since its creation in the 1960s. This theory describes the behavior of elementary particles and explains why they have mass. The discovery of the Higgs boson – a particle that is a key element of the mechanism for generating the mass of elementary particles in the Standard Model – is the most important experiment confirming the theory. The Higgs boson was predicted by Peter Higgs in 1964. The 2013 Nobel Prize in Physics was awarded to François Englert and Peter Higgs for the discovery of this boson.

One of the most promising options for combining all four fundamental interactions at the present time is so-called string theory. The main postulate of string theory is the statement that fundamental interactions arise as a result of vibrations of ultramicroscopic strings with dimensions of the order of 10^{-35} m. This is twenty orders of magnitude smaller than the diameter of the atomic nucleus. Modern science does not have the tools to conduct experiments on such a scale so we cannot check the string theory. Thus, the question of combining all four fundamental interactions remains open.

Before turning to the scenario of the creation of the Universe it seems to us important to narrate one telling story about how discoveries are sometimes made. The essence of the matter is in brief. In 1938 George Gamow predicted the existence of background electromagnetic radiation of the Universe based on theoretical considerations, i.e., radiation emanating not from stars, nebulae, etc., but from "empty" space. In 1946 he put forward a model of the evolution of the universe which explained the origin of this relic radiation. Such a model of the "hot" Universe assumed that once, a long time ago, the temperature of the Universe was much greater than it is now. In 1964, on building their sensitive antenna receiver system, Robert Wilson and Arno Penzias encountered radio noise which they could not explain. It was the background electromagnetic noise of the Universe – CMB radiation also known as "relic radiation". Knowing nothing about the works of Gamow, they just stumbled upon this relic radiation. Meanwhile, astrophysicist Robert Dicke who knew the theory of Gamow specially built an antenna to detect CMB radiation. But Dicke was just six months too late. As a result, the Nobel Prize in 1978 was received by Wilson and Penzias. From this moment on the idea of a "hot" Universe, born as a result of the Big Bang, becomes the main hypothesis of the "creation of the Universe" scenario. The scenario is as follows.

Big Bang. Time 10^{-44} seconds, temperature 10^{32} K. This is the so-called Planck time with the size of the Universe being 10^{-35} cm. Up to this point space, time, radiation and matter were inseparable. Starting from the Planck time their role begins to be different. At 10^{-43} seconds the Universe was in a state with a relative minimum of potential energy (the so-called false vacuum).

This state was absolutely unstable, and the Universe began to inflate at a speed greater than the speed of light. In this case only those sections were informatively interconnected where the distance between them did not exceed *ct*. Such inflation lasted up to a time of 10^{-35} seconds.

Inflation. The inflation model of the Universe is a hypothesis about the physical state and the law of expansion of the Universe at the early stage of the Big Bang (at temperatures above 10^{28} K). It assumes the existence of a period of accelerated expansion compared with the model of a hot Universe. The first version of the theory was proposed in 1981 by Alan Guth. However, the key contribution to its creation was made by Alexey Starobinsky, Andrey Linde, and Vyacheslav Mukhanov.

Separation of gravity. Starting from 10^{-35} seconds at a temperature of 10^{28} K one universal supersymmetry interaction was divided into gravity and the Great Unification.

The separation of the strong interaction. Starting from the time of 10^{-34} seconds at a temperature of 10^{27} K, the symmetry of the Great Unification is broken and a strong interaction is released from it.

The beginning of baryon asymmetry. At a temperature of 10^{16} K and a time of 10^{-12} seconds, quarks and antiquarks are born and annihilate, and the number of particles by one billionth exceeds the number of antiparticles. Later this will lead to the "extinction" of antimatter.

The separation of weak interaction. At a temperature of 10^{15} K, the symmetry between the weak and electromagnetic interactions begins to break and beginning at a time of 10^{-4} seconds and a temperature of 10^{12} K, all four interactions exist independently. Previously free quarks are combined into nucleons – protons and neutrons. The reactions that absorbed neutrinos stop and these particles spread throughout the Universe.

Fixing the number of nucleons. At a temperature of 10^{10} K and a time of 1 second, the conversion of protons into neutrons and vice versa ceases. Their number is fixed in the ratio of 6 to 1.

Paired mutual annihilation of leptons. At a temperature of 10^{8} K and a time of 100 seconds, electrons and positrons as in the case of protons and neutrons are mutually annihilated and a small excess of electrons remains.

Synthesis of the first elements. At a temperature of 10^{7} K and a time of 10^{4} seconds, protons and neutrons merge into nuclei of heavy hydrogen – deuterium and into nuclei of helium.

The most dramatic events in the Universe occurred in the first seconds after the Big Bang. The temperature of the substance and its density fell by more than 20 orders of magnitude, and now the countdown of time has been going on for thousands of years.

The end of the synthesis of elements. After 10^{4} years old neutrons along with protons are spent on the formation of helium nuclei. The remaining protons are hydrogen nuclei.

The end of the era of radiation. The Universe has already cooled down to $3 \cdot 10^4$ K, the intensity of radiation falls, and the main share of energy is already the share of matter.

The Plasma Epoch. Electromagnetism prevails, photons still have such high energy that they do not allow electrons to adjoin atomic nuclei and form atoms. The Universe is still a cosmic gas which is an opaque plasma.

Enlightenment of the Universe. Since the time of 10^5 years photon energy has decreased so much that electrons are now localized around atomic nuclei. Atoms appear. Photons also spread throughout the Universe almost freely creating relic radiation. The Universe becomes transparent and then it continues to gradually cool down.

The time of 10^{10} years and beyond. Cosmic gas forms clusters, celestial bodies arise – quasars and galaxies. Gas clouds form in galaxies, they condense and eventually the first stars appear. Heavier elements are synthesized inside the stars. A star remains after its death under certain conditions and can be condensed. The first planets appear, like ours. Life on Earth appeared over three billion years ago, and about sixty thousand years ago *homo sapiens* appeared – a reasonable man.

The known values of the fundamental world constants thus provide not only the modern structure of the Universe (as noted earlier in Chapter 7) but also its evolution. What determined such an exact correspondence of world constants so that the existence of a complex structure of the Universe including our life became possible?

One of the possible answers to this question is the anthropic principle according to which our Universe possesses the observed properties precisely because these properties allow the possibility of the existence of an observer. The evolutionary biologist Alfred Wallace anticipated the anthropic principle as long ago as 1904:

> Such a vast and complex universe as that which we know exists around us, may have been absolutely required ... in order to produce a world that should be precisely adapted in every detail for the orderly development of life culminating in man.

The discussion of the anthropic principle is beyond the scope of this book. Note only that its most thorough analysis is given in the book "Anthropic cosmological principle" by John Barrow and Frank Tipler.

In our opinion, the anthropic principle has more philosophical, than natural scientific significance. The logic of the development of cosmology as the science of the origin and development of the Universe should lead to its exclusion as an extra entity (see axiom 5). Paraphrasing René Descartes one can say "*cogito ergo mundus talis est*", i.e., "I think, therefore, the world is such as it is".

Dark matter and dark energy. Discoveries in cosmology usually take place in two stages. The first one examines the essence of the question: does this something exist? The second question is: what is it? In the case of dark matter most scientists believe that the first stage has been passed. If we talk about dark energy, scientists have not yet advanced in this problem.

Hidden mass. The first to point out the possibility of the existence of dark matter was Fritz Zwicky. In 1932 he studied the movement of a galaxy cluster revolving around its center. He calculated how much matter is required to maintain their gravitational connection. After analyzing the radiation and calculating the total number of available stars he found that the observed mass is not enough. Zwicky's discovery should have caused great interest. However due to his irritable nature his colleagues did not like him so this discovery, like many others made by him, was ignored.

Spiral mystery. The study of the hidden mass will only resume after 40 years. This will deal with Vera Rubin. Using a highly sensitive spectrograph developed by Kent Ford, Rubin studied the motion of stars in spiral galaxies. Calculating the speed of these stars she noted the anomaly. According to the law of gravity, stars in a galaxy must rotate around the center, the slower they are far from the center. However, Rubin discovered that this is not the case.

Dark matter. Rubin realized that the same mass should be concentrated in the outer halo of spiral galaxies as in the brightly glowing central disk. However, it is impossible to see it because it radiates nothing and participates only in the gravitational interaction. At first these observations were met with skepticism but further research confirmed the findings of Rubin and Ford. Further analysis of gas in clusters of galaxies showed that the mass of dark matter is ten times more than the observed matter. It turned out that Zwicky was right. Gravitational lensing was developed for the more accurate counting of dark matter. This method was also first proposed by Zwicky. Thus it is possible to consider the existence of dark matter. However, the nature of dark matter is not yet known to science.

Strange particles. So far, scientists do not know what dark matter is but they know what exactly does not apply to it. Dark matter has mass but it does not participate in electromagnetic interaction with other matter therefore it does not emit or absorb light. It does not contain the usual elementary particles that we know about; therefore, perhaps it consists of some strange particles that are not yet discovered. Scientists call such particles "weakly interacting massive particles" (WIMP). Attempts are being made to find them but for now the results are still zero.

Dark energy. In addition, there are two new problems. Firstly, the mass of the Universe is much larger than previously thought. This turned up from the analysis of the shape of the Universe which was obtained from the cosmic background radiation maps. Secondly, the rate of expansion of the Universe is increasing. Saul Perlmutter discovered this. So, there must be some kind of interaction, which manifests itself as anti-gravity. The energy associated with this interaction is called dark energy.

Thus, scientists faced the problem of formulating a new theory which explains the nature of dark mass and dark energy.

Fig. 8-2 shows the currently known ratios between the masses of dark energy, dark mass, stars, intergalactic gas, etc.

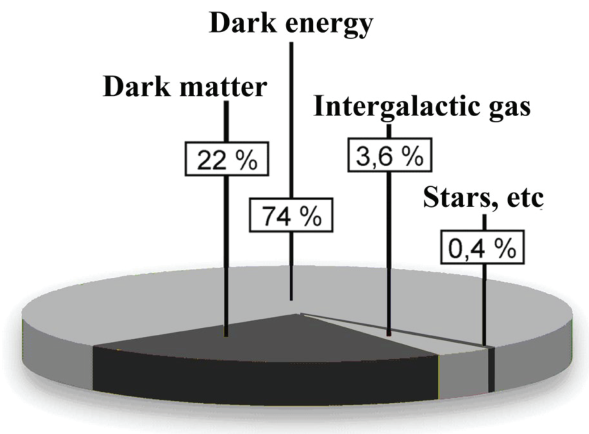


Fig. 8-2.
Relations between the masses of dark energy, dark mass, stars, intergalactic gas, etc.

CHAPTER 9

MACROCOSM. LAWS OF EQUILIBRIUM. THERMODYNAMICS. THE BOLTZMANN CONSTANT

Where is the beginning of that end
which ends the beginning?
Kozma Prutkov

The ancients with their high observation and irrational thinking, came to the conclusion that for the stability of everything that exists, three laws are necessary: the law of birth of the new, the law of conservation and the law of destruction. For example, in Hinduism there is the concept of the Trimurti supreme deity in which the cosmic functions of creation, maintenance and destruction are personified as a triad of deities: Brahma the Creator, Vishnu the Guardian and Shiva the Destroyer. Let's see how modern science fits this view.

According to modern ideas, related to a greater degree with the concept of the Nobel laureate Ilya Prigogine there are two complementary approaches to the description of nature: dynamic and thermodynamic. The first is most appropriate for describing individual objects (bodies, atoms, molecules, and elementary particles) and their interaction with a number of the same or other objects.

The second approach considers significantly larger sets of objects. It is applicable to problems of the microworld with the number of particles of the order of the Amedeo Avogadro number $N_A \approx 6.02 \cdot 10^{23}\,\mathrm{mol}^{-1}$. This approach operates with generalized parameters such as energy, heat and entropy. Such systems are called thermodynamic. Thermodynamics deals with quantities which convey a quantitative description using measurable macroscopic physical quantities but may be explained in terms of microscopic constituents by statistical mechanics.

For thermodynamic systems the expressions used in the theory must satisfy the formal limit procedure, i.e., if the number of particles in the system N and its volume V tend to infinity (which in fact cannot be) then their ratio V/N = const.

There are four axioms for thermodynamic systems. In physics these axioms are called the laws of thermodynamics.

Zero law of thermodynamics. For each thermodynamic (TD) system there is a state of thermodynamic equilibrium which, under fixed external conditions, spontaneously reaches in time. This property is specific to TD systems and is required for them without exception. TD equilibrium as a concept is a state where the macroscopic parameters of the system do not change with time and when there are no flows of matter and energy. In macroscopic theory the zero law is a generalization of everyday experience and observations of TD systems. However, from a microscopic point of view, this statement is far from obvious. Poincaré in 1890 proved the so-called recurrence theorem. This theorem states that the mechanical state of an isolated system does not in any way turn into a certain stable equilibrium state with time, but it is reproduced with a given accuracy after a finite period of time. However, this period of time for a system consisting of one mole of a substance is by the roughest estimates 10^{N_A}. It is clear that the age of the Universe in comparison with this quantity is only a moment. In addition, the states recorded using macroscopic devices no longer represent purely mechanical states. However, this problem of recurrence certainly has a fundamental interest.

The first law of thermodynamics. This is a special case of a global law of energy conservation, adapted for thermodynamic systems. The law of conservation of energy states that the total energy in an isolated system is constant. An isolated system does not exchange particles and energy with the surrounding space. In an isolated system energy can be converted from one form to another but it can neither be created nor destroyed. This law was established in science, becoming the basis of its formation in the middle of the XIX century. The clarification of the meaning and the formulation of this law are generally related primarily to the works of Julius Mayer (1842), James Joule (1843), and Hermann Helmholtz (1847). The term “energy” was introduced into physics by William Rankin in 1853 although the word itself was encountered much earlier in the III century BC by Aristotle.

We write the first law of thermodynamics in the form of energy balance:

$$dE = \delta Q - \delta W \quad . \tag{9.1}$$

Here E is the internal energy, Q is the amount of heat, and W is the amount of work. The symbols d and δ mean infinitely small changes in the corresponding quantities. The difference between these symbols is that d means the exact differential, and the symbol δ means an inexact differential.

Therefore, in an isolated system there is a balance between the values of dE, δQ and δW. The first law of TD is equivalent to the following statement. Perpetual motion machines of the first kind are impossible. A perpetual motion machine of the first kind is understood as an imaginary mechanism that is capable of doing work without consuming energy from outside.

The second law of thermodynamics. This is the law of increasing entropy in isolated systems. The second law of TD establishes two essential aspects.

1. The existence for any equilibrium (more precisely, quasi-equilibrium that is participating in a quasi-static process) TD system of a unique function of the TD state called entropy such that its exact differential:

$$dS = \frac{\delta Q}{T} \tag{9.2}$$

i.e., unlike the amount of heat Q entropy is a potential function and describes the evolution of the thermodynamic system.

2. In mathematics a differential form δQ in the general case is a form of Johann Pfaff:

$$\delta Q = P(x, y, z,...)dx + K(x, y, z,...)dy + R(x, y, z,...)dz + ... \tag{9.3}$$

It is known that if a Pfaff form is defined by only two variables then it always has an integrating factor. Then in the definition of the second law of thermodynamics given above there is no axiomatic statement. However, for three or more variables the mathematical situation changes radically. It turns out that the existence of an integrating factor is not possible with arbitrary functions P, K and R but only if certain sufficiently stringent conditions are imposed on them. Thus the second formulation of the second beginning postulates the existence of an integrating factor for any structure δQ of the Pfaff form; that is, this is a fundamentally axiomatic statement. Trivial cases of one or two variables are not proof of this general statement.

The third law of thermodynamics. In the radical formulation of Planck (1910) it has the form of an initial boundary condition. This condition is formulated as a limit:

$$\lim_{T \to 0} S = 0 \tag{9.4}$$

Let us now dwell on the second law of thermodynamics in isolated thermodynamic systems focusing our attention on non-equilibrium processes.

Thermodynamics is divided into equilibrium and non-equilibrium (for example diffusion occurs in a non-equilibrium system). Non-equilibrium thermodynamics in turn is divided into linear and nonlinear. But we will talk about this later. We first consider the equilibrium thermodynamics which considers quasi-static processes. Such processes proceed infinitely slowly and consist of an infinite sequence of equilibrium states that have very little difference between them. Clearly this is an idealization not a real process. The main advantage of such an idealization is the reversibility of processes; according to the definition, each intermediate state being an equilibrium is completely indifferent to the direction of the flow of the process. Thus, time t as a dynamic parameter drops out of the theory and the process becomes as it were inertia-free.

The second law of thermodynamics also includes the statement that for any non-quasi-static process occurring in the TD system:

$$dS > \frac{\delta Q'}{T}, \tag{9.5}$$

where $\delta Q'$ is the amount of heat absorbed by the system (in a non-quasi-static process) when moving from one state to another nearby. This statement formulated by Rudolf Clausius in 1865 is essentially a consequence of empirical considerations – the principle of maximum work and maximum heat absorption, the discussion of which we omit here.

The second law of TD determines the direction of the flow of the real process and the impossibility of a spontaneous decrease in entropy in an isolated system. Indeed, in an isolated system $\delta Q = \delta W = 0$, and non-equilibrium processes proceed in such a way that:

$$dS > 0. \tag{9.6}$$

This actually determines the direction of the non-equilibrium process and the equilibrium state will correspond to the maximum value of the entropy.

According to this law the system itself tends to a state with maximum entropy which corresponds to a global TD equilibrium. This maximally chaotic state in mathematics is described using the concept of an attractor. Chaos is a state of matter that is characterized by maximum disorder and maximum symmetry, and is the most likely state of the system (see Chapter 3, second example).

This is illustrated in Fig. 9-1 where the gas after the partition is removed spreads throughout the volume, and the symmetry of the system increases.

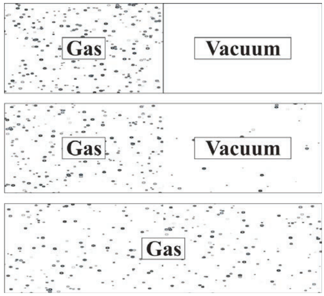


Fig. 9-1.
After removing the septum gas spreads throughout the volume

Imagine a vessel in the form of a cone. We will bring it into rotation and completely and arbitrarily we will lower the balls on its inner surface. Rolling along the surface of the cone each time along different trajectories, they end up at its bottom in a global stable state.

If we now look at the possible trajectories from above, then they represent a system of spirals converging at one point (the lower part of the cone). This is an attractor model. Fig. 9-2 shows schematically the evolution of some variables with asymptotic stability and phase trajectories converging into an attractor. Thus, whatever the initial conditions, the evolution of the system is such that all paths lead to the attractor. Surely you remember the ancient saying: *Omnes viae Romam ducunt!* (All roads lead to Rome). In fact, Rome was a kind of attractor of the ancient world.

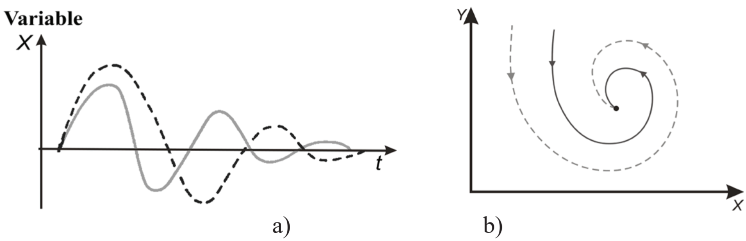


Fig. 9-2.
Asymptotic stability: a) temporal evolution of the state; and b) phase trajectories converging into an attractor

The second law of TD is not a universal law of nature. Nevertheless, in its position it occupies a unique place among other fundamental laws. The fact is that the second law of TD relates to the irreversibility of the unidirectionality of processes in an isolated system. The measure of irreversibility is entropy. The second law of TD is a criterion for distinguishing the future from the past. Thus, the direction of time is determined (time arrow). The probabilistic interpretation of entropy was first given in 1877 by Ludwig Boltzmann. For this he used the idea of determining the most likely state of a system of material points from a thermodynamic point of view. In statistical physics of equilibrium systems there is the Boltzmann formula which determines the entropy S (the characteristic of the macrostate) through the statistical weight Γ (the characteristic of the microstate):

$$S = k_B \ln \Gamma , \tag{9.7}$$

where $k_B = 1.38\ 10^{-23}$ J/K is the Boltzmann constant.

The statistical weight Γ in statistical physics determines the total number of microscopic realizations of a given macroscopic state. In other words, statistical weight is the number of ways in which a given macroscopic state of a statistical system can be realized.

In the macroscopic approach the equilibrium state is unique (this is a consequence of the zero law of TD). This state does not depend on which of the microscopic methods (from among Γ) is realized. Therefore, all possible paths are equivalent in this respect.

Thus, the transition of a system from a non-equilibrium state to an equilibrium state is a transition to the most probable state corresponding to the maximum entropy. Despite the fact that the Boltzmann constant k_B does not define elementary physical processes and is not included in the basic principles of dynamics the following circumstance is important. The Boltzmann constant relates microscopic dynamic phenomena and macroscopic characteristics of the state of a large collective of particles.

Boltzmann established that in the final state with the entropy maximum the probability density $W(v)$ distribution of particles modulo the velocity for an ideal classical gas is proportional to the normal Gaussian distribution:

$$W(v) \sim \exp\left(-\frac{mv^2}{2k_B T}\right), \tag{9.8}$$

where m is the mass of particles, and T is the temperature. A graph of the normal distribution function (9.8) is shown in Fig. 9-3. We will use the properties of the normal distribution later when we discuss the issue related to the dispersion of additive and non-additive quantities.

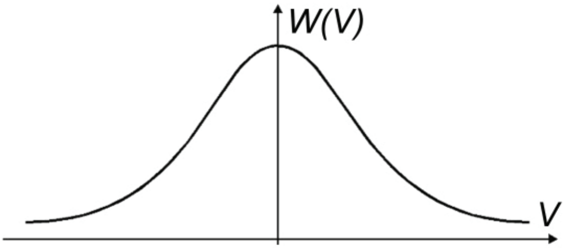


Fig. 9-3.
Graph of the normal distribution function (9.8)

It can be seen from the graph in Fig. 9-3 that in the state of global equilibrium in which the system finds itself the density of probability to detect a particle with zero velocity is maximum. However, importantly, there is a probability of detecting particles with non-zero speed. In the future we will use this fact to justify one of the axioms of biology. Thus, the state of maximum entropy is not a frozen state of chaos, but somehow a mobile one.

Summary. A state with a maximum of entropy has the following extreme features. It is probably the most symmetrical and stable, and has minimal orderliness.

The wording of the different versions of the second law of TD belongs to a number of authors. These are William Thomson (Lord Kelvin) – 1851, Wilhelm Ostwald – 1851, Rudolf Clausius – 1850 and 1865, and Constantin Carathéodory – 1909. However, Nicolas Sadi Carnot came to many fundamental conclusions as early as 1824. Carnot is rightly called the father of thermodynamics despite the fact that he published only one work. In this work however, Carnot laid the foundations of thermodynamics. During Carnot's short life his results were either unknown to, or hardly attracted the attention of, scientists. However, these results were later used by Clausius and Lord Kelvin to define the concept of entropy and formulate the second law of TD.

Chapter 10

Self-organization. Dissipative structures. The Glansdorff-Prigogine evolution criterion

From small reasons, there are
great consequences ...
Kozma Prutkov

Until the middle of the XX century a dispute between a representative of science, an atheist and a representative of religion about the spontaneous emergence of new complex structures would be formally in favor of the latter. Indeed, a representative of a religion could say that the laws of conservation, the law of increasing entropy, are known in science. However, there is no law explaining how a fundamentally new and more complex structure can appear. An honest scientist would have to agree with this. Up to the 1970s one could hear quite logical criticism of the teachings of Charles Darwin. It was calculated that if the evolution of species proceeds by chance then there simply will not be enough time for man to emerge from the simplest organisms since this time is longer than the age of our Earth. The conclusion was unequivocal – the presence of the Creator is necessary. However, theologians in this statement were wrong. In the second half of the XX century a doctrine emerged about the possibility of the emergence of fundamentally new ordered structures from chaos. This spontaneous emergence of new complex structures from the initial chaotic state is called self-organization. In thermodynamics, these new structures were called dissipative. Self-organization can occur if the following four conditions are met. The system must be open, non-linear, and far from the equilibrium state, and there must be feedbacks in the system. To understand the situation that has arisen let's give a summary of the historical information about the main stages of development of this global concept.

An excursus of the history of open systems: the XIX century. The theory of the time evolution of gas in a closed system (Boltzmann); the theory of stability of dynamic systems (Poincaré and Alexander Lyapunov); and the first step in the theory of the evolution of open biological systems (Darwin). It is very characteristic that despite the achievements in the field of thermodynamics and electromagnetism Boltzmann considered the XIX century as the century of Darwin. So, he highly appreciated the principle of biological evolution. What was the basis of Boltzmann's conclusion? The fact is that Boltzmann was one of the few physicists at the time who first understood the importance of Darwin's discovery from the standpoint of the theory of the evolution of open non-equilibrium systems. Thus, already at the turn of the XX century it became clear that the development of the theory of non-equilibrium processes in physical, chemical and biological systems is one of the most important tasks of natural science.

The XX century. The first step in the theory of non-equilibrium processes was taken by Einstein, Marian Smoluchowski and Paul Langevin. They created the theory of Brownian motion. The English botanist Robert Brown observed this phenomenon for the first time in 1827. The reason for the Brownian motion is the jolts from the side of the liquid molecules. According to the Boltzmann kinetic equation the average energy of gas particles in the process of evolution is conserved. This condition is necessary so that in the process of evolution to an equilibrium state, the entropy and with it the degree of chaoticity, is increasing. This is the well-known Boltzmann *H*-theorem, according to which the entropy in irreversible processes cannot decrease. This theorem follows from the analysis of the so-called Boltzmann kinetic equation which describes the behavior of weakly non-equilibrium rarefied open systems. Note that according to the Boltzmann equation it is not the exact value of energy that is saved but only its average value. Thus, energy fluctuations are possible.

In the XX century, mathematicians and physicists made an enormous contribution to the science of open systems. Lyapunov, Poincaré, and later Alexander Andronov, Kolmogorov and Nikolay Krylov have already been mentioned. Kolmogorov's work in 1957 on the entropy of dynamical systems can be considered the forerunner of the science of self-organization. In recent years the work of a number of authors of the Brussels School and above all Prigogine has developed thermodynamics of strongly non-equilibrium systems.

Note that in contrast to isolated or closed systems, in general open systems exchange energy, particles and information with the environment. In open systems dissipative structures can form. The complexity of open systems predetermines the existence of cooperative (coherent) motions of a large number of particles in them. Hence the name of the new science – synergetics, proposed by Hermann Haken. In order to understand some general patterns of the emergence of dissipative structures in the process of self-organization, we will consider the most obvious example.

Henri Bénard cells. Imagine a layer of liquid between two horizontal parallel planes, the linear dimensions of which far exceed the thickness of the liquid layer. If the fluid is isolated, then no external forces (except gravitational) act on it and there is no exchange of particles. In this case, the liquid remains in an equilibrium state for an arbitrarily long time. This state is characterized by a complete macroscopic identity of different parts of the liquid, regardless of their location and the distance between them. Therefore, if you do not take into account the boundaries, the fluid inside our "aquarium" is homogeneous and isotropic which means that the state has maximum symmetry. If we create in such a system, a temperature difference between the lower (T_1) and upper (T_2) surfaces by means of a continuous heat supply then we will bring the system out of equilibrium. As long as the temperature difference $\Delta T = T_1 - T_2$ is small a stationary state will be established in the system due to thermal conductivity. It is characterized by almost linear changes in temperature, density and pressure from the lower to the upper plane. However, as soon as the temperature difference exceeds a certain critical value T_{cr}, we will see how the fundamentally new state is suddenly established. In the liquid, so-called Bénard cells are formed, see Fig. 10-1. Convection rotation of the liquid occurs in each cell. If you look along the horizontal axis, then the direction of rotation of the liquid in two neighboring cells is alternating: then along, then counterclockwise, see Fig. 10-1.

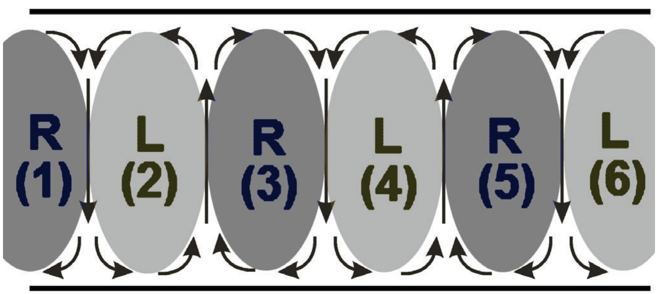


Fig. 10-1.
Bénard cells in the gravitational field

Thus, there is a special transition from an unstructured homogeneous and isotropic system to a structured, i.e., ordered one. This transition is accompanied by a decrease in the symmetry of space (see the discussion of symmetry in Chapter 3, Example 1). Note the principal feature of this transition; it does not occur smoothly, but abruptly. It is significant that when such an experiment is repeated it is fundamentally impossible to predict the direction of rotation of the fluid in the cell. Such unpredictability is a fundamental feature of the formation of dissipative structures. In other words, in the process of self-organization the system may react differently to the same external influences. From the point of view of the mathematical theory of dynamical systems this means that with the same values of the parameters controlling the system several different solutions are possible. The special state of the system in which a multiplicity of solutions occurs is called bifurcation. To illustrate this state, consider the simplest mechanical analogy of a bifurcation.

In Fig. 10-2 the ball rolls down an inclined chute with a branched profile. The main gutter is divided into two gutters by a protrusion. The protrusion plays the role of a critical point. The direction of movement of the ball after the critical point (the point of bifurcation of the gutter) cannot be predicted in advance. After a sharp transition through the critical state (the ball moves left or right) the symmetry of the system decreases. In this example common properties of all dissipative structures are: the reduction of symmetry, greater orderliness and their sharp occurrence. These properties are manifested in the phenomena of self-organization in various fields of science: in physics, chemistry, and biology, and also at the social level.

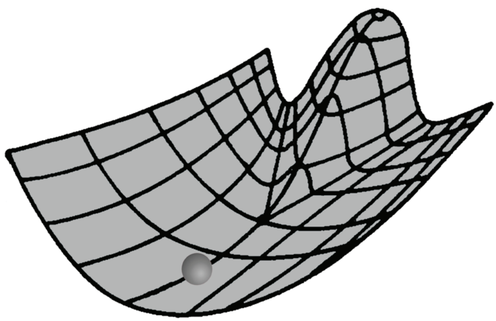


Fig. 10-2.
Mechanical illustration of the bifurcation

However, recent studies have shown that in biological systems everything is much more complicated. Therefore, Prigogine's ideas about dissipative structures for biological systems should be corrected (England 2015).

There are three types of dissipative structures; firstly, spatial structures, for example Bénard cells, and secondly, temporal structures. Such structures arise, for example, in the process of the Belousov–Zhabotinsky chemical autocatalytic reaction named after the discoverer Boris Belousov and the first researcher Anatol Zhabotinsky. Thirdly, spatial-temporal structures arise in nonlinear chemical reactions occurring in a thin layer in the presence of local concentration fluctuations and the diffusion of reagents.

The nature of self-organization is determined by the fact that far from the equilibrium state due to nonlinearity the system is unstable (in the sense of Lyapunov). Therefore, even small fluctuations can lead to a new state, for which the cumulative (coherent) motion of a large number of particles is characteristic. The general theory of self-organization processes is based on the universal evolutionary principle formulated by Paul Glansdorff and Ilya Prigogine (Prigogine and Stengers 1984). Here we only note the most significant factors as a detailed discussion of this theory is beyond the scope of our book.

The principle of Glansdorff–Prigogine establishes the following property of the production of entropy P related to thermodynamic forces X:

$$\frac{\partial P_X}{\partial t} \leq 0 \quad . \qquad (10.1)$$

Entropy production is the increment of the system's entropy in a unit of time due to irreversible processes occurring in the system itself. In TD by analogy with mechanics the concept of TD force X is introduced as a cause of change in the TD state as well as the flow I as a consequence of the action of TD force. For example, the concentration gradient of matter is the TD force, and the diffusion caused by this gradient is the flow. The total entropy production P is the sum of the entropy productions related to forces P_X and flows P_I. However, in nonlinear systems only part of its P_X satisfies general inequality (10.1).

Inequality (10.1) holds true for any relationship between forces and flows in conditions of local equilibrium. Therefore, it is called the universal principle of evolution. According to this principle, in any non-equilibrium nonlinear system with fixed boundary conditions the processes proceed in such a way that the rate of change decreases in the production of entropy due to the change in thermodynamic forces. The equal sign in (10.1) refers to the stationary process.

Determinism, uncertainty and self-organization of dynamic systems. From the time of Newton until the middle of the XX century determinism's justice was not questioned. Indeed, all fundamental physical theories are based on time-reversible equations. This situation continued until the appearance of the mathematical theory of dynamical systems. The creation of this theory was attended by many scientists: Kolmogorov, Arnold, Jürgen Moser, Edward Lorenz, Yakov Sinai and others.

Evidence of a revolutionary change in the understanding of the determinism of mechanical systems is a statement made by the president of The International Union for Theoretical and Applied Mechanics, Sir Michael James Lighthill:

> Here I have to pause, and to speak once again on behalf of the broad global fraternity of practitioners of mechanics. We are all deeply conscious today that the enthusiasm of our forebears for the marvellous achievements of Newtonian mechanics led them to make generalizations in this area of predictability which, indeed, we may have generally tended to believe before 1960, but which we now recognize were false. We collectively wish to apologize for having misled the general educated public by spreading ideas about the determinism of systems satisfying Newton's laws of motion that, after 1960, were to be proved incorrect. In this lecture, I am trying to make belated amends by explaining both the very different picture that we now discern, and the reasons for it having been uncovered so late. (Lighthill 1986, 38)

This statement is caused by the following circumstance. The modern theory of dynamical systems has demonstrated an unexpected fact. Systems described by Newton's equations may not have the property of predictability. The reasons cited by Sir James Lighthill are connected with the exponential divergence of the trajectories of highly unstable chaotic systems described by positive Lyapunov indicators. In mathematical terms, the Lyapunov indicator measures sensitivity to initial conditions which means the following. Each point of a chaotic system is arbitrarily and closely approximated by other points that have significantly different future trajectories. Thus, any small change or disturbance of the current trajectory can lead to significantly different behavior in the future. However, the unusual recognition of Sir James Lighthill is not caused by all this. We will not stop at one of the aspects of the problem associated with self-organization.

The main problem in dynamics is integrating its equations. Since we have the equations of motion of Newton or William Hamilton, naturally we would like to have explicit analytical expressions for variables, i.e., coordinates or speeds as functions of time. At the end of the XIX century

Poincaré showed that not all dynamical systems resemble each other as previously thought. It turns out that there are two types of systems: integrable and non-integrable. For the integrable systems we can eliminate interaction and reduce the task to the problem of free movement. For the non-integrable systems it is necessary to abandon the search for an exact analytical solution. Therefore, the description of dynamics in terms of trajectories is possibly only approximate. In special cases when dynamic chaos is realized in the system, we must abandon the concept of determinism and proceed to a probabilistic description. By dynamic chaos we mean a state with extremely unstable trajectories.

Let us see how this is obtained in the framework of Hamiltonian dynamics where the fundamental quantity is the Hamilton function $H = E + U$ equal to the sum of the kinetic E and potential U energies (Landau and Lifshitz 1976). For conservative systems where the Hamilton function H does not explicitly depend on time it is expressed in terms of generalized momentum $\mathbf{p}_i$ and generalized coordinates $\mathbf{q}_i$ as follows:

$$H = E(\mathbf{p}_1, \mathbf{p}_2, ...\mathbf{p}_N) + U(\mathbf{q}_1, \mathbf{q}_2, ...\mathbf{q}_N), \qquad (10.2)$$

where N – quantity of particles in a system.

The Hamilton function (10.2) is represented in the so-called canonical variables when the kinetic energy depends only on the generalized momentum and the potential energy depends only on the generalized coordinates of the particles. Through (10.2) the canonical equations of motion of the system are expressed. The canonical representation of the equations of motion is rightfully considered the apotheosis of classical dynamics.

For an integrable system there is a so-called canonical transformation in which the Hamilton function in the new variables no longer depends on the potential energy. Such variables are called cyclic or ignorable.

The ability to eliminate the influence of potential energy through canonical transformation into new cyclic variables is the main characteristic of integrable dynamic systems in the Poincaré sense. Consequently, for integrable systems after the Hamilton function is transformed into the appropriate form the interaction between particles is excluded.

Until 1889 it was believed that all dynamic systems were integrable. Nevertheless, there were tasks with unknown solutions, for example a problem of three or more bodies. Difficulties in finding accurate analytical solutions for such problems were associated with technical or computational problems. However, in 1889 Poincaré showed the following. In the general case it is impossible to obtain a canonical transformation that would lead to cyclical variables. In addition, Poincaré showed that most systems are not integrable in principle.

What is the meaning of such a strong mathematical statement? What would happen if Poincaré on the contrary, proved the integrability of all dynamical systems?

This would mean that without exception, the evolution of all dynamical systems with any number of particles is isomorphic to the motion of free particles that do not interact in any way with each other. To wit, these particles can never act as a collective, i.e., coherently. Consequently, there can be no self-organization in principle. This means that life cannot arise in an integrable world.

However, Poincaré not only proved non-integrability but also indicated the reason for the non-integrability of systems. This is the existence of resonances between the degrees of freedom and as a consequence the emergence of the problem of so-called small denominators.

This problem was known in astronomy before Poincaré. But it was his theorem that showed that the main difficulty associated with divergence in solving dynamic problems cannot be eliminated. Small denominators tend to zero and their inverse value tends to infinity. It makes it impossible to introduce cyclic variables for most dynamic systems starting with three bodies (Tabor 1989).

With the advent of the Kolmogorov–Arnold–Moser (KAM) theory the problem of non-integrability and small denominators was the beginning of a new development of dynamics including the dynamics of both coherent movements and chaotic ones. The KAM theory examines the effects of resonances on trajectories. At some points in the phase space of a dynamic system there are resonances, and at others there are not. In the theory of dynamic systems phase space is the space of dynamic variables on which the state of the system depends. Each possible state corresponds to a single point in the phase space. Resonances correspond to rational ratios between frequencies. Frequency ratios outside resonances have irrational values. It is clear that most of the points in the phase space are non-resonant. Resonances lead to periodic motions and the absence of resonances leads to quasiperiodic motion. Consequently, periodic motions are an exception to the general case of complex motion.

The main result of the KAM theory is that there are two fundamentally different types of trajectories. The first are quasi-periodic trajectories. The second are stochastic trajectories arising from the destruction of resonances. The KAM theory does not lead to a dynamic theory of chaos. However, it shows that for small values of a certain parameter, an intermediate mode is obtained. In this mode trajectories of two types coexist: regular and stochastic (extremely unstable).

Fractals, or strange attractors. This system may deviate from its standard state in four different ways.

1. Deviation from the standard state remains limited for any period of time. In mathematics, this case corresponds to Lyapunov stability. The Lyapunov stable state corresponds to orbital stability in the phase space.

2. The state of the system tends to the standard state if time tends to infinity. This is asymptotic stability. In terms of phase space this corresponds to asymptotic orbital stability. Such a state necessarily implies irreversibility, i.e., it does not belong to conservative systems. This type of stability is characteristic of dissipative systems in which dissipative structures can arise under the above-mentioned necessary conditions. Such stability in dissipative systems provides the reproducibility of the regime called the attractor which we discussed above.

3. The state of the system does not remain in the vicinity of the standard state which is unstable. In phase space, this state corresponds to the case of orbital instability. Unstable states can be in both conservative and dissipative systems.

4. If the perturbation does not exceed a certain value then the state of the system remains in a certain neighborhood of the standard state. This state is called locally stable. However, in this case there is no global stability of the system corresponding to the global attractor in the phase space.

In addition to these reactions of the system to external disturbances there is another opportunity leading to a fundamentally different nature of a system's behavior – the transition to chaos. The usual attractor (fixed points or limit cycles) which is a set of points in the phase space characterizes the system's aspiration for an ordered predictable movement. Note that the dimension of the attractor is always less than the dimension of the phase space. In particular, attractors in three-dimensional phase space have only integer dimensions: zero (fixed point), one (line), and two (surface). However, in mathematics objects with dimensions that are intermediate between a point and a line, between a line and a surface, and between a surface and a volume are known. Benoit Mandelbrot called these sets fractals.

We give the simplest example of a fractal, the so-called George Cantor set. Consider the segment [0, 1] on the number axis. We divide this segment into three equal parts and discard the middle part. After that the remaining segments are again divided into three equal parts and the middle part is discarded. This procedure continues unlimited. The length of all the discarded parts is an infinite geometric progression with a denominator of

2/3 and the initial term is 1/3 long. As a result, the length of all discarded parts is 1. However, in the segment after the procedure there are infinitely many points, for example, 0, 1, 1/4, etc. This set of points has no length since there are gaps between any two points. Nevertheless, the dimension of such a set is not equal to zero. In the case of the Cantor set its dimension is $d \approx 0.63$, i.e., the Cantor set has an intermediate dimension between a point ($d = 0$) and a line ($d = 1$). Such sets – fractals, or strange attractors, realize the chaotic behavior of the system when its behavior lacks any kind of determinism and orderliness. In this case we can talk about the probabilistic description of the system (Nicolis and Prigogine 1989).

CHAPTER 11

CRITERION OF RELATIVE ORDERING OF LIVING SYSTEMS. EVOLUTION AND DEGRADATION. THE MAIN REGULARITIES OF THE MATHEMATICAL THEORY OF REARRANGEMENTS

The body's amazing ability to concentrate on itself the "flow of order", thus avoiding the transition to atomic chaos, the ability to "drink orderliness" from a suitable medium seems to be associated with the presence of "aperiodic solids"– chromosomal molecules, they represent the highest degree of order among the associations of atoms known to us (higher than that of ordinary periodic crystals).

Erwin Schrödinger

"What is life in terms of physics?"

Since the basic laws of self-organization in open systems are outlined in the previous chapter, we will consider these features for living organisms in the context of their evolution. We first give a general definition of this concept. By evolution we mean the process of change and development in nature and society. In physical closed (isolated) systems, as we have already discussed, evolution leads to an equilibrium state with maximum entropy and a maximum degree of randomness. In open systems, two classes of evolutionary processes can be distinguished.

1. Time evolution to a non-equilibrium stationary state; and
2. The process of evolution through a sequence of non-equilibrium stationary states, occurring due to a change in the control parameters.

Self-organization and degradation are two possible variants of evolution and in order to distinguish them it is necessary to introduce a new term – "the norm of chaos". Deviation from it ("norms") in one direction or another using medical language can be interpreted as "illness" i.e. degradation which means restoration to the initial state is "treatment" i.e. self-organization. The functioning of a living organism is only possible with a certain norm of chaotic behavior which corresponds to a highly non-equilibrium state. However, the reference point from the equilibrium state (as for example in a simple physical system) in this case does not exist.

Therefore, in biology, economics, and sociology objective information on changes in the degree of chaos is still insufficient to conclude that there is a process of self-organization or degradation. Thereby, it is appropriate to use the terms “the norm of chaos” and “treatment”. It is interesting to note that Plato had the first ideas about evolutionary processes. Since then in one way or another this problem has been raised by many scientists. According to modern concepts for its understanding it is required to define a number of terms. Dynamic chaos is a state that means there are no sources of fluctuations or sources of disorder in the system. This is its difference from physical chaos.

There are two classes of nonlinear systems – dynamic and stochastic. The classification is based on the property of motion reproducibility under the given initial conditions. Reproducible movements are realized in dynamic systems and non-reproducible movements are realized in stochastic systems. However, even if there are no random sources and the process is formally reproducible, the movement can be so complex that the result turns out to actually be unpredictable.

A feature of dynamic chaos is the dynamic instability of motion expressed in the strong (exponential) divergence of trajectories close to the initial moment. Even in relatively simple dynamic systems there are extremely complex motions that are perceived as chaotic due to the impossibility of predicting the result. The mathematical description of such a complex state leads to the concept of a strange attractor.

Dynamic instability can play an important constructive role in open systems. Let us give examples, the idea being borrowed from Yuri Klimontovich with minor changes. First example. An international conference has ended. This is the initial state for its participants. Consider two possible options for their movement.

1. After its termination participants are moving together, not moving away from each other at a considerable distance. For example, they travel by train to London from Cambridge where the conference was held.

2. Then the participants disperse separately to different cities. This is “exponential scatter”. In other words, the movement becomes dynamically unstable.

The question arises. Which of these two options contributes more to the use of new information received at the conference? The first option is useful to some extent as it allows for further discussion of the conference. But it is clear that the second option, when there is a “mixing of trajectories”, contributes more to the progress of science. In this case participants transmit information more rapidly to different places and different listeners. This example demonstrates that dynamic instability and mixing may not lead to chaos but may play a positive constructive role.

We will give examples from medicine only in the form of a statement of results (Klimontovich 1996). Consider the response of a living organism to stress. Mostly, in female individuals the degree of chaos increased, and in male individuals the degree of randomness decreased (that is some orderliness occurred). Returning to the initial state to the "the norm of chaos" implies "treatment". For women this treatment is accompanied by a decrease in randomness, i.e. a process of self-organization takes place, and in men an increase in randomness occurs, that is degradation. Thus, it turns out that for a living organism the meaning of the concepts "self-organization" and "degradation" does not have an unambiguous connection with an increased or on the contrary a decreased degree of orderliness.

It turns out there are only three types of "patients". The first type (male) is a reduced degree of randomness i.e. excessive orderliness. The second type (female) is a not too increased degree of randomness. The third type (super female) is a significantly increased degree of randomness. It is known that many women are hysterical from stress and for men it is extremely rare.

"Soft" and "hard" mathematical models. This section is a simplified presentation of "Theory of Disasters" (Arnold 1990). An example of a hard model is the multiplication table. The simplest example of a soft model is the principle "the deeper you go into the woods, the more trees there are". The usefulness of the mathematical theory of soft models has been clarified relatively recently.

1. Mathematical models of perestroika (rearrangements). We now turn to another very important section that forms in a certain sense "non-linear" thinking, which is so necessary for researchers in our time. From about 1970 information appeared in the press on the creation of a new field of mathematics that was only comparable (as the authors themselves believed) with Newton's invention of differential and integral calculus. Such forecasts of course turned out to be too exaggerated. This new mathematics called the theory of catastrophes arose as a symbiosis of two different branches of mathematics: the theory of singularities of smooth maps by Hassler Whitney and the theory of bifurcations of dynamical systems by Poincaré and Alexander Andronov.

The term "catastrophe" means abrupt responses that arise in the form of a sudden system reaction to a smooth change in external conditions. In the theory of catastrophes problems from various fields of science are solved numerically: from embryology to economics, and from optics to geology. It is important to note that since solutions are always numerical it is impossible to formulate any general patterns of catastrophe evolution. However, a particular case of the theory of catastrophes is the theory of rearrangements. It is in the theory of rearrangements that one can make a number of the most important qualitative conclusions that are the same for any nonlinear system.

Consider a typical formulation of the problem in the theory of rearrangements. Let the system be in a stable state conditionally recognized as "bad", since within the framework of "visibility" there is a more preferable, "better" state. First consider the situation of rearrangements in terms of "housewife".

Suppose that our housewife decides to clean the apartment. It seems that the apartment's condition is bad (it is clear that this is a fairly relative term). When cleaning begins the condition of the apartment deteriorates even more as the disorder increases. And only at the end of cleaning does the condition of the apartment become better than it was originally. After some time, our housewife decides to re-hang the wallpaper in the apartment. What happens with this? It is clear that in this case the condition of the apartment deteriorates much more than during normal cleaning. However, in this case the final condition of the apartment is much better than after the usual cleaning.

What is common in these examples? Any housewife understands two obvious results. Firstly, if you want to improve the condition by means of rearrangements then inevitably you must first get into a worse state. Secondly, the degree of expected improvement is comparable to a preliminary deterioration. Perhaps this is all that a housewife can predict about the patterns of rearrangements. To learn about these patterns in more detail one must already turn to science.

Consider a situation whereby a "bad" state means either an administrative system or a human "illness", and a "good" state is a market economy and "normal health" respectively. The administrative system means the way of organizing social relations characterized by a rigid centralism of the economy. Both the state and man are typical non-linear systems with feedback. This is essential because control without feedback always leads the system to a catastrophe. By feedback in general we mean the following. Suppose there is a system that has an input and an output. There is a signal at the system output. If there is a device that transmits a signal from the output of the system to the input, then this device is feedback. If the feedback signal is transmitted in the same form as it was at the output then this is positive feedback, and if the feedback inverts the signal this is negative feedback.

In some cases, it suffices to destroy only one feedback and the system rushes to a catastrophe. Man, as a state, has quite complex systems with variable feedbacks. Variable feedback is the ability to transmit through the same channel a positive and negative signal. For example, in neurophysiology the role of positive or negative feedback is exercised by various mediators. Let us return to our situation. In Fig. 11-1 along the *Y* axis are the "welfare of citizens" for the state and for a person the "state of health", and on the *X* axis, respectively, are the "enterprise of citizens" and "self-sustaining fluctuations of SFWU". The *Z*-axis in Fig. 11-1 means resistance to rearrangement. Here we have no opportunity to concretize what structural functional working units (SFWUs) are. The dotted line on the graph shows the line of catastrophe (death for a person).

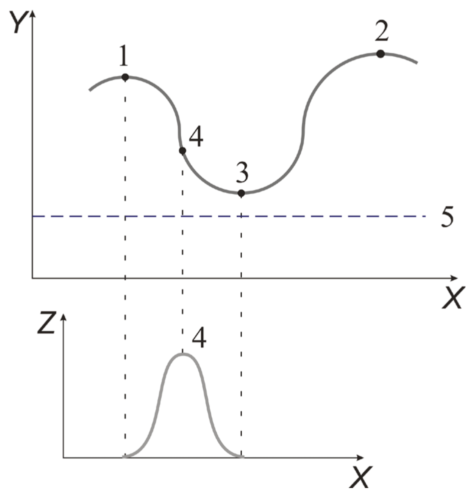


Fig. 11-1.
Rearrangement of the system.
Points: 1 – initial bad state; 2 – good state; 3 – the worst state; 4 – maximum resistance, 5 – the line of catastrophe

So, the patterns are as follows, quoted from the book (Arnold 1990).

1. Gradual movement in the direction of the best condition first leads to deterioration. The rate of deterioration with uniform movement to a better state, increases.

2. As the system moves from the worst state to the best, the resistance to changes in its state, increases.

3. Maximum resistance precedes the worst state that must be passed in order to achieve a better state. After passing the maximum resistance the state continues to deteriorate.

4. As you get closer to the worst state resistance starting from a certain moment begins to decrease, and as soon as the worst state has passed it completely disappears, and the system itself is drawn into a better state.

5. The magnitude of deterioration required for a transition to a better state comparable with a resulting improvement increases with the improvement of the system. A poorly developed system can move to a better state almost without prior deterioration, while a highly developed system, because of its stability, is not capable of gradual continuous improvement.

6. If the system can be dramatically transferred from a bad state to a state close enough to good then it will continue to move in the direction of improvement.

These objective laws of the nonlinear system cannot be ignored. Thus, for example the choice of “treatment” should be such that the maximum deterioration does not cross the line of “death” at the intersection of which the rearrangement is interrupted, and the system tends to catastrophe. With radical changes in the country's economy various troubles for the government are possible before the worst state that is at the maximum of resistance. In the worst state, there is no danger to the government.

The above formulated only the simplest qualitative conclusions. They are not only more important but also more reliable than any quantitative results for a particular model, since they depend little on the details of the functioning of the system.

PART II:

CHEMICAL SYSTEMS

Chapter 12

The transition state lifetime as the fundamental chemical constant

Widely extends chemistry
their hands in human affairs ...
Mikhail Lomonosov

Chemical reactions and their energy. The transformation of one or several initial substances that are different in their chemical composition or structure is called a chemical reaction. Chemical reactions unlike nuclear reactions do not change either the total number of atoms in the system or the isotopic composition of the elements. Chemical reactions occur during the physical contact of reagents, both spontaneously and under the influence of various factors: temperature, catalysts, and radiation, including ionizing, electric current, mechanical stresses, as well as in low-temperature plasma.

Chemical reactions are characterized by the following concepts: the equilibrium degree of transformation determined from the laws of thermodynamics, the reaction rate, and the depth of chemical reaction. The depth of chemical reaction is defined as either the degree of conversion (the ratio of the quantity of the substance which entered into the reaction to that of the initial substance) or the yield of the reaction (the ratio of the amount of the resulting product to that of the initial product).

Photochemical reactions take place under the influence of optical wavelength radiation. The reason for their passage is a change in the electronic structure of molecules when excited by photons. Such excitation is also accompanied by changes in the chemical properties of molecules. A special position is occupied by radiation-chemical reactions that occur due to the absorption of ionizing radiation energy by the substance. Reacting with molecules of the medium and with each other, radiation-chemical reactions lead to the formation of relatively long-lived free radicals and radical ions, as well as various stable reaction products.

Electrochemical reactions are characterized by charge transfer across the boundary between the electrode and the electrolyte. They always go in two directions: cathode and anodic. Chemical transformation in the cathode process is called electroreduction, and in the anodic process it is called electro-oxidation.

Mechanochemical transformations are initiated or accelerated by mechanical stresses. These reactions are possible in both liquids and solids. In liquids cavitation is the effect of shear stresses on the solutions and melts of polymers. In solids such reactions occur for example during dispersion, the action of shock waves or high pressure in combination with shear deformation.

Exothermic reactions are accompanied by the release of heat, and endothermic reactions are accompanied by the absorption of heat.

Catalysis is the acceleration or excitation by catalysts of a chemical reaction with substances that participate in the reaction but are not part of the final products. The catalyst does not shift the chemical equilibrium in the reacting system, it equally accelerates both the direct and reverse reactions. Substances that slow down the reaction are called inhibitors.

The lifetime of the transition state (TS). The TS theory is the simplest and historically the first version of the statistical theory of chemical reactions developed in the 1930s by Eugene Wigner, Michael Polanyi, Henry Eyring, and Meredith Evans. This theory is based on three assumptions, or three axioms.

1. The transition of a chemical system from the initial to the final state is associated with the formation of a TS. For the simplest chemical reaction:

$$A + B \to TS \to C + D \,, \tag{12.1}$$

where A and B are reagents, C and D are the reaction products, and the TS is no longer the reagents but not yet the reaction products.

2. There is a thermodynamic equilibrium between reagents, chemical reaction products and the TS.

3. The speed of a chemical reaction is identified with the decay rate (lifetime) TS – τ.

All three assumptions in the framework of chemistry cannot be rigorously substantiated, and they should in a certain sense be considered axioms of chemistry.

Let us determine the TS lifetime based on the physical meaning of the fundamental constants, taking into account their dimension. Since an elementary act of a chemical reaction takes place at the micro level Planck's constant h, the dimension of which is *energy* × *time,* must be present. The result of a chemical reaction is fixed at the macro level. Consequently, there must be a connection between the micro and macro levels which is realized by means of the Boltzmann constant k_B, the dimension of which is *energy*/*temperature*. Empirical observations suggest that elementary chemical reactions accelerate with increasing temperature. Obviously, the

TS lifetime is inversely proportional to the rate of the chemical reaction. Thus, the shorter the lifetime of the TS, the higher is the temperature. As a result, by comparing the dimensions of the constants h, k_B and the temperature T we obtain the only possible relationship:

$$\tau = \frac{h}{k_B T} . \qquad (12.2)$$

At temperatures that are usual for chemical reactions, the TS lifetime is extremely small: $\tau \sim 10^{-13}$ s. It is clear that direct observation of TS is extremely difficult. If this becomes possible then the statement of axiom 1 will be a proven fact. As can be seen from (12.2), the TS lifetime does not depend on the reagents and products involved in the reaction. Very important result follow from this: the value τ should be considered as a fundamental chemical constant (Khapachev 2000, 3-6).

Thermodynamic equilibrium between reagents and the TS is characterized by equilibrium constant K through which the reaction rate χ constant is expressed:

$$\chi = \frac{K}{\tau} . \qquad (12.3)$$

The constant K is determined by the methods of statistical physics and thermodynamics. According to TS theory the constant K depends on U – the change in the potential energy of the system during the transition from reagents to the TS. The value of U can be interpreted as the activation energy of the reagents. Thus, the reactivity characteristic is the activation energy which practically corresponds to the height of the potential barrier on the multidimensional potential energy surface.

CHAPTER 13

CATALYSIS, ITS CRITERIA AND ESSENCE. THE BELOUSOV-ZHABOTINSKY REACTION

What's in a name? That we call a rose
By any other name would smell as sweet.
William Shakespeare

Catalytic reactions are extremely diverse and numerous. Catalysis is the main part of reactions and in essence the main subject of the study of modern chemistry. Catalytic reactions occur by various mechanisms. We will list them.

1. Free-radical mechanism. This includes: wall catalysis or heterogeneous-homogeneous reactions, catalysis with the development of a chain on the surface of conductors and semiconductors, homogeneous catalysis in solutions.
2. Ion mechanism. This includes: homogeneous acid-base catalysis in solutions, heterogeneous catalysis with bond heterolysis.
3. The mechanism of cyclic transfer of valence electrons in a multiplet-type transition complex.

However, the type of reaction mechanism cannot serve as a criterion for the relation of the process to catalysis.

The main criterion for catalysis is the method of activation. According to the method of activation of the reagent molecules all chemical reactions between valent-saturated molecules can be divided into two large groups. Since the physical essence of activation consists in a complete break ($AB \to A^{\bullet} + B^{\bullet}$) or significant relaxation ($AB \to A \ldots B$) of the initial chemical bonds activation can be carried out as follows. The symbol $A^{\bullet}$ denotes a free radical, which is an atom, molecule, or ion that has an unpaired valence electron.

Firstly, by supplying energy from the outside to the reaction system:

$$AB + Q \xrightarrow{\text{heating}} A \ldots B \to A^{\bullet} + B^{\bullet}, \qquad (13.1)$$

$$AB + h\nu \xrightarrow{\text{photon irradiation}} A^{\bullet} + B^{\bullet}. \qquad (13.2)$$

Secondly, by a preliminary chemical interaction of molecules with a quantum-mechanical system of berthollide type:

$$AB + [C] \rightarrow [C] \begin{matrix} \cdot\cdot\, A \\ \vdots \\ \cdot\cdot\, B \end{matrix} \tag{13.3}$$

according to the multiplet theory (berthollides are compounds of variable composition that do not obey the laws of constant and multiple relations. Here the symbol [*C*] denotes a catalyst. In contrast, daltonids are substances of constant qualitative and quantitative composition). According to the chain theory where *L* is the symbol of the crystal lattice:

$$AB + L^{\bullet} \rightarrow LA^{\bullet} + BL\ , \tag{13.4}$$

$$L^{\bullet} + AB \rightarrow LA + LB^{\bullet} \tag{13.5}$$

Non-catalytic reaction occurs as a result of energy supply from the outside according to (13.1) and (13.2). In catalytic chemistry the activation of molecules and the initiation of reactions are caused by weak chemical interactions which reduce the energy of the initial bonds according to (13.3) – (13.5).

Despite the fact that the two methods of activation of reagents – energy and chemical (catalytic) – are fundamentally different they only represent extremes. In its pure form, catalytic activation is almost never encountered: at low temperatures the reactions are almost never studied, and those best processes of catalysis that nature performs in living organisms are a combination of catalytic and energetic activation but with a clear predominance of the first.

The difference in catalytic and non-catalytic reactions also lies in the presence of a stage in the catalytic reaction that is not present in the stoichiometric reaction – the regeneration stage of the catalytic center. That is, each stage of any catalytic process is a "normal" stoichiometric reaction.

Thus, catalytic reactions are more general and basic reactions with respect to all other chemical reactions.

Essence of catalysis.

1. Catalysis is the most general and common method for performing thermodynamically possible reactions. It consists in lowering the activation barrier by continuously reducing the electron charges and, consequently, the

energy of the initial bonds with the not full-valent chemical interaction of reagents with catalysts.

2. Berthollide type compounds are predominantly used as catalysts. In accordance with this, the main content of catalysis is the interaction of the daltonidic (reagent molecules) and berthollide (catalyst) forms of the chemical organization of a substance. Not all of the catalyst surface is catalytically active but only a small part of it. These are active centers which can be separate *X*-ray amorphous fragments of crystals, small sections of macromolecules, for example, high polymer amino acids in proteins. According to the multiplet theory the minimum number of atoms in the active center is two. In the theory of active ensembles, it can be equal to unity.

3. Four effects characterizing all pictures of a catalysis are a consequence of the not full-valent interaction of molecules of reagents with any catalyst.

3.1. The first effect is the initiation of a reaction that is the emergence of the possibility of redistributing the initial bonds or the generating of free radicals.

3.2. The second effect is the chemical orientation of the reaction, that is the choice of one or another direction of the reaction which depends on the chemical nature of both the reagent and the catalyst.

3.3. The third is the matrix effect. It consists of:

a) providing the catalyst with its surface as a matrix for the reaction (the surface of a solid catalyst, vessel wall or structural unit of solution, for example, a solvate complex), and this increases the number of reagent molecule encounters and the probability of a favorable configuration of the active complex; and

b) the influence that the geometric factor has on the initiation and orientation of the reaction.

Solvation is the molecule of the dissolved substance being surrounded by a shell of solvent molecules.

3.4. The fourth effect is the energy recuperation of a chemical reaction. This energy is not immediately given to the external environment but is captured by that part of the catalyst which is called the weighting agent or aggravator. An aggravator contains active centers and thus acts as an energy trap; the energy goes to the excitation of the active centers. Herewith, the activity of the center increases as the complication of the aggravator according to the exponential law.

4. All four effects appear simultaneously and are mutually conditioned. They seem to reinforce each other. In various cases of catalysis any of them may play a decisive role.

5. It is wrong to define catalysis solely as acceleration of the reaction in the presence of the substances remaining chemically unchanged after the reaction.

Potential barriers restrain many reactions, and they do not spontaneously proceed until a catalyst intervenes. Factual material about irreversible chemical changes (for example, self-adaptation) of functioning catalysts has been accumulated but there will be more about this in the next section on evolutionary catalysis.

6. It is wrong to speak of "anti-catalysis" as is sometimes allowed. The use of additives that slow down the reaction or inhibition has a completely different nature with respect to catalysis.

Catalysis and inhibition can and should be considered as process control methods. But according to the mechanism of action the catalysts and inhibitors are completely different. There are no such inhibitors that would increase the potential barrier of reactions, increasing the energy of the initial bonds, when interacting with reagents just as catalysts reduce this barrier, reducing the energy of the bonds.

7. The division of catalysis into three types: homogeneous, heterogeneous and enzymatic corresponds to the origin and location of the catalytically active center. In the first two cases we are talking about artificial systems containing a catalyst in a homogeneous solution or gas (homogeneous catalysis) and on the interface (heterogeneous catalysis).

In the third case it is meant that the catalysts are protein molecules of enzymes that function in a living organism or are secreted from it. From the point of view of the phase composition of the enzyme in solution, given the large size of protein molecules we can speak of the microheterogeneity of the enzymatic process.

The feature of biochemical transformations is that a single-stage process is replaced with a multiple transformation through several transitional states and through metastable intermediate products with the minimum energy of activation. The following scheme illustrates the difference in the cross-section of the potential energy surface of "non-living" and "living" transformations.

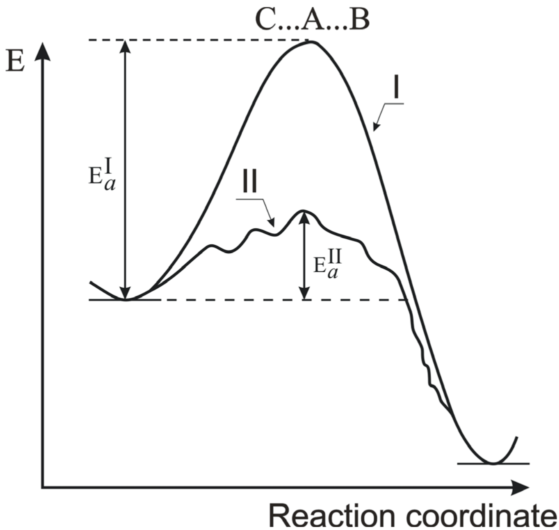


Fig. 13-1.
The distinction of the cross-section of the surface potential energy of “inanimate” and “living” transformations:
I – a one-step chemical process with activation energy E_a^I;
II – a multistage biochemical process with activation energy E_a^{II}.

In addition, in enzymatic catalysis the molecule being converted is not only exposed to several active enzyme groups but also participates in changing their orientation and closes the chain of redistribution of bonds, i.e., the system self-organizes. Similar phenomena of self-organization were observed with inorganic catalysts (a number of palladium-based alloys).

8. The development of metal complex catalysts has confirmed that all three types of catalysis are a single category of phenomena. Almost all known heterogeneous catalytic reactions can occur under the influence of homogeneous complex catalysts and vice versa. The active center of the enzyme in a large number of cases is also a metal complex. The reasons for the special selectivity of artificial metal complexes and enzymes turned out to be similar: it is the spatial structure and location of the ligands surrounding the atom or metal ion.

In biochemistry a ligand is a chemical compound (often but not always a small molecule) that forms a complex with a particular biomolecule (most often a protein but sometimes for example with deoxyribonucleic acid) and produces certain biochemical effects. In the case of ligand binding to a protein, the ligand is usually a small signaling molecule that binds to a specific site of binding on the target protein (for example a receptor). In the case of ligand binding to deoxyribonucleic acid (DNA) the ligand is usually also a small molecule or an ion or protein that binds to the DNA double helix.

9. Typically proteins are considered to be the main functional macromolecules because nature has chosen them as intracellular catalysts. However, as Sidney Altman and Thomas Cech discovered in the work awarded the Nobel prize, ribonucleic acid (RNA) can also play the role of a catalyst. It is reasonable to assume that RNA and DNA are catalytically active.

10. There are reactions in which one of the products serves as a catalyst for the transformation of the starting materials. These reactions are called autocatalytic reactions. For example, the acid formed during the hydrolysis of esters catalyzes this hydrolysis. For some time (the induction period) the rate of autocatalytic reactions is very small but as the catalyst product accumulates it increases and reaches a maximum and again decreases due to the consumption of the initial substance. The autocatalytic character has self-conjugated and some oscillatory reactions play a large role in biochemistry. These processes require more attention and are outlined below.

The Belousov–Zhabotinsky reaction. Phenomena of self-organization leading to the formation of spatial and spatial-temporal dissipative structures can occur when some non-linear chemical reactions take place in which the reaction products are its catalysts and accelerate the reaction itself. This leads to the fact that the rate of formation of the reaction product depends nonlinearly on the concentrations of the reacting components.

A symbolic scheme of a typical one-step chemical reaction can be depicted as follows:

$$A + B \xrightarrow{\chi} C + D \,, \tag{13.6}$$

where A and B are the reagents, C and D are the reaction products, χ is the rate constant depending in a general case on temperature and pressure. There are reversible reactions for which the concentrations of reagents A and B never turn out to be zero. The equilibrium constant K is determined in this case through the ratio of the concentrations of c_A, c_B of reagents and c_C, c_D of the reaction products:

$$K = \frac{c_C c_D}{c_A c_B} \,. \tag{13.7}$$

Thus, a detailed balance is established in the system. This equilibrium is ensured by the flow along with (13.6) of the reverse reaction:

$$C + D \xrightarrow{\chi'} A + B \; . \tag{13.8}$$

The principle of detailed equilibrium ensures the occurrence of reactions (13.6) and (13.8) with identical rates. This situation is typical for isolated systems.

We can make this system open. To do this we will maintain constant concentrations of products c_C and c_D by constantly removing them from the reaction volume, thereby realizing an open non-equilibrium system. Under such conditions the ratio of concentrations in (13.7) is no longer equal to the equilibrium constant *K*. This is a stationary non-equilibrium state of a chemical system, at which detailed equilibrium no longer holds.

If in such a system, there are positive feedbacks then an unstable state occurs. Positive feedbacks in this case reinforce random changes in the rates of direct or reverse reactions. The most typical way of implementing feedbacks in a chemical system is autocatalysis when the reaction product increases its rate of flow.

Let us now consider, by the example of the Belousov–Zhabotinsky reaction, the influence of non-equilibrium conditions on the modes of its behavior. This famous reaction opened a new era in chemical autocatalytic reactions. It was opened by Boris Belousov in 1961. Later, in 1964 Anatol Zhabotinsky studied this reaction in detail.

Belousov discovered the reaction during a simple experiment – the oxidation of citric acid with potassium bromate in the presence of cerium sulfate and sulfuric acid. Cerium, a metal of variable valence, plays the role of a pendulum in it. Two flow directions compete in this reaction. In one case, the cerium ion Ce^{4+} is reduced:

$$Ce^{4+} \rightarrow Ce^{3+} + Br^{-} \; . \tag{13.9}$$

In another case on the contrary the cerium ion Ce^{3+} is oxidized:

$$Ce^{3+} \rightarrow Ce^{4+} \; . \tag{13.10}$$

These reactions are autocatalytic. The free bromine ion Br^{-} acts as a potent inhibitor of the reaction (13.9). Therefore, at first the only reaction (13.9) proceeds until all of the Ce^{4+} is reduced to Ce^{3+}. Then reaction (13.10) proceeds and the process repeats from the beginning. As a result, the solution periodically changes its color, becoming blue (an excess of cerium ions Ce^{4+}) then red (an excess of ions Ce^{3+}). Thus, a chemical clock is obtained on the basis of a peculiar oxidative-reducing chemical pendulum.

Chapter 14

Conceptual systems of chemistry

Everybody must do good,
used in its place.
On the contrary: the exercises
of the best dance master
in chemistry are irrelevant;
The advice of an experienced
astronomer in dancing is stupid.
Kozma Prutkov

Based on the historical and methodological analysis of the development of scientific chemistry Vladimir Kuznetsov established the following patterns (Kuznetsov 1973). According to Kuznetsov the entire history of chemistry fits into the framework of the development and emergence of new conceptual systems which are a collection of laws, theories, and views that characterize the level of knowledge of a chemical phenomenon.

Fig. 14-1 shows a diagram of the development of conceptual systems of chemistry according to Kuznetsov.

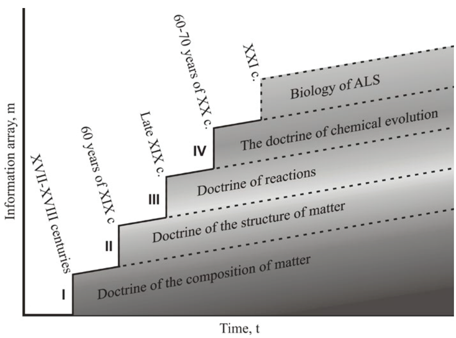


Fig. 14-1.
The development of conceptual systems according to Kuznetsov
ALS – artificial living systems

From the time of the atomic-molecular theory when scientific chemistry actually emerged four conceptual systems are clearly distinguished in chemical science: the theory of the composition of substances, the theory of structure, the theory of reactions, and the theory of chemical evolution. As the main criterion of conceptual systems, a method for solving the main problem of chemistry, the problem of the genesis of properties and reactivity which determines the transition from one conceptual system to another, is chosen.

For a long time, the central problem of chemistry was the ratio of composition and properties, the desire to explain all the diverse properties of chemical compounds by their elemental composition:

composition • → • properties.

In the subsequent period the structure became the central concept of chemistry; it was on its basis that the diverse characteristics of numerous chemical compounds found their explanation and the way to the synthesis of new substances was opened.

The emergence of ideas about the structure led to the differentiation of the concept of property with the release of a new concept – reactivity (or reactive ability – r.a.) or a more general function as a derivative and from the concept of structure and from the concept of properties (Fig.14-2).

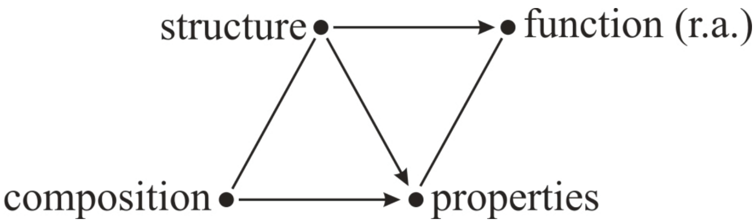


Fig. 14-2.
Scheme of the evolution of chemical studies from composition to function (r.a.)

Reactivity in turn also demanded the taking into account of kinetic factors on which it depends at least to the same extent as it does on structural factors. The categorical scheme in the transition to kinetic theories is complemented by the concept of organization which, similar to the concept of structure, is an attribute of a chemical system but already more complex than a molecule of a kinetic system (Fig.14-3).

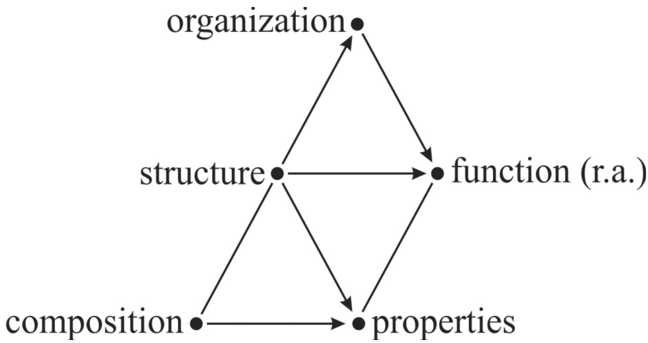


Fig. 14-3.
Scheme of the evolution of chemical studies from composition to organization

The logical conclusion of this scheme is the transition to the study and theoretical description of systems of the highest level of organization, for example, open, self-organizing catalytic systems (Fig.14-4).

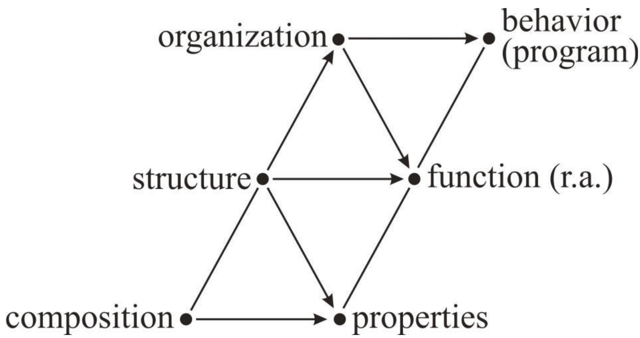


Fig. 14-4.
Scheme of the evolution of chemical studies from composition to behavior (program)

Now a new concept of behavior has emerged that expresses something different from the concepts of properties and reactivity. The problems of chemistry underlying the fourth conceptual system are connected with this. It explores the question of the behavior of complex prebiological systems and the levels of their organization. The key issue here at the present time is the determination of the behavior of systems depending on their organization.

Each conceptual system is prepared by all previous developments in chemistry and is based on all previous systems.

The theory of chemical evolution (fourth conceptual system) emerged on the basis of modern concepts of chemical kinetics, thermodynamics and catalysis (refer to the third conceptual system), namely: theories of a transition state, thermodynamics of irreversible processes and evolutionary catalysis.

Nevertheless, the origins of this new conceptual system are directly in catalytic chemistry. In the 1950-60s empirical foundations were laid for the ideas of self-adaptation or self-movement of the composition and structure of catalysts towards higher levels of organization. A theoretical interpretation of these ideas was given by Alexander Rudenko in the theory of self-development of open catalysis (Rudenko 1969).

From Fig.14-1 it is clear that this conceptual system, being a prerequisite for a fundamental explanation of the mechanism of biogenesis, is a kind of upper boundary of chemistry (there is no border right) in which it exhausts itself with regard to the further development of its own conceptual systems. It thus creates a scientific-theoretical and experimental base for the substantiation of biological laws and the construction of theoretical biology.

CHAPTER 15

CHEMICAL EVOLUTION AND BIOGENESIS

The important thing is not that strictly
but what is true
Andrey Kolmogorov

One of the main problems of modern science is that the question of the origin of life is directly related to chemical evolution. The evolution of chemical non-equilibrium systems is a path leading to spatial, temporal, and spatial-temporal ordering. In this chapter we will try to consider the principles of the emergence, development and self-organization of chemical systems and highlight those signs that are amplified and begin to dominate as chemical systems approach biological ones.

The basis of chemical and biological evolution is morphogenesis i.e., the directed development of structures. Chemical evolution is usually understood as the origin and progressive development of the chemical organization of a substance as a whole, in all its manifestations.

The chemical organization of a substance can be in the form of:

1) daltonid compounds;
2) berthollide compounds;
3) a transition state.

Yuri Zhdanov defined chemicals as discrete points in the continuum of transition states (Zhdanov 1977, 102-14). Based on this there are two aspects of chemical evolution: molecular structural, and functional (kinetic).

Molecular-structural evolution. In this case, as a result of generalizing various approaches to chemical evolution (geochemical, cosmochemical, biogeochemical, and biochemical), evolutionary hierarchy schemes are developed that indicate different directions of substance development, for example: molecules of type $A_1, A_2, ...A_n \rightarrow$ atomic crystals, and etc., as shown in Fig. 15-1.

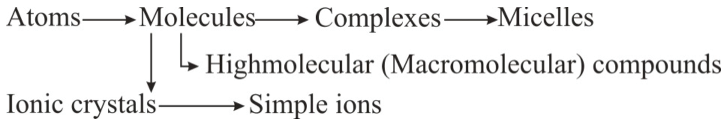


Fig. 15-1.
The scheme of evolutionary hierarchy

In a more general form the sequence of forms of the evolution of substances is depicted in Fig. 15-2.

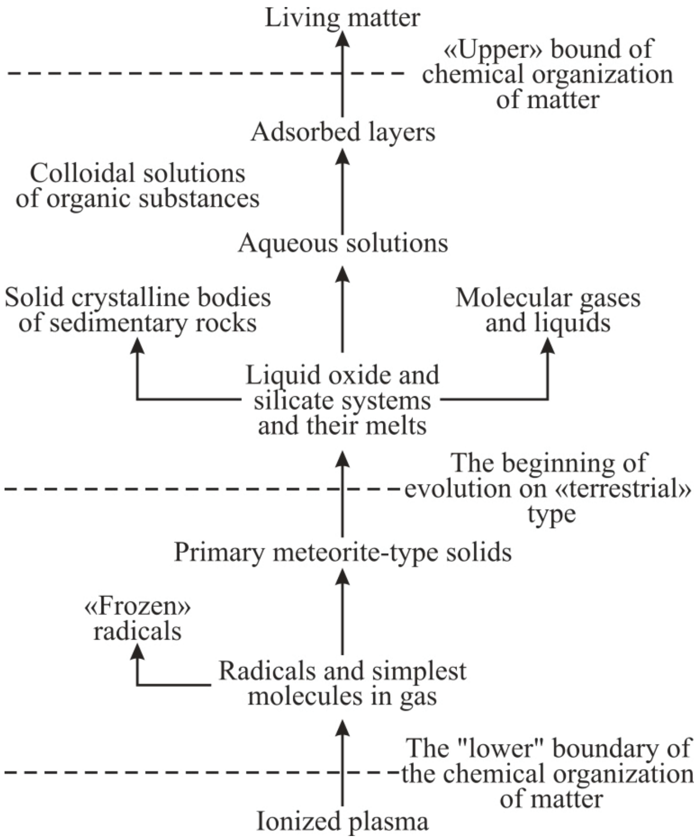


Fig. 15-2.
The sequence of forms of evolution of substance

Thus, according to the general ideas about the levels of organization of a substance, the most highly recognized is the substance – the carrier of life, and then consciousness. Under terrestrial conditions, such a higher level of organization is complex organic compounds – proteins, enzymes, nucleic acids, etc.

Chemical evolution in the direction of live substances in the molecular structural aspect is represented in the form of stages:

I – the simplest inorganic and organic substances (H_2O, CO_2, CO, CH_4, NH_3, HCN, CH_2O, H_3PO_4, H_2S, etc.);

II – small biomolecules or monomers (amino acids, purines, pyrimidines, carboxylic acids, alcohols, monosaccharides, etc.);

III – complex organic substances and biopolymers (proteins, nucleotides, polysaccharides, lipids, etc.);

IV – supramolecular complexes of biopolymers;

V – living cells (or extracellular organisms – viruses).

In experiments on biogenesis, it was shown that almost all molecular components of living matter can be synthesized abiogenically from the simplest substances with the involvement of non-chemical energy sources that are real for the conditions of the primary Earth. These are such sources as: ultraviolet rays, electrical discharges, radioactive radiation, the heat of volcanic eruptions, shock waves, etc.). This period of chemical preparation – a period of intense and diverse transformations, about which one can only build cautious hypotheses – was replaced by a period of biological evolution.

From the above we can draw the following conclusions:

1. Within the limits of the general direction of chemical evolution the main line leading from the level of chemical elements to living matter is distinguished. All other directions can be attributed to side or dead-end directions.

2. The direction of chemical evolution is partly in the chemical itself – the selection of elements and structures that have the greatest complexity, i.e. have the greatest evolutionary potential, information capacity and integrative ability (see below), and partly due to the environment.

3. In the course of evolution, those structures were selected that provided the most advanced types of communication (including feedback) and regulation. Here we have in mind not only the chemical bonds but also any other communication that provides the relationship between systems, subsystems and elements of the systems.

The first and simplest of these structures can be called different phase boundaries. They served as the basis for chemical and physical adsorption. Adsorption introduced elementary ordering into the mutual arrangement of particles. It increased the concentration of particles. In addition, adsorption served as a factor in the appearance of a catalytic effect.

The second structural fragment is the groupings that ensure the processes of electron and proton transfer. These include semiconductor circuits and structures responsible for the so-called transhydrogenation or hydrogen transfer. These types of structural fragments are associated with the need to attract carbon as well as other organogens that can form double bonds and serve as electron donors or acceptors. Organogenic elements include C, H, O, N, P, and S, the total mass fraction of which in organisms is 97.4%. They are followed by 12 elements that take part in the construction of many physiologically important components of biosystems: Na, K, Ca, Mg, Fe, Si, Al, Cl, Cu, Zn, and Co. Their mass fraction in organisms is approximately 1.6%. However, it is believed that almost all chemical elements can be present in living organisms. Thus, 76 elements were detected in human serum at concentrations of 10^{-4} mg/ml and above. These groupings also serve as the beginning of redox or acid-based catalysis.

The third structural fragment necessary for evolving systems is the groupings responsible for energy supply. These include hydroxy-oxo groups, phosphorus-containing and other fragments with high-energy bonds. It was suggested that these structural units also play the role of a catalyst in relation to a number of reactions but most likely their purpose is to lift thermodynamic prohibitions by conjugating disproportionation reactions and breaking macroergic bonds with enzymatic reactions.

The next fragment of the evolving systems is the complex polymer structure of the RNA or DNA type which performs a number of functions inherent to the listed structures, and most importantly the role of a template or a catalytic matrix on which similar structures are reproduced.

One of the very important structural fragments of biochemical systems is the structure of porphine or more precisely the porphyrin structures that already contain metal-nitrogen-carbon bonds. It is known that porphyrin compounds containing iron, magnesium, copper, zinc, and cobalt, play a prominent role as vital fragments.

Part III:

Living Systems

Only, I swear by Zeus, the Athenians,
you will not hear an elaborate speech
adorned with their different phrases
and expressions, I will simply speak,
the first words that have come to me,
because I am convinced of the truth
of my words, and let none of you
expect anything another; and it would
not be necessary for me at my age
to speak before you, the Athenians,
like a young man, with a written speech.
Socrates' apology. After the accusatory speeches. Plato

Biology, physics and chemistry – conceptual sections of natural science. They make it possible to study the structure and functions of living information systems down to the people of the modern type.

The main problem of biology is to understand all the phenomena of living nature on the basis of scientific laws, taking into account that the whole organism has properties that are fundamentally different from the properties of the parts that make it up.

Oddly enough, we cannot yet give a strict definition of life and cannot say how and when it originated. In general, definitions in biology are difficult and usually, the stricter the definition the worse it works. The maximum that is currently possible is to list and describe those features that distinguish it from inanimate matter and express scientific hypotheses about its origin and evolution. We list some of the main features of living systems: nutrition, respiration, irritability, mobility, excretion, reproduction, and growth from the inside (crystals grow from the surface). The listed signs of life are only observable manifestations of the main properties of living matter, that is to say, its ability to extract, transform and use the energy received from outside. They are all known from the school biology course and do not need comments. In addition to these, there are of course other no less “main” ones.

Information about the origin and evolution of man is obtained from research in the fields of paleoanthropology and genetics, and also in evolutionary psychology, allowing us to answer many questions that have long been of interest to a reasonable person. Some of these answers will be presented in subsequent chapters. The authors of course do not pretend to give a full and comprehensive presentation of the data of human evolutionary biology, a detailed analysis of the concepts of anthropogenesis as well as other latest achievements of modern biology.

In this regard it is appropriate to recall the decree of the Russian emperor Peter the Great:

> An infantry officer passing by the cavalry unit must dismount and hold his horse to the bridle so that his appearance does not cause ridicule among these cavalrymen.

We ask both professionals and ordinary readers to assume that the authors dismounted in this part of the book, as strictly speaking in the previous ones.

Chapter 16

Biological axioms and biological constants

Mirum non est, quod ita habet super nos talis multum potentia; post omnes, quid nos vivere, est temere.
Seneca

We know only one version of life, the one that is on Earth. We don't know which of the properties of life are mandatory for any life at all. Even so, let's risk naming a few of these properties. In this chapter they will be formulated in the form of axioms. In addition, we will first formulate two basic statements that are common to any variant of life. The first is the presence of hereditary information. The second is the active implementation of functions aimed at self-maintenance and reproduction as well as obtaining the energy necessary to perform all this work.

All life on Earth solves these problems with the help of three classes of organic compounds: DNA, RNA and proteins. DNA provides storage for hereditary information. Proteins perform all types of active work. The division of labor in DNA and proteins is quite strict; proteins do not store hereditary information, and DNA does not perform active work. In biology unlike physics there are almost no rules without exceptions. In this case there are exceptions too but these are very rare.

Finally, RNA molecules serve as intermediaries between DNA and proteins ensuring the reading of hereditary information. With the help of RNA, proteins are synthesized in accordance with the instructions written in the DNA molecule. Some of the functions performed by RNA are very similar to the functions of proteins (active work on reading the genetic code and protein synthesis). Other RNA functions resemble DNA functions (storing and transferring information). And all this RNA does with the active assistance of proteins. At first glance, RNA seems to be the "third wheel".

We now give one fundamentally important experimental fact that we associate with the manifestation of the "zero axiom of biology." We introduced such a numbering because in due time Boris Mednikov proposed biological axiomatics consisting of four axioms (Mednikov 1982). We will set out these axioms later after the zero axiom.

The zero axiom

In 1847 Louis Pasteur made a fundamental discovery which for a long time could not be explained and was appreciated only in the XX century with the advent of quantum mechanics. Pasteur found that in living organisms, amino acids and sugars are optically active; that is, they rotate the plane of polarization (two of the electric field strength vectors **E** and magnetic **H** fields around the velocity vector **v**) of the optical radiation incident on them. The following fact turned out to be surprising. All of the natural amino acids are left-handed stereoisomers, i.e. the **EH** plane rotates to the left, see Fig. 16-1a; and all natural sugars are right-handed stereoisomers, i.e. the **EH** plane rotates to the right, see Fig. 16-1b.

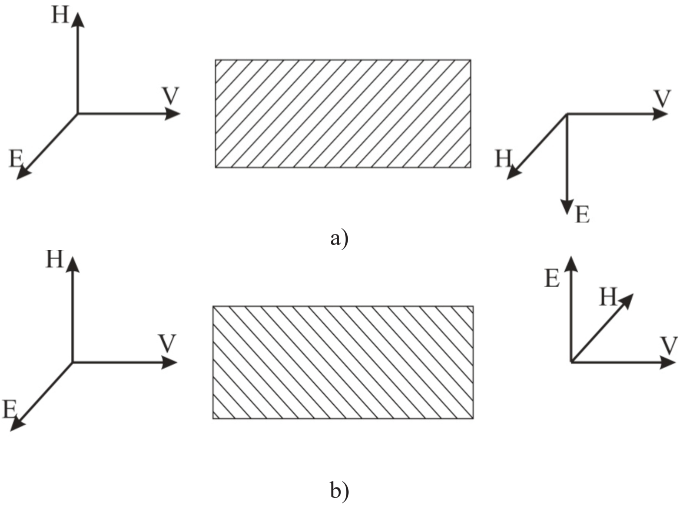


Fig. 16-1.
Rotation of the polarization plane in an optically active medium:
a) left-hand stereoisomer; and b) right-hand stereoisomer

During the artificial synthesis of these biochemical compounds racemic (optically inactive) mixtures are formed. They have an equal concentration of left- and right-handed molecules which is, from a thermodynamic point of view, the most favorable state. Therefore, the discovery of Pasteur is not only one of the important criteria that distinguish the living from the inanimate but it also requires an explanation of this asymmetry. In other words, any scientific hypothesis explaining the origin of life must take into account the conditions under which a chiral system, that is a system with broken symmetry, arose from an initially symmetric system.

One of the possible explanations for this fact as a result of which chirally pure living systems could be formed in an evolutionary way, is as follows. The possibility of forming a system with reduced symmetry arises as a result of bifurcation during self-organization.

Thus, the emergence of life should be considered as a “favorable” catastrophe that occurred at the bifurcation point. It is clear that with this approach the theory of the explanation of life should proceed in the field of condensed matter physics, synergetics, and chemistry. In addition, self-reproduction at the molecular level of chirally pure systems being the problem of these theories becomes another sign of the life.

At present, evolutionary theories of the origin of life are being replaced by theories of a jump based on the theory of bifurcations and catastrophes. Thus, in principle it is possible to calculate the time of the origin of life. The combination of these ideas formed the idea of a “biological explosion”.

We now formulate the zero axiom in the following form (Khapachev 2000, 3-6).

Zero axiom. The natural origin of life is associated with the appearance of the chiral asymmetry of its main carriers which is manifested in the selection of certain forms of chiral carriers.

In this form, the formulation of the axiom is functional, that is, the formulation does not depend on what the carrier of life is based on. The following is important. If life arose in the process of self-organization, then the need arises for the selection of certain forms (left or right) of chiral carriers. In the artificial reproduction of this process such a violation could not have happened.

We now turn to the formulation of the remaining axioms of biology first formulated by Mednikov.

The first axiom

John von Neumann set the following problem based on the results of the works of Norbert Wiener on cybernetics and Claude Shannon on information theory. Is it possible to build a machine that following its own instructions would create an exact copy of itself? In other words, is it possible to build a self-replicating automaton? The problem was solved and resulted in the following. You can create such a machine but there is a certain threshold of complexity below which it cannot reproduce its own kind. The limiting complexity turned out to be not so big, about 10^6 bits, i.e. the machine must consist of no less than 10^4 elements. Neumann's theorem does not give an answer about what these elements are, how to assemble them, etc. The importance of the theorem as we shall see lies elsewhere.

At first glance, it seems there are no obstacles for the “biological reproduction” of machines. From the solution of the problem it turned out that the “daughter” machine would be fruitless since it does not already have a self-reproduction program. Therefore, for the appearance of the third generation it is necessary to provide a copying device in the “mother” machine which also inherits a copy of the self-reproduction program. Thus, according to von Neumann it is not the structure that is inherited but a description of the structure and instructions for its manufacture. As a result, the entire reproduction process consists of two different operations: copying a program called a genotype, and constructing the organism itself – the phenotype. Thus, the first axiom can be formulated as follows.

The first axiom. All living organisms should be a unity of phenotype and genotype (i.e. a program for building a phenotype). Only the genotype is inherited from generation to generation.

This axiom is also functional in this formulation. It is not associated with any specific chemicals that form the basis of life. Under terrestrial conditions the basis of the phenotype is proteins and the basis of the genotype is nucleic acids. Here we do not dwell on the recently discovered special form of heredity. The so-called prion proteins are able to transfer information about their spatial form from one protein to another without the participation of DNA. In another galaxy or planetary system life can be built on a different structural basis but according to the first axiom principle common to the whole Universe. That is, the principle of separate copying of the phenotype and genotype remains unshakable. However, life on the basis of only one phenotype or one genotype is impossible since neither the self-reproduction of the structure itself nor its self-maintenance can be achieved.

The second axiom

In 1927 an outstanding biologist Nikolai Koltsov formulated a statement which was later given the status of the second axiom of biology by Mednikov. This principle remains unshakeable to this day despite the fact that since the time of Koltsov, ideas about the nature of hereditary molecules have changed cardinally. This is due to the fact that like the previous axioms the formulation of the second axiom is functional.

The second axiom. "Hereditary molecules" are synthesized by a matrix method in which the genes of the previous generation are used to build the genes of the future generation.

In the 1950s the structure of the "hereditary molecule" was deciphered by Francis Crick, James Watson, Maurice Wilkins, and Rosalind Franklin. They also showed that the hereditary molecule itself has a mechanism for matrix copying. For our life the hereditary molecule was DNA (for some viruses, RNA).

Thus, the second axiom postulates that life (in its statics) is a matrix copy followed by the self-assembly of copies. The principle of matrix copying itself is based on the complementarity property of nucleotide residues. Complementarity is due to the hydrogen bond.

The composition of RNA consists of four nitrogenous bases: A (adenine), U (uracil), G (guanine) and C (cytosine). In the DNA molecule thymine (T) is present instead of uracil (U). Moreover, A–T and G–C are mutually complementary. Thus, DNA molecules (like RNA) are capable of self-copying. However, catalysts are needed for this – proteins or ribozymes. Inherited information that is stored in DNA as a sequence of nucleotides can be copied to messenger RNA (mRNA) and vice versa. Copying accuracy is ensured by the property of nucleotides called complementarity. Against each nucleotide of the initial molecule (matrix) in the synthesized copy (replica) there can only be one strictly defined nucleotide from the four possibilities. Cytosine (C) is always opposite guanine (G); and adenine (A) is opposite uracil (U), or thymine (T) (the replacement only in the DNA molecule). When a new replica is synthesized on this replica it will be an exact copy of the original molecule.

The genetic code is a universal way for all living things, by means of which the primary structure of a protein molecule (amino acid sequence) is encoded in a DNA molecule (or RNA). Each amino acid is encoded by three nucleotides (a codon or triplet). There are only four nucleotides in the DNA so they can form $4^3 = 64$ different triplets. There are only 20 amino acids in proteins so the genetic code is redundant; many amino acids are encoded not by one but by several interchangeable codons. The reading of genetic information occurs in two stages. First, the information is copied from DNA to RNA (transcription). This operation is performed by a special enzyme, DNA-dependent RNA polymerase. The resulting transcription RNA molecule containing the "instructions" for protein synthesis is mRNA. The implementation of this "instruction", that is protein synthesis (translation), is carried out by ribosomes. The transfer of genetic information from generation to generation is carried out thanks to replication. Replication is the process of synthesis of the daughter DNA molecule on the matrix of the parent DNA molecule. During maternal cell division, each daughter cell receives one copy of the DNA molecule which is identical to the DNA of the mother cell. DNA replication is performed by an enzyme complex consisting of 15–20 different proteins called a replica.

The third axiom

It was found that in order to cause a single mutation, it is required to bring the minimum energy

$$E_{\min} \sim 2.5 \div 3\,\text{eV} \tag{16.1}$$

to the DNA. The average energy of thermal motion of molecules surrounding DNA at temperatures that are common for a living organism is approximately $E_{av} \approx 2.5 \cdot 10^{-2}$ eV. That is, at physiological temperatures DNA is quite stable. But the problem is not as simple as it seems at first glance. As is well known, the speeds of molecules during random motion are unequal even in the state of thermodynamic equilibrium. For example, for an ideal gas they obey the Maxwell distribution which is a special case of normal distribution (9.8), see Fig. 9-3. Consequently, there are always molecules with energies that are sufficient to disrupt the structure of a gene and cause a mutation. In fact, such changes in genetic programs are a consequence of the statistical distribution of molecules in energy. Thus, mutations have the following properties: they are random, unpredictable and not directed. Therefore, mutations can be adaptive only by chance.

However, the thermal motion of molecules does not play a dominant role in the process of mutagenesis. Much more important for mutagenesis are high-energy radiation photons (ultraviolet, *X*-rays, and gamma-rays), fast elementary particles and molecules of substances capable of reacting with DNA (chemical mutagens).

In order for radiation or a particle to cause a mutation it is necessary for them to be in a small area near the DNA. The size of this region is known as the radius of the effective exchange of R_{eff}:

$$R_{eff} \approx 10^{-9}\ \text{m}\,. \tag{16.2}$$

It was first measured by Nikolai Timofeev-Resovsky and Karl Zimmer. Therefore, since the order of magnitude of R_{eff}: refers to the microworld the Heisenberg uncertainty principle (6.3) operates.

Consequently, the process of mutagenesis in this case also has a probabilistic nature. A similar probabilistic picture arises when considering other mutagens. Summing up, we can formulate the following principle.

The third axiom. In the transfer process from generation to generation genetic programs for many reasons change randomly and undirectedly, and only accidentally do these changes turn out to be adaptive.

This axiom is based on the principles of statistical physics and the Heisenberg uncertainty principle. In addition, its existence is due to the presence of two characteristic parameters E_{min} and R_{eff} which are a kind of fundamental biological constant (Khapachev 2000, 3-6). It is important to understand that the role of these fundamental constants lies not in their numerical value but in their conceptual value. The existence of these biological constants determines the possibility of random and undirected changes in genetic programs. Thus, they determine the minimum energy and the effective distance for the implementation of the mutation.

Imagine that $E_{min} = 0$. Then mutations come from any particle with an arbitrarily small energy, i.e. they go too fast. Such an organism is obviously not able to exist. A similar conclusion will be obtained if $R_{eff} = \infty$. On the contrary, if $E_{min} = \infty$ then mutations are impossible which means that the species cannot change and adapt to possible changes in the environment. A similar conclusion will be obtained if $R_{eff} = 0$. As it can be seen, life is impossible in these extreme cases. These extreme cases actually mean the absence of biological constants.

Note that unlike physics, in biology there are practically no rules without exceptions. Therefore, the third axiom formulated above is valid in the first approximation. Some exceptions to the third axiom are given in the book of Alexander Markov (2014).

The fourth axiom

This axiom, called the principle of amplification, owes its origin to the eminent geneticist Timofeev-Resovsky. It is easiest to understand this principle from the example given by Vadim Ratner.

Suppose there is a fertilized egg cell – the carrier of the mutation of the gene encoding an important functional carrier for life – an enzyme without which a living organism cannot survive. In the process of growth and development of the organism the egg cell became 10^{15} cells respectively, and the genes multiplied. Each gene produces 100 mRNA molecules, each of which synthesizes 10^2 enzyme molecules. Each molecule of the enzyme performs 10^4 acts of its own (important for the life of the organism) reaction. As a result, we have $10^{15}\cdot 10^2\cdot 10^2\cdot 10^4 = 10^{23}$. Thus, the possible number of acts of the enzymatic reaction is comparable to the Avogadro number which means that this is already a macro level. This is the magnitude of the increase in the results of a single quantum jump, a single mutation that exists at the micro level. Natural selection can already work with the macro level. It is important that it acts, not directly on genetic programs but on the phenotype in which each change is amplified 10^{23} times. Thus, we can formulate the principle of amplification as follows.

The fourth axiom. Random changes in genetic programs in the formation of the phenotype are magnified and subjected to the selection of conditions by the external environmental.

It should be noted that due to the increase of random changes in phenotypes the evolution of living nature is fundamentally unpredictable. One can only say that natural selection allows breeding mainly for the descendants of those individuals that are best adapted to environmental conditions. However, how they will be adapted one can only guess.

CHAPTER 17

DIGEST OF CONCEPTUAL ACHIEVEMENTS OF MODERN FUNDAMENTAL BIOLOGY

'We be one blood, thou and I', Mowgli answered.
Rudyard Kipling

In this chapter we will present, in our opinion, the most interesting conceptual discoveries in biology over the past half-century.

The RNA world

Until the beginning of the 1980s there was an unsolvable problem in biology. Which molecule appeared first: DNA or protein? RNA was usually forgotten due to the fact that for a long time it was believed that RNA could neither store information like DNA nor function as a protein. Over time, it turned out that for many viruses, hereditary information is not stored in DNA but in RNA. At first it was considered an exception to the rule. A revolutionary event for all biology occurred in the 1980s when ribozymes were discovered – RNA molecules with catalytic properties. Thomas Check and Sidney Altman received the Nobel Prize in Chemistry for this discovery in 1989. Among the ribozymes were found catalysts for the replication of RNA molecules, of their own and others.

Now, RNA has become in a certain sense no less important a molecule than DNA because only RNA can perform two main tasks at once; although DNA copes better with the task of storing information, and proteins are better at "work". Consequently, on the basis of RNA there can be a living organism that has neither proteins nor DNA, and all functions are performed solely by RNA molecules. As a result, it became clear that RNA-organisms could acquire proteins and DNA later in the process of evolution but at first, they did without them.

As a result, the RNA-world theory emerged which is apparently the most outstanding achievement of theoretical thought in biology in recent decades. In this theory the first living organisms were RNA-organisms without proteins and DNA. The prototype of the future RNA organism was probably an autocatalytic cycle formed by self-replicating RNA molecules. Candidates for this role can be precisely those ribozymes that are able to catalyze the synthesis of their own copies.

The possibility of inheriting acquired traits

From a certain point on, there was an opinion that genetics denies the possibility of inheriting somatic mutations. Are there any exceptions to this rule? Somatic mutations are changes of genes in certain cells during the period of individual development of the organism. There is a traditional view that changes in the cells of the body cannot affect the genes of germ cells. In most cases, this statement is most likely true. However, nature is immeasurably more complicated than any of our models, rules, and theories. And in this case, there are exceptions.

In 1885 August Weismann formulated the principle which was later called the “Weismann barrier” (WB) and in the XX century acquired the status of a dogma. In essence, WB means the following. Mutations in somatic cells (cells of the body) are never transmitted to germline cells (reproductive cells), i.e. body cells cannot transmit information to germ cells. Thus, no external influences, no acquired experience, no exercises of the organs of the body can lead to at least partially non-random changes in heredity.

As can be seen, the Weismann principle is fundamentally different from the Jean-Baptiste Lamarck principle. This principle is formulated as follows. Acquired traits are inherited; the determining factor of heredity is the impact of the external environment.

Indeed, in the XX century it was established that hereditary information was recorded in DNA molecules by a special code. This information transferred in one direction – from DNA to RNA, from RNA to proteins. It is proteins that mainly determine the appearance of the organism (phenotype) and moreover perform many different functions. Characteristically, at first no mechanisms were found to transfer information in the opposite direction, i.e. from proteins to RNA, or from RNA to DNA.

Above we have already said that there are viruses in which hereditary information is written on RNA molecules. These viruses have special enzymes which are able to perform reverse transcription. That is to say, they can copy information from RNA to DNA. DNA created in this way can integrate into the chromosomes of the host cell and multiply along with them. Information transfer from proteins to RNA (reverse translation) has not yet been found in nature. It is possible that such a phenomenon does not exist in nature.

However, Masayuki Nasimoto in 2001 was able to experimentally show the fundamental possibility of backward translation. As a result, the process is the direct opposite of translation. Information is not copied from RNA to protein but conversely from protein to RNA. Nasimoto suggested that using such “reverse tRNA” reverse translation could exist in the RNA world. It can be assumed that there was once a symmetry between direct and inverse processes and there was an equilibrium between the processes of direct and inverse translation. Later, the symmetry was broken, the equilibrium shifted towards the direct translation and proteins lost their ability for coding RNA. It is possible that this was due to the principle of nucleotide complementarity.

Comparatively recently, several other molecular mechanisms have been discovered that in principle can be used to pass on acquired traits. These mechanisms are not directly related to mutations, i.e. with the change of information recorded in the structure of DNA molecules. Therefore, such heredity is called epigenetic.

Thus, the facts show that acquired traits can sometimes be inherited. However, this rarely happens so it is not at all typical. In addition, this happens only with some specific categories of hereditary traits. Moreover, these traits can be considered “acquired” only with a certain degree of conditionality.

Despite the fact that nature has purely “technically provided” the possibility of transferring the acquired properties to descendants, this usually does not occur. Apparently, it's just irrational.

Mitochondrial Eve and Y-chromosomal Adam

At the turn of the XX-XXI centuries, a global study of mitochondrial DNA (mtDNA) and Y-chromosomes of modern humans was conducted. The results showed that all modern humanity comes from a small population that lived in East Africa about 160-200 thousand years ago. To date it has been established that about 60-100 thousand years ago a small group of sapiens came out of Africa, and their descendants gradually settled the whole Earth. Thus, all of modern non-African humanity, partially mixed with Neanderthals and Denisovans, are the descendants of this group of immigrants. The Denisovan hominins are an extinct species or subspecies of archaic humans in the genus Homo which was discovered in 2010 in Altai.

It turned out that if we build an evolutionary-mtDNA tree of modern people based on experimental data and move along its branches from top to bottom (from the present to the past), then all the branches eventually converge at one point in time and space. This is East Africa 160-200 thousand years ago. Thus, the concept of mitochondrial Eve (mtEve) appeared in science (mitochondria are transmitted through the maternal line). After that, in a similar way there was Y-chromosomal Adam (only men have a Y-chromosome and transmit it from father to son). Y-chromosomal Adam and mtEve are the most recent common ancestors (MRCA) from whom all currently living people occur in patrilineal and maternal lines. Eve and Adam lived at about the same time and in the same place.

These results were very ambiguously perceived by ordinary people, various pseudoscientific speculations appeared as few people understood their true meaning. However, there is nothing surprising in the existence of a scientific Adam and Eve.

We formulate the corresponding theorem on Y-chromosomal Adam and mtEve (Khapachev et al. 2018, 24-9).

Theorem (first formulation). With any number of different haplotypes in the original ancestral group of people, after a certain time there will remain the haplotypes of one Y-chromosomal Adam and one mitochondrial Eve.

A haplotype is a set of linked genes on one chromosome of a certain diploid individual which is inherited by all of the descendants on the descending line either through the Y-chromosome from the father for the male descendants, or through the mtDNA from the mother for the female descendants.

Thus, if the theorem is true then any individuals (separately men or women) of the modern population contain DNA fragments having a common origin. Thus, we arrive at the second formulation of the theorem.

Theorem (second formulation). Any homologous (that is having a common origin) DNA fragments somewhere in the past inevitably converge at one point, that is at one ancestral DNA molecule. This wording was given by Markov (Markov 2011), however unfortunately he did not call it a theorem.

This point does not necessarily coincide with the point of appearance of the species. Moreover, if we take different homologous fragments of DNA each of them will give its own “point of convergence”, different from the others. To prove the theorem in the second formulation we need the third axiom by Mednikov formulated above.

We present the proof of the theorem in the second formulation. Suppose there are two representatives of different haplotypes. Let these haplotypes be inherited from the distant past to the present, and ensure that their genealogies do not overlap. The presence of ancestors in each line in the amount of 2^N (N is the number of generations) shows that such a number of individuals simply could not be in the past. This means that somewhere the genealogies (by the haplotype of an individual) once crossed. But more is important to us. The main thing (for the Y-chromosomal Adam) is to cross the pedigree male lineages. Let one of these male ancestors in the pedigree be M, and the other K. If they too, have never intersected anywhere this means that at one or at different times, different individuals in different lines had exactly the same mutations. According to the third axiom that mutations are random, this is almost unbelievable. Consequently, the lines M and K intersect and share a common ancestor. The same reasoning is true for mtEve.

In his book, Markov (2011) illustrates the theorem in the first formulation, that is how it works and how it raises conclusions about the existence of scientific Eve and Adam. In addition, some fantastic situations are considered in which Eve or Adam will never appear.

Mirror neurons

Mirror neurons are probably one of the most important discoveries in neuroscience over recent times. Mirror neurons were detected in monkeys for the first time in the 1990s and later in humans. Mirror neurons are a special group of brain cells that are excited when performing a certain action and when monitoring the execution of this action by another being. Let us give examples of the work of mirror neurons.

Imitation. Mirror neurons provide a “simulation” of other people's actions. Imitation and location to the person often “go hand in hand”. So mutual imitation often occurs when parents and their infant communicate. The mirror-neural system carries out an internal projection into our brain not only of the actions of other people but also their existence as such. People have long used this property of the human psyche without understanding the reason for its existence. Imitation can make us slow or quick, quick-witted or stupid, capable or not capable of mathematics, rude or polite, aggressive or cooperative, and so on. It is important to note that mirror neurons contribute to automatic imitation effects which we sometimes experience unconsciously and which limit our independence. So, for example they force us to submit to social impulses even if we are not aware of it. Fashion, advertising, sales and teaching are namely based on this principle.

Empathy. The activity of mirror neurons explains empathy. Empathy is the ability to put ourselves in the place of another person, to understand emotions and intentions, feelings, and the desires of our interlocutor, making us socially integrated and more successful.

There is a biological explanation for all this which is due to the following circumstance. Our brain is able to mirror the aspects of another's internal state at the level of one cell. Mirror neurons react to beauty; it can be art and science, causing pleasure and triumph. People themselves must make their choice on what and whom to surround themselves with, what to broadcast and what to reflect on.

Nature has made people to be physiologically created for empathy and connected with each other at a basic, pre-reflexive level.

Altruists and Simpson's paradox

There is at first glance a strange statistical effect called the paradox of Edward Simpson. Due to this paradox a situation may arise that intuitively seems impossible. In each individual population, the percentage of carriers of “altruistic genes” is steadily decreasing (altruists always lose out in competition with their selfish kinsmen) but if we look at all the populations in general it turns out that globally the percentage of altruists is growing. This is the principle of action of the paradox.

In Fig.17-1 a hypothetical example of the Simpson paradox action is given. The figure made is based on paper (Chuang, Rivoire, and Leibler 2009, 272-5).

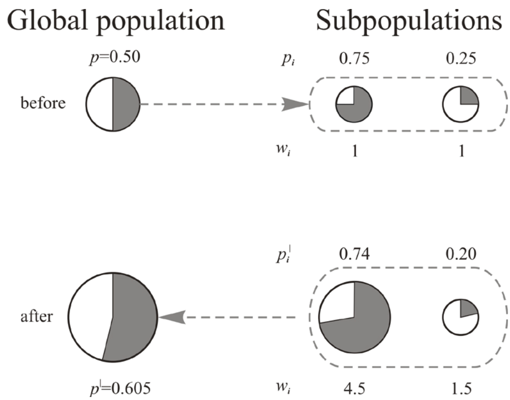


Fig. 17-1.
A hypothetical example of action of the Simpson's paradox

In Fig. 17-1 marked, p_i and p_i' are the proportions of altruists, w_i is the population size, the number of each of the subpopulations at the beginning is normalized to one, and the number of subpopulations is $n = 2$.

In the initial population there were 50% altruists and 50% egoists (circle on the left above). This population was subdivided into two subpopulations with different ratios of altruists and egoists (two small circles in the upper right). During the growth of each of the two subpopulations the altruists were the losers, their percentage decreasing in both cases. However, those subpopulations in which there were initially more altruists grew stronger due to the fact that they had at their disposal more "socially useful product" produced by altruists (two circles at the bottom right). As a result, if we calculate both of the two grown subpopulations, we will see that the "global" percentage of altruists has grown (the large circle at the bottom left). The calculation of the share of altruists is carried out by the formulas:

$$p = \frac{\sum p_i}{n}, \qquad (17.1)$$

$$p' = \frac{\sum w_i p_i'}{\sum w_i}. \qquad (17.2)$$

The main conclusion that can be drawn from the above-mentioned work is as follows. An incentive for the development of intra-group cooperation and altruism is intergroup competition. Analysis of mathematical models showed that altruism could only develop in conjunction with xenophobia, i.e. hostility to strangers. It turns out that such seemingly opposing properties of a person as kindness and militancy developed in a single complex: neither of these features alone would contribute to the reproductive success of their owners. Such is the nature of humanity.

Thus, natural selection operates in parallel on two levels: the individual and the population. Under certain conditions it can ensure the development of altruism even when in each individual population it favors the egoists, and altruists are condemned to extinction.

Human physical fields

Currently, no one is surprised at the ability of animals to anticipate earthquakes, to find water in arid places or roads in their habitats, to heal themselves and each other, and many other qualities not common to most people. However, until now the existence of such abilities in some people called sensitives or extrasensory seems either empty fantasies or something supernatural. Meanwhile, numerous scientific studies have shown that any biological object in the process of vital activity generates a complex picture of physical fields and radiations, and also transmits and perceives the information embedded in these fields and radiations.

The picture of these fields reflects the work of the physiological systems of the body ensuring its homeostasis. A person is given the opportunity to see other people in the reflected light of a narrow spectral range of frequencies of an electromagnetic field. Visualization of the physical fields and radiation of biological objects not only significantly expands our “vision” but also provides an opportunity to look deep into the body and brain to observe the dynamics of physiological processes in their own “light”. For a person there are six main types of fields and radiations. We will give their characteristics according to an article by Godik and Gulyayev (1990, 75-83).

1. Infrared thermal radiation characterizes skin temperature determined by capillary blood flow. In humans this radiation is strongest in the middle infrared wavelength range (3–14 μm) where its intensity is in the order of 10 mW/cm^2 which is more than 100 W from the entire surface of the body. The characteristic depth of absorption of this radiation in biological tissues is about 100 μm.

2. Radio thermal radiation carries information about the dynamics of thermal fields of internal organs and the brain. This radiation is very weak, its intensity in the decimeter wavelength range is of the order of 10^{-12}W/Hz cm^2. Unlike infrared thermal radiation the characteristic depth of absorption of this radiation in biological tissues is in the order of several centimeters which is why it carries information from the depths of the body.

3. Acoustic thermal radiation in the ultrasonic range of wavelengths characterizes the temperature distribution inside the body with a higher spatial resolution than radio thermal radiation since the ultrasonic wavelength is much smaller than the wavelength of the electromagnetic field emitted from the same depth. In addition, low-frequency acoustic signals carry information about the physiological mechanics of internal organs (heart, lungs, etc.). In the frequency range 1–10 MHz biological tissues are sufficiently transparent for acoustic waves. Its intensity is very small and amounts to about 10^{-16} W/cm^2 in the 100 kHz frequency band. The wavelength in this range is of the order of 0.1 cm, which is much less than the wavelength of thermal radiation emitted from the same depth.

4. Optical chemiluminescence associated primarily with lipid peroxidation provides information on the saturation of tissues with oxygen, the antioxidant status of the organism, etc.

5. The electric field reflects the bioelectric activity of the brain, heart, muscles and other internal organs. In addition, the electric fields around a person are associated with triboelectric charge (charge resulting from friction) on the stratum corneum of the epidermis, which has high ohmic resistance (10^9–10^{11} Ohm/cm^2) and therefore reflects the physiological "seismicity" of the torso.

6. The magnetic field created by bioelectric sources is very weak, about 10–100 times smaller than the geomagnetic field. The magnetic field more directly than the electric field reflects the distribution of the bioelectric activity of the brain and internal organs, since it is practically not shielded by the diamagnetic tissues of the body. In other words, the bioelectric "landscape" inside the body through the "magnetic window" is visible without distortion as through transparent glass, while through the "electric window" it is visible with distortion as through a stained-glass window.

7. A chemical "micro-atmosphere" should be added to the six physical fields listed above, formed by exhaled gases, fumes through the skin in the process of imperceptible perspiration, etc.

Thus, in each of the named fields the physiological information is contained in the space-time distribution of signals, i.e. in their dynamic images, and can be used in the early stages of the diagnosis of various diseases.

It has been established that in the animal world receptors of electromagnetic, acoustic and other radiation exist and are effectively used, having a close connection with both the cerebral cortex and its subcortical structures. For the perception of information of this kind it is only necessary to create conditions for coordinated control of the amplitude, phase and frequency of those radiations induced by nerve cells and protein molecules.

If such conditions arise then in principle, there are no prohibitions for directional radiation that carries energy over long distances without significant attenuation. Similarly, by causing the necessary phase shifts of the oscillations when they are perceived by the relevant structures of the brain, and adding them we can amplify and highlight weak signals from a certain source to the receiving system. This, in fact is a well-known radio telescope electric-scanning principle which finds its application in the work of the central nervous system.

Receiving information about the surrounding reality through sensory channels, a person often remains at the level of unconscious perceptions. Each of us has been in a situation where, without realizing what information we have received we say: "I do not know why, but I feel that it is so".

Researchers believe that some people are able to perceive local distortions of various physical fields that are well known to science:

electromagnetic, electrical, magnetic, radiation, and acoustic. If we add to this that well-known organs of sense can perceive much from the outside world and thereby "generate" information about it in the brain, then the source of information on the basis of which the manifestation of certain reactions of a person takes place, loses its mystique.

Beauty and the brain

A joint study of neuroscientists and mathematicians conducted in recent years has shown that the perception of "beautiful" mathematical formulas affects the same brain region as the perception of painting and music. These works became one of the first attempts to understand the concept of mathematical beauty with the help of a rigorous scientific method.

Bertrand Russell (Russell 2008) said:

> Mathematics, rightly viewed, possesses not only truth, but supreme beauty – a beauty cold and austere, like that of sculpture, without appeal to any part of our weaker nature, without the gorgeous trappings of painting or music, yet sublimely pure, and capable of a stern perfection such as only the greatest art can show. The true spirit of delight, the exaltation, the sense of being more than Man, which is the touchstone of the highest excellence, is to be found in mathematics as surely as poetry.

Earlier, at the very beginning we made a bold statement that the "beauty" of mathematics is absolute unlike the beauty of painting or literature. And how absolute is it for mathematicians themselves? Is it possible to compare the concept of beauty in mathematics with beauty in poetry, music, and visual arts?

In 2014 an article of a group of authors was published (Zeki et al. 2014). Thus, specialists in very far apart scientific fields united to try to explore the mathematical "beauty" with a rigorous scientific approach.

Earlier in 2004 Semir Zeki observed with the help of functional magnetic resonance imaging (fMRI) a certain correlation between the perception of beauty in the visual arts and the work of the brain section called the medial orbito-frontal cortex (mOFC). The mOFC is the part of the cerebral cortex that is located roughly behind the eyes. Further, in 2011 Zeki studied the perception of music on the mOFC.

Finally, in 2013 Zeki and Michael Atiyah formulated the natural question: is the perception of mathematical "beauty" related to the same brain region?

Mathematicians were shown 60 formulas, among them were famous theorems, and fundamental identities and definitions. At first, the subjects rated each of them on a scale from –5 (ugly) to +5 (beautiful). After 2-3 weeks they were again shown all 60 formulas but in a different order. As before, the reactions of the subjects were observed on fMRI. Unlike experiments with the perception of art (this is important!) subjects were asked to evaluate the clarity of each formula. This was done in order to cope with the trap, which can be formulated as follows. You cannot judge the beauty of what is not clear.

The results of the subjective assessment of the beauty of formulas from the list turned out to be more curious than significant from a scientific point of view. The most "beautiful" (score 3.375) was named as Euler's formula:

$$\exp(i\pi) + 1 = 0 \quad . \tag{17.3}$$

The most "ugly" (score 1,687) was the Srinivasa Ramanujan formula:

$$\frac{1}{\pi} = \frac{2\sqrt{2}}{9801} \sum_{k=0}^{\infty} \frac{(4k)!(1103 + 26390k)}{(k!)^4 396^{4k}} \quad . \tag{17.4}$$

Perhaps formula (17.4) turned out to be the most "ugly" because there was no more cumbersome and asymmetric expression in the list. However, this seemingly "ugly" but extremely interesting formula is very useful for engineers. With it you can quickly get approximate values of π. Even the first term of the series (for $k = 1$) gives π with an accuracy of six decimal places.

Abstract beauty can be found in the formula of Ramanujan. It connects the fundamental constant π describing the properties of a circle with completely incomprehensible numbers: 9801, 1103, 396 and 26390. It is not clear why they can calculate π. So is it "ugly"? But what about Hieronymus Bosch? See his triptychs, for example, "The Garden of Earthly Delights" in the Prado Museum in Madrid.

Thus, the study of Zeki and Atiyah, based on a rather limited experiment proves only that the perception of mathematical beauty in one of its aspects is somewhat similar to the perception of some facets of artistic or musical beauty.

CHAPTER 18

PHYSIOLOGICAL BASES OF VITAL ACTIVITY. LIVING INFORMATION SYSTEMS

Multi multa scinut, nemo – omnia.

According to modern concepts any living organism is a self-organizing nonlinear open metabolic information system with developed feedback obeying the laws of control, on the basis of which it is able to maintain a non-equilibrium homeostatic state and perform purposeful actions to meet its many needs. This definition applies to subcellular, unicellular and highly organized multicellular organisms including humans but their main characteristic is the presence of signs of mental activity associated with the emergence of the nervous system and the development of the cerebral cortex.

Reflections on the nature of mental activity have continued probably since the activity began. But it was possible to reach agreement on its source relatively recently after the achievements of the natural sciences of the XX century. At this time psychology went beyond descriptive human science. It began to play an active role in defining methods and techniques for influencing a person, his natural strengths and potencies so it immediately joined the sphere of neuroscience which studies a complex system of relationships: nature – man – society. This system of relationships is manifested in the vital activity of individual subjects, various social groups and associations as well as in the activities of *homo sapiens* as a whole.

The progress of the natural science and technical disciplines contributed to the modern concept of mental activity. It allowed the consideration of the life activity with the involvement of the conceptual apparatus of system analysis. From this point of view, a living organism forms a unity with the external environment being an element of a higher order system – the biosphere (or the noosphere according to Vladimir Vernadsky).

At the same time, the organism itself is a complex system of interconnected and interacting elements or SFWUs which provide it with growth, development and full-fledged activity. Thus, the organism is an

SFWU at the level of the biosphere; organ systems (cardiovascular, respiratory, digestive, excretory, endocrine, sexual) and an SFWU at the organism level; organs are the SFWUs of these systems and so on down to the molecular and submolecular levels.

The integrity and functional unity of the organism as well as its interaction with the external environment, are provided by the nervous system. The nervous system controls the work of all other SFWUs of the body combining and coordinating their activities. In addition, with the help of sensory organs and special sensory endings situated in the internal organs the nervous system constantly perceives irritations from the external world and ensures one or another activity of the organism carrying out its interconnection with systems of a higher order – nature and society. Thus, the nervous system is the management system – the regulator, and with the help of this system the organism as an integral SFWU enters into unity with the external environment in which it grows and develops. This is why the branch of science related to the relationship between the organism and its environment became known as neuroscience.

Today the central position of neuroscience is the following axiom: all conscious and unconscious mental processes, in principle, are consequences of physiological processes in the nervous system and can be considered as its function.

Neuroscience includes all areas of knowledge related to the nervous system and aimed at studying the nervous processes that ensure the interaction of the organism with the external environment. Various aspects of this interaction are investigated at the molecular (neurochemistry and biophysics), cellular, tissue, organ and organism (neuromorphology) levels of structural organization. Studying the ability of a structure to perform certain actions, i.e. the study of its functional properties, is within the competence of neurophysiology and physiology of higher nervous activity (HNA) including mental activity.

Becoming an integral part of neuroscience and engaging in that area that studies the behavioral responses of higher animals and humans, arising in response to various factors of the internal and external environment, psychology adopted another neuroscience position that the source of mental activity lies in the informational reflection on the basis of the nervous system, i.e. in the property of structural interaction that arose during the transition of matter from inanimate – insensitive to sensitive – to alive.

Currently the concept of "psyche" is interpreted very broadly. It includes such processes as thinking, consciousness, and memory, as well as emotions and intelligence – forms of HNA that condition and modify behavior.

Behavior is a form of vital activity that determines the likelihood and duration of an organism’s contact with environmental objects that can satisfy its innate and acquired needs.

Informational reflection presupposes the existence of an SFWU that cannot just experience external influences and change its state accordingly but also builds special relationships with the external environment carrying out behavioral reactions, i.e. performing special functions. Thus, any information system including a living organism is a functional system with a specific structure and ability to perform various actions, changing its structure. At the same time, the external stimulus affects the state of the system not directly but indirectly: it brings the system into an active state and “includes” internal action programs which are then carried out independently of the properties of the stimulus itself and constitute a particular function.

In the process of the transition of matter from insensitive to sensitive and further to thinking matter live information systems emerged which evolved from simple information systems – single-cell organisms with elementary irritability to complex information systems – multicellular organisms with reflexes based on the nervous system, and then higher animals and humans with mental reflection. Neurophysiology which deals with complex information systems borrowing the term “reflex” from physics has endowed it with new content. As applied to a living organism a reflex means reflection on the basis of the nervous system or neurophysiological reflection. It is associated with the excitability of nerve cells and the generation of nerve impulses.

So, the initial form of informational reflection, characteristic of all living things, is irritability – the body's ability to respond to specific reactions in response to a specific stimulus. These reactions are already manifested at the cellular level and occur due to the energy of the cell itself, and the energy of the external stimulus serves only to “launch” the internal process. This property shows a feature of an information system in which the intrinsic characteristics of the stimulus are not decisive for a particular function. The external stimulus stimulates, excites the internal program of self-movement, leading to the active state of the functional system that organizes the behavior, regardless of the properties of the stimulus itself. Often insignificant in their own characteristics – energy or material – the impact can have a huge informational and signal value for the reflecting system.

With the emergence of multicellular organisms and the formation of nervous tissue, and then the nervous system such a form of informational reflection appeared as a reflex, and finally, the next step in the evolution of living information systems is associated with the development of the cerebral cortex which is the substrate of mental reflection. In humans the process of perception of the external world seems to repeat the history of the development of the psyche from elementary irritability to conceptual thinking and consciousness associated with the activity of the cerebral cortex. At the same time, between the reactions of the cell and the purposeful behavior of a person lies a huge layer of phenomena resulting from the interaction of specialized structural formations of the nervous system, organizing behavior and united according to the principle of hierarchy. The hierarchy is the "service ladder", i.e. the location of the SFWU in the order of submission and transition from the lowest level (or rank) to the highest. This principle is manifested both in the unification of any SFWU of a separate organism, and in the unification of social animals including people into social groups, various communities, and, finally, in such an association which we call "humanity".

On our planet humanity is the most developed living information system that cannot only learn about the world, master it and defend itself against its damaging factors but also change objects and environmental phenomena at its discretion in order to fully satisfy all its needs.

The need-informational approach to the activity of the nervous system organizing and regulating behavior is based on a large number of scientific facts obtained in studying the structure and functions of the nervous system of animals and humans, in studying their behavioral responses, and also in analyzing the results of the activity of the healthy and diseased human brain. To some extent this approach made it possible to find out the ways in which the nervous system coordinates the needs of the organism with the conditions of the internal and external environment and organizes behavioral acts aimed at achieving a specific goal.

The nervous system is a regulatory system that controls the work of all other systems of the body carrying out the processes of its life support and behavior in the external environment.

The activity of the nervous system aimed at combining and coordinating the work of individual parts of the body and the regulation of life support processes is considered to be the lowest nervous activity. All structures of the nervous system including the cerebral cortex take part in this activity.

The activity of the nervous system aimed at the interaction of the organism with the external environment, i.e. on the organization of behavior, is called the higher nervous activity. Mental activity is an integral part of the HNA. The HNA is so-called because of its complexity, and not because it is connected with the higher parts of the central neural system

(CNS) since both the higher cortical structures of the brain and the upper and lower sections of the trunk, the spinal cord and peripheral formations of the nervous system take part in the organization of behavior.

At the same time the behavior due to the activity of the nervous system is a continuum of results. Continuum in this context means a variety of results achieved during life. The continuum of results or the behavioral continuum is a continuous sequence of behavioral acts performed by the body throughout life. A separate behavioral act is a segment of the behavioral continuum from one result to another.

In organizing behavior, action initiators are multiple needs, and the ultimate goal of any behavioral act is to satisfy them.

The presence of needs determines the constant activity of various functional systems and, in the first place, the nervous system, and the satisfaction of needs reduces this activity. In other words, the activity of a particular neural ensemble (block) naturally increases with the implementation of behavior aimed at achieving a certain goal (final result), and decreases when the goal is reached, i.e. the need is met.

The concepts of activity and expediency are associated with the ideas about anticipatory reflection, about the future activity of individual neurons, neural ensembles, all components of the management nervous system and the organism as a whole.

In a developed management system, there are components responsible for obtaining information from the environment (external and internal); components responsible for planning actions and forecasting results; components of command systems that motivate certain executive bodies or systems as well as structures related to the action acceptor, i.e. receiving information about real results and compliance with their expected results (controllers). If the result is unsatisfactory the action acceptor sends impulses to the structures responsible for planning actions and creating specific programs for new actions.

Reprogramming thus occurs and the achievement of the desired result.

Fig. 18-1 presents with some changes the scheme of the operational architectonics of the central nervous system according to Pavel Simonov (Simonov 1986).

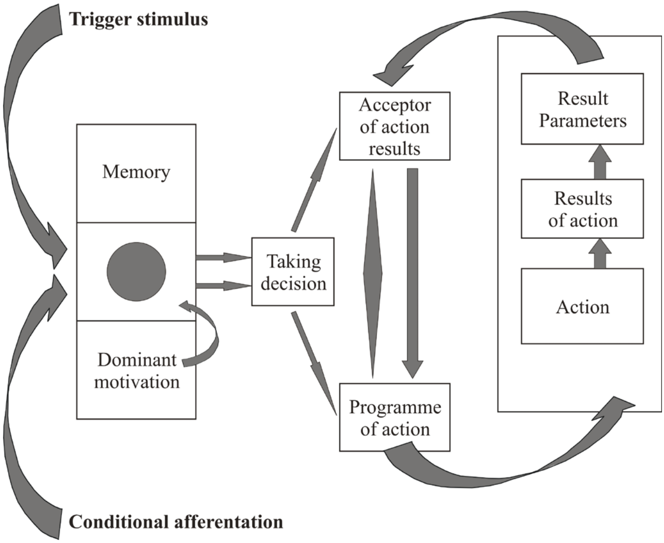


Fig. 18-1.
The scheme of operational architectonics of the CNS

A behavioral act occurs in the following cases.

1. In the presence of a trigger stimulus and in an appropriate setting (reinforcing afferentation), i.e. the first stage of the behavioral act is the actualization of the need (the emergence of the motive and the presence of the appropriate situation). The motive is the actualized need that arises when the memory of the targets is activated. If the need has not become a motive, then there is no incentive for action.

2. Upon activation of the neural mechanisms necessary for making a decision on the action, i.e. when motivation occurs; at the second stage, the corresponding material is extracted from the memory or the search behavior is performed. Motivation is a physiological mechanism for activating traces (engrams) stored in memory of those objects and actions that are able to satisfy an existing need; if there is no necessary information in memory then mechanisms of search behavior are activated according to the principle of "trial and error".

3. When activating the structures responsible for assessing the likelihood of the satisfaction of needs (emotional background) for drawing up action programs and in the presence of mechanisms for implementing actions. That is, the third stage includes making decisions about actions, drawing up action programs and realization actions.

If any of these components which are necessary at each stage of the organization of behavior is absent then the behavioral act is not realized. For example, according to expert evaluation of the emotional sphere if the probability of satisfying a need is so small that a negative emotion appears, programs are not compiled, and commands for action do not arise.

Thus, motive is necessary for the implementation of the first stage of the behavioral act. Motivation is needed for the second stage. At the third stage actions are programmed and implemented.

The ability to perceive changes in the external and internal environment and the ability to change their behavior to achieve a specific goal are two basic properties of all living information systems (living organisms) from the simplest to the most complex. However, information systems that have a nervous system and especially a brain, in their abilities to sense and act are far superior to other, simpler organisms. This is due to the fact that with the development of the central nervous system in the purposeful behavior of living information systems such options of action appeared that are characterized as "modeling", and among the control mechanisms, along with the "programs" and "blocks" there appeared an "outposts" management which is based on the forecast development of the situation in the external environment.

The simplest mechanisms by which all control laws are implemented are based on two basic principles: management "on perturbation" and management "on deviation".

The selectivity of informational reflection is manifested by management "on perturbation" when the system can accept, reject or change (strengthen or weaken) the signal from the external environment "at the input" and give it the value of the incentive. At the same time protective reactions are realized. In our context the signal is a nerve impulse and the incentive (stimulus) is the motive for action, the cause of action. It is important that all information flows that may have a signal value, i.e. perceived as a signal in a particular situation, exist as irritants.

Irritants are those changes of the external environment that exist objectively and do not depend on our perception. Not every irritant is a signal. There are different irritants that are not perceived by the system and, accordingly, do not become signals. Irritants that carry information necessary for both orientation in the external environment and assessing the state of the organism itself acquire a signal value if they cause the generation of nerve impulses in the receptor cells of sensory systems.

From this it follows that the irritants themselves are not yet information; they acquire this quality only after they become signals – semantically significant information elements for those information systems that perceive them.

The signal turns into a stimulus when it becomes a motivation to action, i.e. in those cases when one or another action occurs after the perception of the signal. Not every signal becomes a stimulus, as not every irritant is a signal. Thus, a chain is built: an irritant – a signal – a stimulus where the broadest concept is "irritant", and the narrowest is "stimulus".

When management on deviation uses biological feedbacks (BFB) outrunning reflection is realized. The management system detects changes in the parameters of the actions of other functional systems "at the output". If such parameters differ from "given", BFB is turned on, the deviations are corrected and the parameters return to the necessary ones. With the participation of positive and negative BFB purposeful independent behavior is formed including foreknowledge.

The combined effect of positive and negative BFB allows you to organize adaptive behavior including the processes of self-organization, for example compensatory-adaptive reactions (CAR).

Fig. 18-2 and Fig. 18-3 provide with some changes the schematic representation of management methods according to Eugene Popechitelev (Popechitelev 1997).

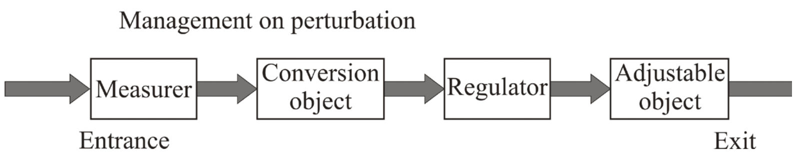


Fig. 18-2.
Management on perturbation

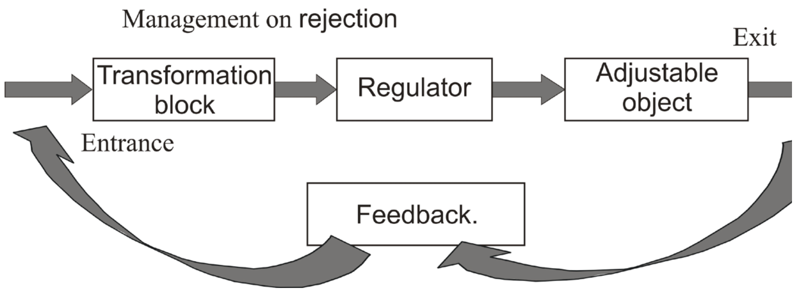


Fig. 18-3.
Management on deviation

Other management mechanisms in the body that evolve during evolution include the following.

1. “Thresholds schemes” i.e. the connection of command systems to executive bodies only if there is a certain threshold level of the “input” signal. Subthreshold or ultraboundary levels can be perceived by sensor systems but they do not lead to action because of deviation by command systems. To realize a management action with the help of a “threshold scheme” it is necessary for the signal to acquire a threshold value and become a stimulus.

2. “Program management” is a specific program of action. It depends on the “output” signal which has the value of a stimulus and includes a well-defined program of action.

3. “Block management” in which the action is realized by connecting standard programs into single blocks (for example when walking or running, in the limbs the programs of the action of flexor muscles and extensor muscles are simultaneously included).

4. “Outpost management” which is based on predicting the results of the action and the development of the situation in the external environment, i.e. the management of future-oriented behavior. For example, when a cat is hunting a bird it jumps not to where the bird is at the moment but to where the bird will be when it takes off. Thus, the cat acts with foreknowledge of the future.

The behavioral actions implemented by the managed system are divided into three groups.

1. “Reactions” where each action is considered as a simple mechanism of a behavioral act which is included in any situation in the transition of a certain signal irritant in the starting stimulus. Such mechanisms include irritability and simple reflexes carried out on the principle of “start – brake” or “signal – response”.

2. “Stereotypes” in which the behavior is based on certain programs, when each program was repeatedly worked out by the system and became standard for it (typical). That is, the organism responds to the action of an external stimulus with a standard set of behavioral acts (automated behavior). This is how complex unconditioned and conditioned reflexes (CR) are performed.

3. “Modeling” in which each behavioral act takes into account the changing parameters of the external environment, the current state of the body, the objectives pursued, and the predicted results. That is, a model of the “required future” is drawn up and a correspondence is established between existing and possible situations and behavior.

In this option of action, a comprehensive analysis of information from the external and internal environment is required to recognize a specific situation.

In this connection, signals coming from the sense organs take on special significance. These signals that occur in the receptor neurons of the sense organs and are transmitted via afferent fibers to the nerve centers are called sensory signals. The process of transmitting information along the afferent pathways of the nervous system is accompanied by their multiple transformation (transcoding) at all hierarchical levels and ends with the recognition of the sensory image. It is clear that the identification of the sensory image includes the collection of information, and its analysis and synthesis. For this, various nerve elements are combined into functional systems, which have their own sensors, communication channels, signs, detectors, etc. All these elements are necessary for semantic information to be involved in the processes of higher nervous activity.

The sensory system is a combination of peripheral and central formations that provide coding (reflection) of the physicochemical characteristics of external irritants, an assessment of their signal nature and biological significance. These formations are involved in the organization of individual behavioral acts and the entire adaptive, purposeful behavior. In addition to sensory information, the organization of behavior involves information stored in memory. To assess the incoming information, the so-called "expert systems" are included in the structure of the behavioral act which are attention and emotion.

In Fig. 18-4 the overall structure of the behavioral act is given with some changes according to (Simonov 1986).

The motive, or the motive cause of action is always this or that actualized need. In animals, it is a vital or social need; in humans, added to this is a large spectrum of ideal needs.

On the basis of needs there are instincts (complex and most complex unconditioned reflexes (UCRs)), drives and CRs as well as emotions that arise with needs. In addition, people add attitudes and ideals that are acquired in ontogeny based on thinking and consciousness. With the help of emotions, dominant and subdominant needs are highlighted if at some point in time several competing needs are actualized.

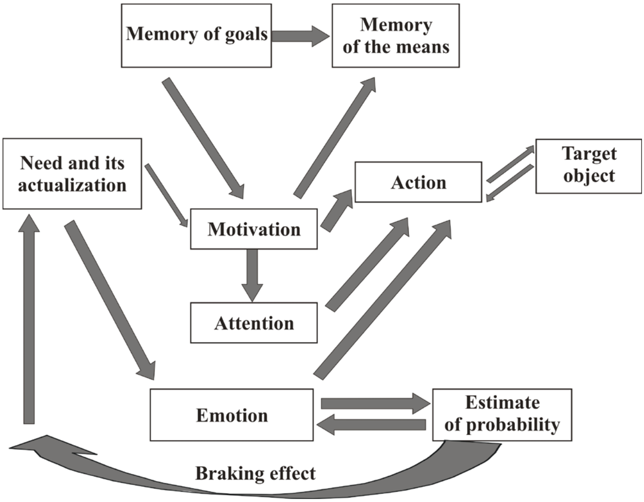


Fig. 18-4.
The overall structure of the behavioral act

When they talk about competition of needs or about "the struggle of motives" they actually mean those emotions that are caused by the needs that are simultaneously actualized at a given moment in a particular person.

If motive is the cause of the action, then the mechanism that drives the action is motivation. Thus, motivation is the activation of the activities of brain structures involved in the realization of behavioral acts aimed at meeting the needs and, accordingly, at eliminating or minimizing negative emotions and obtaining or maximizing positive emotions.

From a neurophysiological point of view, motivation is the mechanism of activation in the memory of engrams (memory traces) of those external objects that are able to satisfy the need of the organism and the actions that are necessary for this to occur. Thus, the emergence of motivation is closely related to the selective excitation of the nerve centers involved in the organization of targeted search behavior.

There is no motivation without needs, but it is quite possible to meet a need that has not become motivation. For example, a person has an acute need for vitamins which causes him a state of discomfort, i.e. negative emotions that are a reflection of unmet needs. Without knowing the reasons for his discomfort, the person is not motivated. It is necessary to go to the doctor and find out the motive then there will be motivation, the memory of the means is activated, and the person starts to act.

Chapter 19

Metabolism and homeostasis

Est in media verum.

The basis of the vital activity of all living beings is their ability to metabolize. Life requires energy consumption, and a variety of chemical reactions that take place in cells and intercellular space serve as a source of energy in the body. All chemical activity that ensures the mobility of cells and their endings, excitability, conductivity and growth, restoration and reproduction, is called metabolism. The energy required for life processes is produced in the body with the continuous breakdown of proteins, fats and carbohydrates which has a long series of intermediate steps. If this splitting occurs in the presence of oxygen then the end products are carbon dioxide and water, and this process is called respiration or biological oxidation. At the same time chemical energy is bound in a biologically useful form – in the form of energy of organic compounds, such as for example, adenosine triphosphate.

The energy obtained is again included in the process of synthesis of organic substances, that is, the same carbohydrates, proteins, fats and other compounds necessary for life, and is also used to process nutrients coming from outside. A nutrient is any substance consumed by the body that can be used to obtain the energy needed to consume, and repair body tissues, or to support physiological processes. After entering the body, nutrients are either included in the composition of new elements or oxidized, delivering energy to the body. Some of this energy is used to synthesize new substances, some is consumed in the process of the functioning of organs and cells (transmission of nerve impulses, contraction of muscle cells, etc.), and some is released as heat.

Nutrition or trophicity is a complex multi-stage manifestation of the activity of the organism. It is composed of the processes of the search and absorption of food; digestion; the absorption of nutrients and their entry into the internal environment of the body; balancing the assimilation reactions of arrived simple molecules with the processes of dissimilation (disintegration) of complex molecules that make up the internal environment of the cell; and the timely removal of intermediate and final products of exchange. The combination of these successive or simultaneously occurring processes is aimed at growth, and maturation at preserving the structure and function of cells, tissues and organs, and consequently of the whole organism. Disruption of the relationship between these processes can lead to the degradation and death of the whole organism.

Depending on the nutrient supply, organs, tissues and cells may experience various trophic conditions, which are named in accordance with generally accepted terminology, in the list below.

Eutrophy is optimal nutrition, that is, a relationship between the level of utilization of nutrients flowing to the cells and the rate of removal of decay products, in which there are no deviations from the normal (morphological) structure, the physicochemical properties and the functions of cells and normal growth ability, development and specialization.

Hypertrophy is an enhanced nutrition. It is expressed as an increase in the mass of cells (true hypertrophy) or in an increase in the number of cells (hyperplasia) usually with an increase in their function. Hypotrophy is a weakened diet associated with a decrease in mass and number (hypoplasia) of cells such as physiological hypotrophy of various organs with low mobility, a very common state of the body at present. Atrophy is the lack of nutrition and, accordingly a gradual decrease in the mass of cells and their disappearance. Dystrophy is a qualitatively altered malnutrition that leads to morphological shifts and changes in the physicochemical properties and functions of cells, tissues and organs, their growth, development and specialization. There are local, systemic and general dystrophies, congenital and acquired as a result of the damaging effects on the body of external and internal factors. Dystrophic changes can be reversible if the harmful factors cease to act, and irreversible, resulting in cell death if the dystrophy from the very beginning was incompatible with their life. In many cases severe damage or the death of organs and tissues is the result of dystrophic changes as a result of chronic effects on the body of certain pathological factors. Trophic changes can develop as independent phenomena or as symptoms of various diseases.

With the development of a number of standard and specific physiological processes (inflammation, tumor) the phenomena of hypertrophy, hyperplasia, hypotrophy, hypoplasia, and atrophy can be observed simultaneously. Often these changes in the trophic state follow each other. The founder of neuroscience, the Nobel laureate Santiago Ramón y Cajal was the first to point out the following circumstance. Often one can observe severe competition for nutrition among cells in the tissue, a “struggle for existence” among cells in inflammatory areas or among elements exposed to tumors.

In a civilized community the struggle for existence is significantly reduced through public administration aimed at the division of labor and the distribution of benefits that create conditions for common interests. Similarly, in a living organism thanks to special control of the nervous system as well as the distribution of functional roles the struggle disappears or becomes moderate, manifesting itself only when the total nutrition of organs or cells is seriously disturbed by internal or external causes.

Still Hippocrates noticed the connection between changes in individual organs and parts of the body. He believed that the organs "sympathize" with each other in relation to their nutrition. Jakob Winslow developed this idea and first suggested that the mutual influence ("empathy–sympathy") of internal organs against each other in which the disease of one of them involves other organs in the painful process is carried out by a "sympathetic" nerve. Then, for more than 200 years of history the doctrine of "nervous trophism" developed. The main conclusion of this doctrine is that the regulation of trophism in the body is carried out by a reflex path ("reflex" – reflection) through the nervous system and one of the most important elements of the trophic reflex is its department traditionally called sympathetic.

The trophic reflex as well as any reflex consists of three links: the afferent ("bringing") part, the nerve center, and the efferent ("diverting") part. But in this case, it is not about a simple reflex arc but about a branched reflex which receives multiple structural and functional support in all its links and is carried out simultaneously in various afferent, central and efferent ways. Naturally, in any event affecting one or several links of the trophic reflex almost all morphological structures and physiological mechanisms related to the trophism of cells, tissues, organs and the body as a whole are involved in the process.

Metabolism is regulated in such a way that the internal environment of the body optimally matches the external conditions but the level of metabolic processes has a certain relatively constant range corresponding to adaptive reactions. External conditions tend to cause changes, and the body continually adapts to this ensuring the reaction of the internal environment within a constant range. Such a desire for relative constancy of the internal environment is called homeostasis.

The concept of homeostasis was originally formulated by Walter Cannon for the theoretical justification of the constancy of the internal environment of the body and a number of its physiological functions. Then the term "homeostasis" was understood as the constancy of the composition of blood, lymph, and tissue fluids which consists in maintaining the osmotic pressure at a certain level, the total concentration of metabolites, individual ions, and the acid-base state. Later the concept of homeostasis was mathematically developed by Wiener from the standpoint of an open system as a stationary state that is provided by regulatory mechanisms with feedback.

The most common definition of homeostasis in the literature is the following. Homeostasis is the relative constancy of the composition and properties of the internal environment and the stability of the main physiological functions of the body, due to complex regulatory interactions at the molecular, cellular and organismic levels of organization of the living. This definition belongs to Gregory Kassil, Yuri Veltischev, Boris Tarusov, and Vladimir Frolkis.

In this case the concept of homeostasis already includes both the phase and cyclical courses of the reactions. Such a formulation of course is a development of the initial ideas, when the constancy of the internal environment was put in the foreground, and any change was considered as a violation of homeostasis. Nevertheless, at present the term homeostasis is used even more widely; however, this does not mean the invariance of the state but the indispensable fluctuation around a certain level.

Progress in the fields of neurobiology, clinical and experimental medicine and psychology is accompanied by the emergence of new concepts, the renewal of traditional ideas, and the use of terms of related disciplines with a new shade of content. High rates of development and continuous enrichment of ideas about the nature and mechanisms of physiological phenomena make the process of “semantic drift” of concepts especially tangible.

One of the independent new directions in the non-drug improvement of normal (healthy) functions and the correction of impaired functions based on the targeted activation of the body's reserve capabilities is the use of BFB. In a general sense, feedback is the effect of the results of the functioning of a system on the nature of this functioning which is schematically depicted in Fig. 19-1. Feedback which leads to an increase in the output signal at a constant signal at the input is called positive, while that causing a decrease is called negative. In general, the feedback gives the regulators signals about the result of the management action.

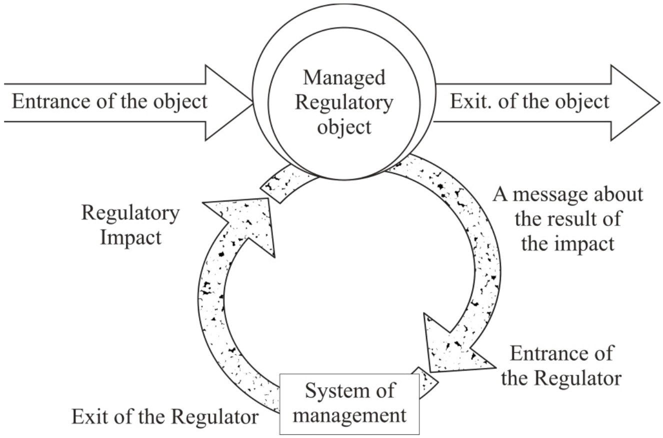


Fig. 19-1.
The generalized scheme of the device with feedback

According to modern concepts the organism is a complex dynamic system consisting of a complex of interconnected and interacting elements that become more complex and rise along the hierarchical ladder. At the same time, it is an element of a higher order system (biosphere) and forms a unity with the external environment. At each hierarchical level direct and feedback links are made between its constituent elements and the SFWU. The organism is an SFWU at the level of the biosphere; specific and non-specific functional systems (cardiovascular, respiratory, digestive, nervous, endocrine) are SFWUs at the level of the organism; organs are SFWUs of functional systems, and so on down to the molecular and submolecular levels.

The level of fluctuations of homeostatic reactions (the width of the range) and the reserves for its maintenance are determined by the activity of an SFWU which exists due to the energy produced in the process of metabolism. The degree of activity of an SFWU which provides a dynamic balancing of the intake and expenditure in the body of flows of substance and energy depends on the intensity and direction of the metabolism program. The optimal state of homeostasis is provided by evolutionary hereditarily programmed mechanisms of the constitutive and inducible metabolism.

Constitutive metabolism is aimed at the production of substances that form the physical basis of an SFWU, and participate in the reactions necessary for its life support. It maintains homeostasis in isolation from any external influences for a given SFWU due to its internal regulatory mechanisms. Inducible metabolism is directed to the synthesis of substances that provide the various functions of an SFWU which are necessary for performing specialized work in the system as well as for carrying out the full-fledged vital activity of the body in accordance with environmental conditions, that is for implementing adaptive and compensatory-adaptive reactions.

Inducible metabolism maintains an adequate homeostasis in the presence of external influences on the organism as a whole through the interaction of internal and external regulatory mechanisms. The inclusion of inducible metabolism leads to the launch of biofeedback acting on constitutive metabolism in order to stimulate (positive communication) or block (negative communication) the synthesis of substances used for life support. Under certain conditions substances synthesized under the control of external regulatory systems can act only according to the principle of negative feedback and block the metabolism aimed at supporting a particular SFWU if it is necessary to optimize the work of a functional system or organism as a whole. A generalized scheme of homeostasis in a healthy and a diseased body exposed to an extreme irritant or a pathogenic agent is presented in Fig. 19-2.

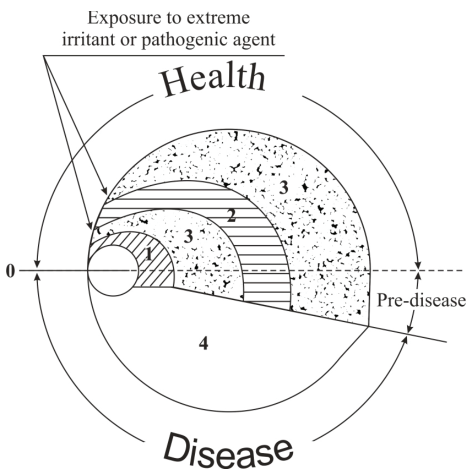


Fig. 19-2.
Exposure to an extreme irritant or a pathogenic agent:
1 – basic homeostasis (constitutive metabolism);
2 – adaptive and homeostatic reactions;
3 – compensatory adaptive homeostatic reactions;
4 – pathological reactions

For the organism as a whole as for an open system, external regulatory stimuli are environmental factors. For functional systems of the body (organs, tissues, and cells), external regulatory stimuli consist of the results of the interaction of nerve, neuroendocrine, endocrine, immune and other influences, each of which represents a specific control loop. It is important that any circuit is self-regulating and somewhat autonomous but at the same time for each of them the other homeostatic circuit is also a regulator providing an adequate output signal.

The highest integrating and regulating role in providing reactions that maintain homeostasis at the organically-systemic level is played by the nervous system. Regulation is based on a complex hierarchy of mutually subordinate "substations" in the central and peripheral nervous system, ranging from cortical-subcortical brain structures to axon terminals operating in each working SFWU up to cell groups or even individual cells. When one of these substations is damaged especially the brain structures, a cascade of compensatory-adaptive reactions occurs entailing a complex of morphofunctional rearrangements in the brain itself, the rest of the nervous system, and then in organs and tissues of other systems' life support.

According to modern concepts in the nervous regulation of homeostasis the most significant reactions are not so much the nerve-impulse influences of the "start-brake" type as the neurotrophic influence reactions that are carried out due to the release of special substances – neurotransmitters in the nerve endings. Advances in the study of the synthesis and redistribution of substances in neurons, the transneuronal transfer of macromolecules and low-molecular compounds in the system of interneuronal, neuroglial and neuro-tissue relations, have shown that these mechanisms are the basis of the overall process of interaction between various formations of the nervous system and its regulatory effect on the tissue substrate which is called nervous trophism

In this regard the close attention of researchers once again attracted one of the main systems of regulation used by the body to coordinate its activities in accordance with the requirements of the internal environment and external conditions which is called the autonomic nervous system (ANS).

The ANS performs general regulation by changing the activity of two of its balanced divisions – the sympathetic and the parasympathetic, and using transmitters. It is able to almost instantly mobilize the internal resources of the body for a holistic vegeto-psychomotor reaction to any situation arising from the outside or inside.

On the one hand the ANS functions in such a way that it continuously maintains some relatively constant predetermined level for each physical or chemical parameter of the internal environment exciting or inhibiting various physiological functions of the body in order to minimize deviations of individual parameters and, despite significant fluctuations in the conditions of the environment, to ensure the relative constancy of homeostasis. On the other hand, the level of activity of systems that ensure the interaction of the organism with continuously changing environmental conditions must vary widely to ensure an adequate adaptive response to newly arising unusual or emergency situations that generate special requirements for the composition of the internal environment.

The emergency mobilization of internal resources and maintaining the necessary level of functional activity of organs and life support systems for the self-defense of the body in such conditions are only possible when using a single regulatory system that organizes and manages the coordinated actions of other systems. This is the ANS system which as a divergent system with one “entrance” and many “exits” to tissues and target cells, has extensive zones of influence and integrates and coordinates various activities of the organism.

The advantages of the ANS are that it affects not only the tissues and organs of the main life-support systems but also other neurons, communicating with all hierarchical levels of the nervous system and going beyond specific sensory, motor or associative functions. In addition, it is very important that the effect of the mediators used by the ANS as molecular mediators depends on the conditions in which it is carried out, and may have the opposite direction in different conditions of existence.

It is known that in the early stages of phylogenesis the ANS served mainly to accumulate and conserve energy. Further, at the later stages of evolution as the sympathetic division gradually formed and developed its ability to mobilize the internal resources and expand, the range of CAR of the body began to manifest itself. Summing up the accumulated data and summarizing the results of basic research in this field Leon Orbeli formulated the idea of the ANS as an adaptation-trophic system and noted that the ANS regulates the level of activity of an organ by influencing the processes occurring in it.

The doctrine of nervous trophism contains the main link that combines the theoretical problems of physiology and pathology with the clinic. At present the concept of the neuro-dystrophic component of disease mechanisms is beginning to occupy a proper place in doctors’ thinking, and the idea that it is not a disease to be treated but a sick organism including its nervous system, has become the leading concept in the minds of many clinicians.

Chapter 20

Adaptive, compensatory-adaptive, and pathological reactions of the body

It is not the strongest of
the species that survives,
nor the most intelligent,
but the one most responsive
to change.
Charles Darwin

As an element of the biosphere that forms unity with the environment the organism is the most advantageous system in which the mode of action automatically changes in order to get the best management. Such an automatic adjustment of a living organism to continuously changing conditions of the external and internal environment is called adaptation. An organism is an adaptive system that maintains performance in the case of unforeseen environmental changes. This happens by changing the properties of the managed object as well as the control objectives, by searching for the most appropriate state. Manifestations of adaptation are learning and self-learning, management and self-management processes. It should be emphasized that the adaptation process, that is the achievement of the most advantageous mode of operation, always takes place under certain specific environmental conditions. That is, you can adapt to a disease. And a sharp deliverance of the patient from any long-lasting disease can lead to serious disharmony in the work of all his vital systems and functions. In other words, one should be very careful in treating chronic diseases to which the body has practically adapted.

The stability of the system and its ability to adapt and compensate for impaired functions depend on the number of feedbacks between all its levels and the consistency in their activity by time and place. Based on the integration in the systems of management and coordination activities of the command and execution systems, there is a necessary level of homeostasis.

The living system is constantly in a vibrational state, and the diapason width of various oscillations determines the stability of the system and its ability for adaptation, and compensatory reactions in the conditions of life. The expansion of an adaptive system and CAR is particularly pronounced in a state of tension, when the body tends to increase the period of oscillation in order to extend the recovery time, to accumulate the energy and plastic reserves necessary for their subsequent increased spending.

The system of maintaining homeostasis by the body is extremely complex in its mechanisms and based on the almost infinite variety of its reactions. At the same time all of them are formed from body functions differing from each other only by the originality of combinations of certain functions in each specific case, and are united by a common feature of the compensatory-adaptive essence and homeostatic orientation. Adaptive reactions provide homeostasis in everyday conditions with moderate fluctuations of functional loads due to changes in the rates of biological reactions within the physiological norm. CAR provide homeostasis in terms of exceeding the threshold value of impact (under the action of extreme irritants or damaging agents) on the same basis. That is, these are the same adaptive reactions but more pronounced and aimed not only at ensuring the vital processes, but also at eliminating damaging factors or the resulting defect with the subsequent restoration of functions. At that, any change in the functional state is based on structural changes.

The material basis of CAR and the adaptive reactions are two main types of structural rearrangements: rearrangements reflecting degenerative (destructive) and regenerative (restorative) processes. Inherently reactive, they are both functional and organic. On the quality basis these are divided into degenerative-dystrophic and regenerative-hyperplastic structural changes.

Degenerative-dystrophic changes can be a manifestation of both the breakage (destruction, even death) of structures under the influence of various factors, and an expression of an extreme degree of functional stress (dystrophy or atrophy, or death).

Reparative-hyperplastic changes are aimed at compensating the resulting defect and the subsequent restoration of functions. They are carried out at the expense of intrastructural reorganizations which increase the activity in the SFWU part after the damage as well as in other structures related to it or qualitatively different but replacing it in functional terms. Both types of structural changes in terms of severity, that is on a quantitative basis, may be reversible or irreversible.

These changes take place in the conditions of the norm. They are polar but balanced forms of the same process – the physiological renewal of the

SFWU of the body. In conditions that exceed the threshold value of adaptation but do not go beyond the compensatory capabilities of one or another SFWU the balance of these two qualitatively different forms of structural changes is also maintained although their intensity acquires a different degree of expression due to increasing functional activity. If the functional stress turns out to be extreme or does not eliminate the effect of the pathogenic factor, then the limiting degree of the CAR comes, then the complete exhaustion and loss of the ability to maintain the necessary balance of destructive and reparative rearrangements and, as a result the displacement of functions in one direction or another with appropriate clinical and anatomical manifestations in the form of various symptom complexes and diseases. See Fig.19-2.

Violation of internal differentiation between various control and functional systems of the body or between the organism and its environment leads to pathological homeostasis reactions, to the growth and disorganization of self-sustaining fluctuations at the organism-system level, i.e. to management without feedback. The critical state of the control parameters (the bifurcation point) after which the body spontaneously goes into one of two possible states – improvement or death, is well known to clinicians when they see a patient in "crisis". In an unstable condition when the signs of pathological reactions are fairly obvious it is especially difficult to fight for the patient's life since conventional methods of treatment in which the results are proportional to the efforts, no longer work. Earlier (in the theory of rearrangements) we considered the structure of the development of pathological reactions in terms of self-sustaining fluctuations in an SFWU, and emphasized that if the maximum deterioration of the condition crosses the conditional "death" line the changes prevail, and the nonlinear system (in our case, a sick organism) rushes to death.

Various structures of the nervous system are activated to implement adaptic and compensatory adaptive reactions, primarily these are the CNS and the peripherical nervous system (PNS). The CNS includes those parts of the nervous system that lie inside the skull and the spinal column: the brain and spinal cord. Nerves come and go to the brain and spinal cord through the holes in the cavities of the skull and spine. If they are located outside the skull or spine they become part of the PNS. Some departments of the CNS and PNS can work completely independently even though they are integral parts of these systems or operate with very limited control by the central nervous system. Such departments and their constituent elements form an ANS which is mainly responsible for the regulation of the internal environment: it manages the work of the heart, lungs and other internal organs.

The brain itself consists of a number of easily distinguishable sections. Within the divisions there are many arbitrary boundaries between the subdivisions established by specialists – brain researchers who did not always use clear landmarks. Nevertheless, among the rather confusing neurological nomenclature generally accepted structures are distinguished just as a geographic map highlights states with their own borders. These structures are called regions, complexes or formations which in turn are divided into smaller sections – fields or nuclei depending on the closeness of the separate structural units – and neurons are located in them.

The brain has two paired formations – the right and left hemispheres: the cerebral cortex together with the subcortical white matter. All the formations under the hemispheres are called stem structures since they constitute the brain stem which is similar to the trunk of a tree, and the hemispheres are like its crown.

In the "crown" the main areas are: the occipital lobe (mainly responsible for vision); the temporal lobe (hearing); the parietal lobe (reaction to sensory stimuli and movement management); and the frontal lobe (coordination of functions of other areas of the cortex). The cortex and the sections of the trunk beneath it are divided into five parts: the forebrain (in the trunk are the basal nuclei), the diencephalon, (its most important structures are the thalamus and the hypothalamus), the midbrain and hindbrain which include the pons and cerebellum (small brain), and the medulla oblongata which directly passes into the spinal cord.

The fields and cores of all the listed objects perform various functions: they transmit, redistribute, or save the received information playing the role of a relay, a distributor, or a bank; and control, integrate and test it in order to successfully manage the subordinate structures with the help of nerves and endocrine glands.

In this regard it should be noted that despite a similar histological structure of any part of the same zone of the cortex (for example, the sensorimotor) the physiological functions are distributed within it in different densities. For example, approximately 3/4 of the area of the motor cortex of the human brain is occupied by the process of managing the tone of the facial muscles, the muscles of the hand and the phalanges of the fingers (especially the thumb), as well as of the localization and erogenous zones. And only 1/4 of the territory of the entire motor area remains to manage the prevailing mass of muscles of the body, arms and legs. Firstly, this has considerable importance for human life as a public individual with all means of individual communication (means of communication, transmission and receipt of information) such as voice apparatus and facial expressions. Secondly, the level of general culture (life, production, art, and social activity) of human society is such that the individual does not so much require the subtle control of trunk muscles for climbing trees, rocks, and crawling along hard-to-reach narrow crevices for example, as for the excellent handling of household and work tools which would be almost impossible without the fine control of the muscles of the hand and fingers. Thirdly, this is what is inherent in any living object – the instinct to reproduce and preserve offspring. However, not everything is so simple and unambiguous as described above. Nature left the opportunity to vary the emphasis on managing the tone of different muscle groups and other SFWUs. This possibility is realized due to the fact that there are many more afferent (sensitive) nerve cells than efferent (motor) cells, and therefore there is always a competitive "struggle" between sensitive neurons for "mastering" the attention of a motor neuron and the implementation of a particular motor act.

The entire range of the above complex management functions of the central nervous system is carried out using such a structural unit as a neuron. Separate nerve neuron cells do not function as isolated units, like liver cells or kidneys. Many billions of neurons work by receiving signals from any other cells and transmitting them further. Transmitting and receiving cells are combined into neural circuits and networks. A separate neuron with a divergent structure can send signals to thousands of other neurons. Similarly, a neuron can receive input information from other neurons using one, several, or very many input connections if paths converge on it.

Neurons are usually in two states: quiet (inhibition) or excited (arousal). They turn on and work when they receive a signal that exceeds their level of insusceptibility – the threshold. In this case both temporal and spatial summations occur – the accumulation of signals. In a neuron the number of signals that it receives may vary considerably from a few single signals to several thousand. A neuron is a flexible, reliable and very economical management structure and at the same time a powerful storage base of memory.

Places of connection of neurons are specific areas on the surface of nerve cells where contact occurs, called synapses. The process of transmitting information in these places is synaptic transmission. When neurons interact by means of synaptic transmission, the signal sending the presynaptic cell secretes a specific substance onto the surface of the perceiving postsynaptic neuron. This substance is called a neurotransmitter. It serves as a molecular mediator for transferring information from the transmitting cell to the receiving one. The neurotransmitter thus closes the circuit, carrying out a chemical transfer of information through the structural gap (synaptic cleft) from the transmitting cell to the receiving cell at the point of contact.

The typical structure of a synapse (interneuron contact) is as follows. At the end of the axon (terminal), neurotransmitters are produced with which neurons communicate to each other. When an electrical nerve impulse arrives at the terminal on the axon, neurotransmitters from synaptic vesicles are ejected into the synaptic cleft. Here they interact with receptors located on the membrane of the "receiving" neuron. In addition to receptors there are proteins on the membranes of neurons that pump neurotransmitters from the synaptic cleft.

Neurons are electrically excitable cells which have the ability to regulate their internal electrical potential. In the normal state this potential is negative and due to the state of the inside of the cell. During a brief period of excitation, the inside of a neuron in less than 10^{-3} seconds becomes positively charged. This period from the negative state of the "contents" of a cell to a short-term positive is called an action potential or a nerve impulse. The positive state does not last long because the excitation reaction is self-regulating, and the neuron returns to its original state with a negative potential inside before the next signal. The action potential spreads along the process of the neuron which is called the axon, and as a "wave of activity" reaches all the synaptic endings of the neuron. The main advantage of the electric conduction of the pulse along the axon is that the excitation spreads quickly enough over long distances without any attenuation of the signal. However, the rapid electrical transmission which functions so well in the axon, ceases to work on reaching the synapse: here a chemical transfer comes into play. Without going into the biological causes of this we can simply state that the chemical bond in the synapses provides the most complete transmission of information. So, when communicating with each other you convey the main content of your speech, emphasizing or reinforcing the meaning using stress, voice timbre, and facial expressions. When nerve cells communicate the basic units of information are transmitted by specific chemical mediators – synaptic mediators. If we continue our analogy with the method of communication between people, then we can say that some chemical mediators (sometimes called transmitters) convey "facts" while others convey additional semantic shades or accents.

The human brain differs from the animal brain not only in its mass but also in the incredible interconnectedness of its parts. These relationships greatly expand our ability to capture and evaluate the information that we perceive and later retrieve from memory. This in turn gives us the ability to develop analytical strategies and estimate the results of past experience which far exceeds the capabilities of a brain endowed with fewer connections.

Chapter 21

Higher nervous activity and mental functions. Consciousness

The history of the human soul,
even the smallest soul,
is hardly no more curious
and no more useful than
the history of an entire people,
especially when it is the result of
the observation of a mature mind
on itself and when it is written
without a vain desire to arouse
participation or surprise.
Mikhail Lermontov

The whole world around us is reflected in the human brain. Contact with this world and its effects on the body are possible thanks to highly specialized nervous apparatus called sensory systems.

In the sensory systems of the brain there are a number of levels of information processing. Each level is constructed in such a way as to ensure the most complete and accurate communication between unambiguous elements of the same level, and wider interactions between all elements of each level. In the most simplified form, any sensor system has two main channels; a specific channel that has a direct signal switching to certain points of the cerebral cortex and an associative channel that receives signaling from the result of the interaction of various specialized channels of a given sensory system.

The scheme of both channels includes several levels of signal processing: receptor, stem, thalamic and cortical. Between the levels in both channels there are two-way communications that provide the most complete processing and “imprinting” of information, the comparison in the operational memory of traces of available information with new signals.

All of this together provides a full recognition of the image and an assessment of its adaptive significance. The associative structures of the brain are the centers that ensure the unification and integration of all sensory systems since every integral function of the brain is polysensory. For example, the function of visual recognition is carried out in the interaction of visual, motor and vestibular sensory systems. Consequently, we are dealing with a complex reflex activity, through which a person most fully manifests the world around him for himself. The main elements of CRs are: memory (structural trace), perception (reflection of the present) and an adaptive reaction which makes sense as an organism's adaptation to the foreseeable future. Apparently, in the conditioned-reflex act the presence of a basis is manifested in which elements of the past, present and future are already laid. By applying and testing in practice the correctness of these reflections a person comes to the truth.

The creation of multiple and diverse forms of intercellular connections reflects the evolutionary law of morphological progress in the nervous system. With the increase in functional adaptation in neurons the formation and lengthening of a new nerve ending, the size of the nerve cell body and the diameter of the axon are not associated with specialization but are proportional to the richness and spread of end branches, and consequently to the abundance and variety of connections. All fluctuations in topography and morphology as well as the theory of dynamic axon polarization itself (a strict excitation direction from an axon), are apparently determined by three principles of economy: saving matter (the development of the shortest path between two related areas), saving time (a dynamic consequence of the previous paragraph) and saving space. Each peripheral irritation produced by the sensitive processes of a single nerve cell (dendrites) spreads to the avalanche centers. In other words, the number of neurons involved increases from the periphery to the large brain, in the convolutions of which (tertiary sensitive fields) lies the base of the cone of the common sensory inflow and the beginning of new connecting pathways. In the last century this anatomical and physiological law was formulated by Cajal as the law of the unity of perception and nervous avalanches and was successfully used by him to explain the mechanisms of such psychological processes as attention, memory, association of ideas, etc.

Cajal claimed that adaptation, intellectual development, and professional improvement of functions by exercise (physical education, speech, writing, playing the piano and other types of activity) are explained both by the gradual thickening of the pathways excited by the passage of impulses and by the formation of new endings (the growth and development of new dendrites and the distribution and branching of sensory collaterals), capable of improving the adaptability and length of contacts and even creating entirely new connections between initially independent neurons. According to Cajal, mental abilities and their most noble manifestations (talent and genius) depend not on the size or number of cortical neurons but on the richness of their connecting endings or in other words on the complexity of associative pathways over short and long distances. The position that was earlier defended by Theodor Meynert and Paul Flechsig, is that the abundance of white matter indicates a wealth of connections and consequently a higher intellectual rank. Subsequent studies fully confirmed this position. Scientists now believe that structural changes in the nervous system can be the result of learning processes and storage of memory traces. In other words, behavioral changes resulting from experience are developed through learning and memorization and can be fixed at a structural level.

There are two different ways of learning and remembering information depending on what needs to be learned. Procedural knowledge is knowledge of how to act. At the same time, acquired skills remain at a high level for quite a long time. However, there is no awareness that learning has occurred and there is no awareness and recollection of the ways to solve the set task. It is this second, declarative knowledge that requires the special processing of information and is associated with consciousness and thinking.

Everyone knows what unconsciousness is. This is the state when we are not aware of our thoughts and actions, for example when we are stunned by a blow to the head or faint. The concept of consciousness is so multivalued that it does not have a simple definition. Traditional neurophysiology asserts that consciousness is a function of the brain, in the cortex of which information coming in through the senses is combined with information extracted from the storerooms of our memory. This allows us to intelligently interpret specific (visual, auditory, tactile, etc.) sensations. We respond to them differently depending on how current information interacts with emotions and memories, targets, self-esteem and the value scale (hierarchy of needs).

However, due to the fact that neurophysiologists are still not able to definitively identify those brain mechanisms that are responsible for consciousness, its nature remains the subject of eternal and incessant debate in which philosophers, theologians, cybernetics experts involved in artificial intelligence, and neurobiologists are involved at all levels. One of the issues that causes controversy is whether, after all, consciousness is a specific material function of the brain or a separate intangible process, the embodiment of the spirit or soul. Without pretending to address this issue, it is still possible as the first working definition of active consciousness, as a state in which we are aware of our thoughts and actions, to accept the following. Consciousness is active declarative knowledge about our mental and physical activity.

Declarative knowledge includes fixing individual and public experience, information about when and where events stored in memory took place, and the possibility of their internal reproduction in the form of memories, structuring one's own and other people's experiences, allowing one to reflect on the phenomena and relationships between them, making plans, and giving actions the meaning which follows from the goals. The ability for declarative knowledge of course does not mean that all actions of an organism are constantly recognized. Many types of mental and physical activities are carried out without connecting consciousness which is only one of the complex mental processes that reflect reality.

It is no coincidence that the Nobel laureate Ivan Pavlov united the cortical divisions of analyzers where information is decoded into the first signal system of reality. Mental processes such as memory, consciousness, planning, etc., according to Pavlov relate to the second signaling system. The joint activity of both signaling systems is a higher nervous activity, and the differences between animals with a brain are in how developed these systems are, and what role mental processes play in life processes.

Mental processes include all forms of reflection of reality: sensations, perceptions, ideas, emotions, will, mind, intellect, etc. In essence, in content they correspond to the basis on which the inner world of a person is formed.

At the level of sensations, the images that form in the cortical parts of the analyzers seem to be similar in all animals with a great gray mantle of the crust. They can be identified by a general reaction to an irritant but in addition to this, images are associated with genetic information and accumulated individual experience, making each of us unique and inimitable. An active search for an organizational principle that would ensure approaches to the phenomenon of consciousness led to at least one indisputable statement. A species-specific type of behavior unique to *homo sapiens* is the use of language associated with the important process of sequential processing of information in one of the hemispheres of the brain.

If consciousness is really connected with language, then it doesn't appear in a person from birth. It develops as life experience is gained and a "vocabulary" is acquired which allow one to reflect on the phenomena and connections between them, plan actions and predict results. And most importantly, there arises the ability to self-analyze, an awareness of one's self, the transition to self-awareness in the widest sense of the word and to the public consciousness.

According to some psychologists, in particular Julian Jaynes, such forms of consciousness have arisen in the history of the human race surprisingly recently as a result of certain changes in language and culture that occurred somewhere around the VII century BC. Characteristically, originally in Greek the word “psyche” meant “life”, “living state”, and “soma” meant “corpse” or “lifeless state”. Only thanks to the writings of Pythagoras and other thinkers, psyche began to denote “soul”, and soma – the body. Jaynes believed it was a mistake to assume that it was only a change of words. A change in words is a change in concepts, and a change in concepts is a change in behavior.

The whole history of religion, politics and even science convincingly testifies to this. Without such words as “soul,” “freedom,” and “truth,” there would have been other roles, different culminations in the drama of human history. In addition, Jaynes believed that the nervous substrate of consciousness is sufficiently plastic so that, on the basis of training and culture, a transition can take place to self-consciousness. This statement cannot be confirmed by scientific research since experimenting on the living human brain is simply impossible. At the same time, scientific facts that are associated with special problems of consciousness, thinking, and ultimately human behavior were nevertheless obtained and have the greatest value.

Extensive use of the results of studying the brain and animal behavior, reliable information about the functioning of a healthy and diseased human brain, in some ways allows us to find out the ways in which the nervous system coordinates the body’s needs with the conditions of the external and internal environment and organizes a behavioral act aimed at achieving goals. The generalization of the principles of the more or less studied systems can really lead to an understanding of complex processes that are close to thinking.

Today we propose to adhere to the point of view that all conscious and unconscious mental processes in principle are derived from processes in the nervous system and can be considered as its function. If such methods are invented and data are presented that convincingly show that spiritual activity is not a property of the human brain then we will have to abandon one erroneous theory in favor of another, less erroneous one.

At the present stage of development of biology, it has been ascertained that conditioned reflex human activity has a lot in common with the conditioned reflex activity of animals. In humans CRs are produced in the same way as in animals, with the help of starting and reinforcing stimuli. Reinforcing stimuli are the unconditional reactions of salivation, respiration, changes in heart rate, and motor, galvanic skin, defensive and other reactions. As well as in animals, human CRs are capable of fading, they are subject to external inhibitory irritants, dynamic stereotypy, etc. At the same time the HNA of a person including mental reflection is characterized by a number of essential differences.

In quantitative terms, among them are the following differences.

1. High speed of all types of learning.
2. The possibility of developing high order CR (associations).
3. Developed ability to differentiate.
4. Quick alteration of behavioral stereotypes.
5. Developed ability to predict and plan.
6. The predominance of social over vital needs.

Qualitatively, man’s HNA is characterized by a stronger development of mental reflection (animals only have certain of its elements), the ability to store, process and transmit information using symbolic sign systems (languages), and the awareness that learning has occurred. The most significant in human activity is the verbal sign system, i.e. speech (oral and written). Other signs (symbols) are used: numbers, notes, drawings, gestures, facial expressions, etc. Man is able to give conventions to certain concepts, operate with these symbols and transmit information using external media (from petroglyphs to computer disks). On this basis man develops abstract thinking, consciousness, self-awareness and social consciousness, abilities for artistic self-expression, artistic and scientific creativity, as well as the ability to transfer knowledge and experience from generation to generation to replenish and develop social relations, art, and science, that is for all that we call cultural and historical values.

The listed features of man’s HNA are reflected in Pavlov's ideas about the first and second signaling systems associated respectively with concrete and abstract conventional signals. These signals are triggers in the implementation of HNA and are important incentives. Unconditioned stimuli that have a “reinforcing” value are not included in the structure of signaling systems.

According to Pavlov.

1. The first signaling system is a system of temporary connections that occur in the brain when exposed to signals with a specific content. They reflect certain signs of objects and phenomena, as well as events of the surrounding world within the current moment.

2. The second signaling system is a system of temporal connections arising in the brain under the action of abstract signals (words and other symbols that encode generalized concepts and are relevant both for analyzing past experience and predicting the future).

Pavlov distinguished three levels of HNA.

1. The UCR.

2. The CR of the first signal system associated with specific signals (for example, a dog’s response to its name, a person’s response to a phone call).

3. The CR of the second signal system. This is associated with abstract signals, i.e. symbols.

The first two levels are common to humans and animals. The third is peculiar, mostly only in humans although higher animals have certain beginnings of the CR of the second signaling system which arises only after training by humans.

The works of ethologists show that animals use sign systems quite well: “tongues” of sounds and smells during ritual behavior (for example, in the mating season); special signals for the transfer of certain information (for example, signals of imminent danger, of the direction and distance to the source of food or water, of territorial law, of a “leading” or “subordinate” position). However, the “language” of animals is a collection of specific signals and refers to the first signal system. Abstract signals only as a last resort, and only after training by man can be perceived by some highly organized animals. For example, there is a case when a couple of chimps managed to teach sign language to deaf and dumb people. This included 90 signs denoting objects, actions, and qualities. Using this language, monkeys communicated with experimenters and even among themselves but the complexity of the problems solved by animals using sign systems did not exceed the level available to a two- or three-year-old child.

Currently, apart from the levels that Pavlov singled out within the framework of the physiology of HNA such mechanisms of mental reflection as emotions, memory, thinking and consciousness are also being studied. Pavlov's foundations for the clinical physiology of HNA as the science of pathological changes in HNA and sensory systems, and in particular the teachings on neurosis and mental illness were widely developed.

In a healthy organism any manifestation of activity is determined by expediency (the so-called conformity postulate), i.e. any behavioral act is aimed at achieving a specific goal. From the point of view of management theory, the organism is an optimized system that automatically changes its behavior (methods of action) in order to achieve better management. In this regard the processes of self-organization and development are determined by the three principles of economy.

1. The economy of matter (the development of the shortest channel of communication between two interacting areas).

2. The economy of time of information transfer.

3. Space saving – structuring, i.e. the acquisition of certain forms of carriers of functional systems allowed to fit into the limited volume of an organism.

At the same time the organism itself is also limited to certain sizes. For optimal functioning, for example, human growth is considered normal in the range from 1.5 to 2 m, other sizes of the human body predetermine failures in the work of various functional systems.

The management activity of the nervous system is also regulated by three principles.

1. Homeostatic orientation. This is the elimination of conflict in the interaction of the organism with the external environment including with society; elimination of long-term stress conditions; and establishing a balance between various physiological and mental processes occurring in the body, i.e. maintaining dynamic homeostasis.

2. Hedonistic expediency. This maximizes the feeling of pleasure (comfort) and minimizes suffering (discomfort).

3. Pragmatic expediency. This is an optimization of behavior aimed at achieving the goal (benefit and success are achieved with the lowest cost).

The existence of the organism in constantly changing environmental conditions leads to the fact that external irritants cause changes in the body, and the body is forced to maintain its relative constancy within a certain range continuously adapting to each specific situation. Personality also adapts to the changing conditions and requirements of the environment only at the level of mental reactions. Mental reactions are based on physiological processes and have the same regularities. They are also aimed at preserving homeostasis including "mental homeostasis" bearing in mind the spectrum of adaptive reactions of the psyche.

Adaptation as one of the most important properties of an organism developed in the process of evolution, manifests itself in different ways but always with one goal – to achieve the most advantageous mode of operation in specific existing environmental conditions.

It is possible to adapt to illness or an emergency, therefore abruptly relieving a person of any disease or a too fast transition from an emergency situation to normal conditions can lead to serious disharmony of all his functions including mental ones. This is why such serious attention is given to clinical (medical) psychology which deals with patients suffering from any somatic diseases. At the beginning, clinical psychology which was actively developed by psychiatrists was aimed at studying the deviations in the intellectual and personal development of such patients, as well as at correcting maladaptive behavior. However, over time, the scope of clinical psychology was expanded by studying the mental state of patients with somatic diseases in terms of their adaptation to the conditions of the disease. Manifestations of adaptation are the processes of learning, self-study, management and self-government, i.e. in fact it is the processes of self-organization.

The stability of any functional system including the human body is one of the most important characteristics of its activity. Stability reflecting the ability of the system to adapt and compensate for impaired functions depends on the number of feedbacks between all its levels and consistency in their activity by place and time. On the basis of integration, the necessary level of homeostasis is ensured in the management and coordination systems of the actions of the command and executive structures. A live information system is constantly in an oscillatory state, and it is the width of the range of different oscillations that determines the stability of the system in the conditions of life. Particularly vividly, stability manifests itself in a state of tension when the body seeks not only to expand the range of adaptive and compensatory-adaptive reactions but also to increase the period of oscillations in order to prolong the recovery time and accumulate the plastic and energy resources necessary for their further increased spending. Thus, stability is one of the characteristics of complex systems. A brief definition of stability is the ability of a system to maintain its internal properties under conditions of disturbing influences.

The area of stability of the system is the boundaries of changes in its parameters at which the system performs its functions quite effectively. The concepts of the physiological and mental norm are connected with efficiency. Performance indicators (from the point of view of the observer) are the main criteria of the norm, they reflect the degree of fitness of one or another functional system to fulfill the tasks assigned to it. Efficiency itself as a functional characteristic is associated with the quality of the performance of the objective function and the magnitude of the costs that went towards the goal.

Performance indicators can be expressed both quantitatively (for example, by numerical values, which can describe internal processes when measuring certain parameters) and qualitatively (according to the principle "better – worse"). In favorable conditions the performance indicators have the same value in terms of counteracting adverse effects, when the system uses additional funds seeking to stabilize any process or parameter the performance indicators decrease. Nevertheless, in physiology and psychology there are concepts about the "conditional norm" which includes CAR and adaptive reactions, and about pathological processes when it comes to maladaptation and decompensation.

Speaking about humans, one cannot but touch upon the question of concepts that are characteristic of *homo sapiens* such as consciousness, subconsciousness and superconsciousness. Let us briefly discuss their characteristics.

Consciousness is a specific human form of reflection of reality, the formation of an internal program of actions based on operating with declarative knowledge. Received and transmitted knowledge allows a person to separate himself from other animals and non-living systems. To realize it means to acquire the potential opportunity to communicate and transfer your knowledge to another including other generations in the form of cultural monuments. Consciousness unites all that is communicated or can be communicated to other people. The communicative origin of consciousness determines the ability of a mental dialogue with oneself, judgments about one's own actions, planning actions and predicting consequences, that is, the emergence of self-awareness.

In relation to communicative consciousness, speech is the most important leading means of declarative knowledge but one should not forget about the grandiose system of artistic images that, without being completely verbalized certainly belongs to the sphere of consciousness. The human perception of art involves both the "primary language" of sensually direct concrete images and the "secondary language" of preverbal concepts (the result of the activity of the first signal system) and verbal concepts (the results of the activity of the second signal system) inherent exclusively to man. The HNA of man includes not only the first signal system, common to all higher animals, and specific human speech, i.e. the second signaling system, but also the third factor associated with the process of imagination, mental representation, the so-called imaginative factor directly connected with consciousness. Experimental and clinical studies of the functions of the large hemispheres have confirmed that the preservation of the connections of the cognitive zones of the cortex with the speech structures of the brain is an indispensable condition for the functioning of consciousness.

The subconscious is a kind of unconscious mental activity that is directly related to consciousness and includes everything that was conscious or can become so in certain conditions. This group of phenomena includes good automation and therefore ceases to be aware of the skills as well as the motivational conflicts that are repressed from the sphere of consciousness. The sphere of the subconscious includes deeply learned social norms. In the subconscious mind are those manifestations of intuition that are not associated with the generation of new information, and involve the use of previously accumulated experience. It is the subconscious that is the source of many conclusions where it is often impossible to explain exactly what external signs of an object or phenomenon led to this conclusion if you do not conduct a mental reconstruction of a previously realized experience. There is also a direct path that bypasses the rational control of consciousness. These are mechanisms of imitation behavior which over time become internal regulators of actions and often form a personality to a greater extent than the intellectual sphere. The subconscious mind is always guarding the procedural and declarative knowledge obtained and well learned be it an automated skill or a social norm. Thanks to the conservatism of the subconscious the individually learned (conditioned-reflex) knowledge acquires the imperativeness and rigidity inherent in the UCR.

Superconsciousness is the generation of new, previously non-existent information by the recombination of traces of impressions received from outside bypassing the control of consciousness, although conscious experience and reserves of the subconscious can serve as material for recombination activity. The sphere of superconsciousness includes the initial stages of creative activity (sometimes even in the form of dreams) – the generation of conjectures, hypotheses, and creative insights. The original formulation of the problem and its setting before the cognizing mind occurs in the subconscious and the conscious, but fundamentally new information not directly derived from previously accumulated experience is generated by superconsciousness. Then it is transferred to the level of consciousness where the secondary selection of information takes place through logical assessment and verification in practice. Creative intuition contains two types of superconsciousness; intuition-guesswork is a production of hypotheses and intuition-judgment is a product of axioms, i.e. direct discretion of the result (truth) that does not require formal-logical evidence. In their genesis lies something fundamentally common, namely the lack of information necessary and sufficient for a logically flawless conclusion. In the first case (intuition-guessing) this information is not yet available but it can be obtained in the course of testing the hypothesis. In the second case (intuition-judgment) the necessary information cannot be obtained at all since the existing level of individual and public consciousness is not yet able to provide a secondary selection with the help of formalized evidence. In this case the resulting new information remains at the level of intuition and cannot be formalized, i.e. this is an axiom.

Psycho-heuristic intuition-guesswork within the framework of a formalized system of axioms can thus be discursively reduced to some basic assumptions taken as axioms or postulates. Intuition-judgment is precisely the one that differs from intuition-guesswork which fundamentally does not boil down to any axioms since it has an axiom character and in some cases it is an axiom.

Establishing axioms is always an act of intuition-judgment. The most important and fundamental for the whole of such scientific judgment is the judgment of the sufficiency of the experimental test, of the provability of experience which is always inevitably limited. However, we use this judgment in all cases when we resort to the criterion of truth.

CHAPTER 22

STRUCTURE AND CLASSIFICATION OF HUMAN NEEDS

Nosce te ipsum
The inscription on the pediment of the temple of Apollo at Delphi

The idea of the basic needs of man has developed on the basis of scientific data on the cerebral mechanisms of animal behavior both in terms of ethology and in terms of developmental biology. Extensive use of the results of studying the brains of animals, reliable information about the functioning of a healthy and diseased human brain to some extent, allows us to discover the ways in which the nervous system coordinates the needs of the body with the environmental conditions and organizes a behavioral act.

The basic evolutionary needs of all living information systems are the needs for substances, energy and information. These needs are necessary for the growth and development of any living information system from a cell or subcellular organism (viruses) to a multicellular living organism with a psyche.

In phylogenesis the needs inherent in living information systems evolved along with the information systems themselves, and in humans they are the most diverse. Along with that, they are based on a single genetic material. The phylogenetic prerequisite for the development of human needs is constituted by the requirements that cause UCRs. Innate human needs are basically the same as in animals, however they are qualitatively transformed by the cultural-historical process of developing thinking and consciousness.

In ontogenesis, the needs of each person based on the same needs for all of us, develop depending on the life story. Therefore, the behavior due to UCR is the same for all animals of the same species, and the behavior based on CR and mental activity is entirely individual for each subject.

Before turning to human needs, we note that from the point of view of the information theory the most complex UCRs (instincts) are innate, specific to each animal species, a form of informational reflection which is an integral behavioral complex that includes the motivating and supporting components. For example, in the implementation of eating behavior the motivating component is the feeling of hunger, and the reinforcing component is the external object which is food.

The most complex human UCRs based on the satisfaction of its innate needs, are the same as in higher animals and can be divided into three groups: vital, zoosocial (role) and self-development and social development reflexes.

Vital UCRs: food, drinking, respiration, thermoregulation (regulation of body temperature), reflex regulation of sleep and wakefulness, defensive reflexes (the so-called reflexes of "biological caution" or runaway and "biological aggression" i.e. attacks), the power saving reflex and some others.

There are two criteria for the assignment of the reflex to the group of vital UCRs.

1. If the reflex does not work and the need is not satisfied, this leads to the death of the organism. Vital related reflexes can thus satisfy the needs necessary for the sustenance of one individual.

2. The second criterion is the method of realization of the reflex: the vital reflexes can be realized without the participation of another individual of the same species, they are individual in nature.

The specific form of the realization of a reflex depends on the situation. For example, in a hare driven into a corner while fleeing it is replaced by attack. In fact, this is the same defensive reflex in its two manifestations. Usually, the hare is not peculiar to aggressiveness, but when there is nowhere to run, it can rush at a predator and often defeats it.

An example of a person's innate vital reflex is the fear of heights. It is well manifested in newborn monkeys and human cubs despite the fact that in their individual experience they are not familiar with the danger of heights. When trying to raise a child to a height, he instinctively clings to any object that can keep him from falling. Infants hold tightly onto the finger stretched out to them and can hold on for a very long time (grasping reflex). In this form, they manifest innate reflex protection against falling from heights.

The second group of UCRs is zoosocial (role) reflexes. They are associated with sexual, parental, territorial, and hierarchical behavior and are aimed at preserving the species. This also includes the compassion reflex, i.e. a reaction to the emotional state of another individual, the so-called phenomenon of emotional resonance (in psychology – empathy).

Zoosocial reflexes can only be realized through interaction with other individuals of the same species. Entering a group relationship, a certain individual always plays any role: the marriage partner, the parent or a cub, the owner of the territory or the newcomer, the leader or the led. Any interaction between two or more individuals of the same species is "social behavior" which is based on zoosocial reflexes.

The third group of UCRs is the reflexes of self-development and social development which include the following.

1. The research reflex is the need to obtain information that enables animals and humans to master new territories, gain new knowledge and skills, and use new and unusual situations.

2. The reflex resistance to coercion and its variety including the reflex overcoming the obstacle (Pavlov called them reflexes of freedom).

3. Reflexes of "preventive armament" which have no pragmatic significance, do not implement actions aimed at specific goals, would prepare an individual for possible behavior in a newly emerging situation, and create a program of actions aimed to the future. These include play and imitation reflexes.

The third group of UCRs shows the outrunning nature of information's reflection. For reflexes of the third group the main feature is that they are not associated with individual or group adaptation to habitual situations and environmental conditions. Most often they are focused on extreme conditions or as they still formulate, on extraordinary impacts. Figuratively speaking the third group of UCRs turns to the future and focuses on the development of new situations and habitats. Imitation and game reflexes are based on the ability to manage events (to master events), and to be competent (competence drive). These UCRs occur in both animals and humans. They are associated with the acquisition of skills and abilities that have no practical value in the present tense although in the future they may become necessary for the successful satisfaction of vital and social needs. For example, in young animals the formation of behavior associated with the extraction of food is carried out with the help of the game reflex of the young which are fed by the parents, and in the game they are only learning to get food on their own.

Training skills and abilities occur both in the game and in imitation. With the help of the imitative reflex young animals acquire certain skills reproducing the actions of a demonstrator – a parent, leader, or master. It is important to note that imitative behavior promotes learning only at an early age. For example, monkeys not trained to build a nest in childhood lose their ability to learn this skill as they become adults. No matter how much they observe how the nest is built they are not able to master this skill. People with social deprivation (lack of communication with their own kind) have children-animals who cannot be taught human communication, in particular, taught to speak after 8-10 years. This is why we talk about the decisive role of the personal example of parents (caregivers) in the development of many skills in a child.

A brief acquaintance with the needs that underlie a basic UCR of higher animals and humans programmed at the genetic level allows us to approach a number of issues that do not have a satisfactory solution within psychological knowledge, and sometimes are not posed by psychology due to an underestimation of their key preservation of mental and psychosomatic human health. One of the reasons for this state of affairs is the task of systematically studying the fund of innate and acquired human needs which has not been fulfilled so far. Nevertheless, the idea of basic needs has taken shape, and there has been a very successful attempt to classify them. According to Simonov, basic human needs can be classified as biological (vital), social, and ideal (Simonov 1991).

Biological (vital) needs for food, water, sleep, thermal comfort, protection from external adverse effects, etc., are designed to ensure the individual and species existence of man as part of nature. They generate many secondary and tertiary needs such as material needs for clothing, housing, means of transportation, and protective structures. Among the biological needs is the need to save power, which motivates a person to search for the fastest and shortest way to achieve his goal and get the best result from the least energy and material costs. The principle of saving power lies at the basis of improving skills and inventiveness however it can transform into laziness and become paramount in the hierarchy of needs.

Social needs include the need to belong to a particular social group (community) and occupy a certain (not necessarily dominant) place within it, to enjoy the attention and respect of others, and to be the object of their affection and love, i.e. a need for communication with their own kind. Despite the fact that in different people this need varies in intensity, its complete absence is a clear criterion for mental disorders or even mental illness. Interacting with the social environment, a person seeks two goals: to merge with the community, becoming a fully-fledged member, and at the same time highlighting his “I”. Based on this, social needs include the following.

1. The need for social security.
2. The need for affiliation (“fear of loneliness”).
3. The need to achieve a certain social level.
4. The need for self-actualization: independence, self-identification, and self-esteem.

In the structure of self-esteem 20% is occupied by the external side (as estimated by others) and 80% by the internal side. With the help of external factors, the lack of internal self-esteem (self-sufficiency) is compensated.

The development of a child is evidence of the independent origin of the social needs of man. Thorough research has shown that the need for attachment and the fear of loneliness are not derived from either the need for food or early sexuality, as Sigmund Freud believed. Moreover, if the functions of satisfying biological needs (feeding and hygienic care) and communication in the form of smiles, conversations, and games are divided between two people, then the child becomes more attached to the person who enters into psychological contact. At the same time, he remains indifferent to an adult who is passionlessly satisfying his first vital needs.

The ideal needs of man based on the need for information include the need for knowledge of the laws of existence and the development of the surrounding world and its place within it, both by appropriating existing cultural values and by opening completely new ones, unknown to previous generations. These needs exist without regard to the pragmatic satisfaction of vital and social needs, although they are closely related to them. They are directed to the future and are realized not only at the level of consciousness but also at the level of superconsciousness. In contrast to animals, man seeks to understand the rules and laws that govern the world around him.

On comprehending these laws, man lays them at the basis of the models of the world that he creates, be they scientific theories or works of art. Ideal needs underlie creativity, and they encourage people to participate in such forms of knowledge of the world as science, art and religion. Science satisfies the human need for the knowledge of objective truths. It belongs to the rational sphere of thinking and includes the discovery of scientific facts, the creation of hypotheses and theories that allow you to fix the object or phenomenon and identify the properties and relationships, regardless of subjective perception.

Art claims to be subjective truth while religion satisfies the need to search for absolute truth. However, art and religion unite the fact that they are engaged in the search for truth that cannot be explained with the help of a rational approach. Art and religion belong to the irrational sphere of thinking but rely on different theses. Art relies on the thesis “I feel” and religion relies on the thesis “I believe”. While the main thesis of science is “I know”.

In the structure of biological, social and ideal needs there are two types: the needs of conservation and the needs of development. At the same time, there is a distinction between the needs of individual preservation and development (selfish needs – for oneself) and the needs of social preservation and development (altruistic needs – for others).

Conservation needs will be met within the existing current living conditions and individual and social norms. Development needs exceed the norm and are aimed at expanding the limits of habitats and changing living conditions for the better (improving the quality of life). For example, in the sphere of social needs, individual self-preservation and self-development are manifested in the desire to preserve, consolidate, or improve their social behavior, while the public requires improving social norms or improving positions of a social group. There are legal (laws and punishment for their violation), moral (public morality and etiquette, including censure for their violation), and aesthetic social norms (concept of the beautiful and ugly). The latter group of norms is the least regulated, but their non-observance can also cause public censure, rejection of the subject from the group or his rejection to one or another community. In ideal needs, the norm is the knowledge achieved by now. The ideal need for conservation is satisfied by the appropriation of knowledge, and the need for development induces a striving for the unknown. The ratio of needs "for oneself" and "for others" in the personality structure allows one to single out rational egoism and rational altruism in normal conditions, as well as deviant (deviating from the norm) behavior – egocentrism and unreasonable altruism. Concerning the observance of social norms, "innovators" and "conservatives" (ordinary people) stand out. Addicts, extremes and "conformists" differ in deviant behavior. In the animal world, any "extremism" is not welcome. The female will never mate with a partner whose behavior shows signs of "extremism" or any other "wrong" deviant actions.

The needs of conservation and development give rise to two basic forms of emotions: negative and positive.

The mechanism of conservation (survival) does not need positive emotions. It can exist at the expense of one negative emotion, when eliminating discomfort by meeting a particular need is a "reward" for a person. Positive emotions are a mechanism of development based on the need to overcome obstacles for the achievement of a goal, the primary initiating behavior as well as the need for "armament". While negative emotion is a starting stimulus, an urge to action positive emotion as pleasure from overcoming or from acquiring new knowledge, skills, abilities, or even material benefits, is a reinforcing stimulus. Therefore, the need for overcoming and the need for "armament" are considered additional to vital, social and ideal needs.

The ability to overcome obstacles is usually called the will. From the point of view of neurophysiology, the will is the activity of the subject due to the need to overcome an independent and additional motive that determines the behavioral act. The pleasure of overcoming is the most striking indicator of the will. At the same time the activity caused by the obstacle in a certain situation and of certain subjects, can push the initial behavior into the background. Then you can meet with unreasonable stubbornness, i.e. with this type of behavior when overcoming has become an end in itself, and the original motive has lost its meaning or has even been forgotten.

A property that correlates negatively with the will is suggestibility which includes a predisposition to hypnosis. This allows you to use suggestibility as a test to assess the volitional qualities of the individual. Numerous studies have shown that suggestibility does not depend on the stability of the psyche but on the type of thinking. Rational thinking reduces the level of suggestibility and irrational thinking increases it. In ontogenesis the highest degree of suggestibility is observed in people under the age of 25-30 years and especially in children and adolescents from 6 to 18 years. On the one hand, it contributes to education, and on the other hand, it causes the danger of a long delay in imitative behavior in children or a return to it in adults. In this case there is complete subordination to the "leader" – the demonstrator, the loss of self-consciousness and a sense of personal responsibility for their actions and deeds. A sense of personal responsibility, as well as a mechanism for predicting the consequences of an action, is formed in the process of the development of the human brain, its thinking and consciousness. In a person with a developed mindset and consciousness, suggestibility is reduced. The danger is much less with analytical thinking which allows you to include mechanisms for a comprehensive and repeated analysis of your actions and their consequences, so that a person is able to independently confront unfavorable situations and look for a way out of discomfort without relying on the help of an "all-knowing leader".

Independent (by origin) needs for "armament" and "overcoming obstacles" along with other needs participate in the regulation of behavior and contribute to the formation of personality. The hierarchy and combination of needs for each person have their own unique character. According to the set of needs, people can be fundamentally similar, but in their hierarchy and combination they can radically differ. Along with the individual characteristics of a person's somatic and psychological constitution, the set and hierarchy of his needs contribute to the personality structure, forming the physiological, psychological, and social components of individuality.

Most of the components of individuality are finally formed by the age of majority and practically do not change during life. Deviations occur in conditions of long-term emergency stress, life-threatening events, severe and long-term somatic and mental diseases, and may concern all aspects of personal individuality including needs. Normally, for any living creature and especially for a person with the largest set of needs, those activities that simultaneously lead to the satisfaction of a complex of needs, and achieve the highest degree of qualification and professional skill, are of particular value.

For example, work is capable of satisfying a material need for earnings, a social need for defining a public niche or achieving a prestigious position in a social environment as well as an ideal need for acquiring additional knowledge if a person seeks to introduce elements of creativity into his work.

Apparently, social and ideal needs are fundamental in the formation of people into ethnicity.

CHAPTER 23

PSYCHOSOMATIC DISORDERS

Though this be madness, yet there is method in't.
William Shakespeare

In psychology between the conditional norm of the personality and its disorders (in the old use of the term – psychopathies) are the so-called accentuations which are sharp (pointed) manifestations of mental processes that do not reach a pathological degree. In fact, these are pronounced CARs which are located on the border between the conditional norm and pathology.

There are still disagreements among psychologists, psychiatrists and clinical psychologists on the interpretation of manifestations of the norm, accentuations and pathological symptoms. The same applies to deviant behavior; deviant in our context – anomalous.

Deviant behavior is defined as a system of actions or individual actions that contradict the norms accepted in society and manifest in the form of unbalanced mental processes, low adaptive capacity, disruptive self-actualization processes or in the form of a deviation from moral, aesthetic or ethical control over one's own actions. In order to judge the deviations in behavior and about the degree of pathology, it is necessary to determine the "starting point", i.e. that individual or social norm in relation to which a particular behavior is deviant. It is important to remember that the same symptoms can correspond to both psychological manifestations of individual personality traits (standard variants) and psychopathological changes associated with the disease. In assessing a symptom, it is impossible to base it on the definitions given by the interlocutor, especially if a person with a high level of development and broad knowledge in philosophy, psychology and medicine is being examined. The attempts of the subject to not only describe his condition but also independently analyze it increase the risk of an incorrect assessment of one or another mental process when the particular manifestations of the psyche in a particular situation are difficult to differentiate from the clinical symptom of "mental disease". For example, the subject may complain of apathy and melancholy while objectively he has hypopathy and anxiety. In this regard it is very important to understand the terminology and try to give the most accurate formulation of a particular concept.

The similarity of negative and positive mental changes in healthy and sick people causes difficulties in the differential diagnosis of the psychological characteristics of the person and psychopathological manifestations of the disease. Until now it has remained unclear whether the emergence of psychopathological phenomena in healthy people is possible or whether during the period of their appearance such a person cannot be considered completely healthy and one should at least recognize that he has neurotic disorders. However, knowledge of the physiology of the HNA and sensory systems allows us to understand that many of the so-called psychopathological symptoms cannot be definitively categorized as pathological manifestations since they are often situational manifestations of the normal variants. Despite the fact that there are still no clear differential diagnostic methods for distinguishing many painful and non-painful manifestations of the psyche the most important criterion for assessing the options for the norm of mental processes is their relatively expedient nature and the absence of an excessively long fixation that does not meet the requirements of the current reality or the needs of an individual's development. For example, the choice of loneliness, the rejection of all kinds of communications, isolation from the outside world with all its apparent pathology (symptoms of autism) are not always psychopathological phenomena. At those stages of personal development, when it comes to nominating goals or recruiting and ranking needs, and the desired ideas have not yet been formed, immersion into the world of their own experiences fenced off from the outside world limiting communication functions, a lack of real adaptation to the requirements of life with a tendency toward dreams and fantasies (i.e. thinking that costs without a constant influx of information from the outside) is a phase of development, although at a particular point in time it can be assessed as a stable personality trait that is introversion. In adolescents, signs of autism most often do not relate to pathological conditions. At the same time such symptoms in early childhood or in adults combine not only a reluctance to establish contact with others but also the inability to do this, and the ability to enjoy solitude cutting off all external information, refers to pathological manifestations.

The second important diagnostic criterion for assessing the standard options is the presence of criticism of the manifestations of a particular mental process. The productive symptoms a person is able to understand and describe should be remembered. Productive symptoms mean the addition of new signs or qualities that were previously absent. Therefore, there is the possibility of a critical attitude to the newly appeared mental changes. Negative symptoms (negative, deficient, that is, disappeared) in many cases are not amenable to critical analysis. A person does not notice or does not understand that he has lost certain qualities, since there is a so-called "syndrome steal of psyche" syndrome. It is assumed that along with actively unconscious symptoms such as delusional or overvalued ideas there are conditionally unconscious symptoms or not full consciousness, and not at any time (hallucinations, obsessive rituals, etc.). For example, the unconscious is abulia (a state of pathological lack of will in which the patient is not able to perform an action, the need for which is realized), and a perceived hypobulia; the same applies to couples autism – introversion; apathy – hypopathy, etc. From this it follows that it would be more logical to explain the manifestation of certain deviations in different terms referring to the psychopathological symptom or the accentuation of character defined as the variable norm of the psyche (the set of variable norms of the psyche determines the diversity of individuals). An example of this is the description of autism and introversion. If you read the definition of these two concepts that psychologists give, they are very similar. However, the mechanisms of the emergence of these two states are fundamentally different. By estimating a personality according to the sociability parameter it is possible to come to the wrong conclusion about the presence of signs of autism in the introvert. To avoid this, it is necessary to trace the development of signs in dynamics and remember that introversion is a variable norm which is associated with the individual (constitutional) level of development of certain brain structures and the features of their interaction as well as with the characteristics of education and self-actualization of a person. Introversion, unlike autism, does not occur suddenly, is capable of compensation under the influence of external factors and can be leveled in certain conditions and specific situations. Autism is a stable state that is not subject to significant fluctuations and is not dependent on environmental conditions. In young children, this condition is congenital, in adults it occurs suddenly. It is considered in the framework of deficiency symptoms and in fact it is a process of "robbing" a healthy psyche and causing damage to it. Therefore, in the psychological analysis of CAR, adaptive and pathological reactions, it is extremely important to assess the changes in the mental functions of a particular person over a long period of time taking into account the hereditary and ontogenetic factors of personality development.

In the psychological plan the pathological behavior differs from normal in the following parameters.

1. The tendency to generalization, i.e. the ability to manifest itself in a variety of situations and cause a variety of reasons.

2. The tendency to acquire the properties of a stereotype, i.e. to repeat the same, often inadequate behavior, for different reasons.

3. The tendency to exceed the limits of deviations from generally accepted behavior, never exceeded by the social group to which the subject belongs.

4. The tendency to lead to social disadaptation.

In general, it should be remembered that the presence of certain symptoms which are often referred to as psychopathological is not a sign of mental disorder but may be a normal variant if the symptom is accompanied by a critical attitude towards oneself and an adequate assessment of the environment when a person does not mix up unusual subjective feelings and fantasy with ordinary reality, and the behavior does not go beyond social norms.

For many diseases the development of homeostatic reactions is implemented in such a way that if there are already noticeable morphological changes in an SFWU the clinical signs of impaired function can be relatively scarce or absent since structural changes have a high degree of compensation which is provided by the intensification of regenerator-hyperplastic processes not only in damaged place but also nearby, and even at a distance from it. In this case the pathological process can remain compensated for a long time and clinically manifest only then when the phase becomes the decompensation of a disease, that is, when CAR has the maximum degree of voltage which is close to complete exhaustion.

Practically, this means that the first subjective and objective clinical manifestations of the disease should be treated very carefully, remembering that much more often they are signals not of the onset of the disease but in fact the already far gone decompensation phase. The scheme of development of a disease is represented in Fig. 23-1.

The beginning of the disease, on the one hand, can be violations of the processes of perception and the processing of afferent signals about the state of the musculoskeletal system, skin, vestibular, visual, auditory and other sensory systems (exteroception), the state of visceral organs, metabolic levels and the body’s needs for this moment (interoreception). On the other hand, the onset of the disease may be a violation of efferent regulatory processes aimed at the adequate completion of adaptive and compensatory responses to changes occurring in the internal and external environment. At the same time, signals from interoreceptors may not have an independent meaning until a certain point, but interacting with exteroceptive signals they become an important component of the body's analytical-synthetic activity.

Under normal conditions afferent and efferent impulses of the visceral sphere are a powerful factor of behavior acting as elements of various functional systems that provide vital activity.

Under the conditions of pathology visceral signals suppressing the individual reactions of all other signals can acquire a dominant value and determine the vectors of the organism's behavior. This category of pathological phenomena includes psychosomatic (or psycho-physiological) disorders whose development is accelerated by causes directly related to the behavioral or social factors surrounding the patient. The combination of these factors contributes to the onset of recurrent manifestations of certain symptoms: dysfunction of the gastrointestinal tract, certain forms of arthritis, chronic skin diseases, asthma, hypertension, etc. In such cases we are dealing with an exaggerated reaction of the organism, arising mainly with the participation of the ANS.

In order to quickly and fully restore somatic and psychosomatic functions various methods of targeting cortical and subcortical centers, afferent and efferent links of specific and nonspecific functional systems are used. Currently, traditional methods of pathogenetic therapy and pharmacology are being successfully used as well as new ways to eliminate some defects of sensory functions with the help of mechanical feedback devices that can replace faulty elements of functional systems or improve the elements necessary to optimize a particular activity. By incorporating elements reflecting the state of certain physiological parameters into the functional system of external feedback, positive results are obtained on the self-regulation of homeostasis in healthy and sick people which are then fixed in the process of training.

In these cases, by means of natural processes new functional connections are formed which are the basis of self-regulation necessary to achieve the desired results. These new functional relationships are supported by information about the degree of probability of satisfaction of a particular need and thus are included in the sphere of the subject’s emotional activity.

The achievements of the last decades in the field of psycho- and neurophysiology exploring the relationship between thinking, memory, learning ability and emotions, allowed the determination of the basis of these processes (the brain structures responsible for them); the mechanisms of interdependence between these structures and their functions; the direction and course of development of the nervous system's reactions to various external and internal influences and made it possible to assess the role and importance of the rational and emotional sphere of the organism's activity in adaptive behavior.

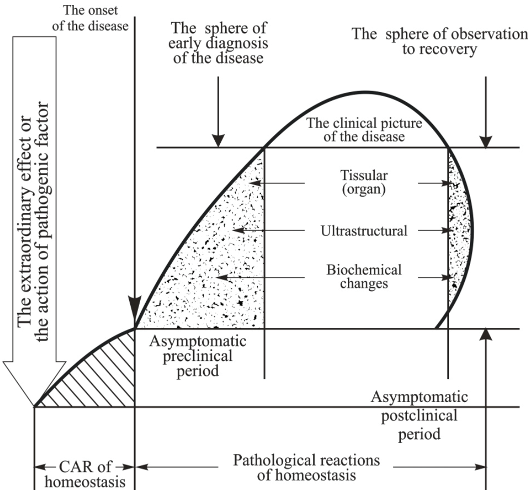


Fig. 23-1.
Scheme of development of the disease

It was found that a number of brain formations such as the cortical divisions of the sensory systems, the anterior and temporal portions of the cortex, the limbic cortex, the nucleus of the almond-shaped complex and the hypothalamus, are a neuroanatomical substrate for the emergence and realization of complex interrelated reactions, consciousness and memory, emotional reactions and somatic reactions and visceral organs in response to changing environmental conditions, which creates a holistic psycho-motor reaction of the body to any effect.

Very important data have been obtained on the neurochemical organization of the brain in experimental animals and in humans in clinical studies. They allowed the establishment of the same basic neurotransmitter systems: noradrenergic, serotonergic, dopaminergic, peptidergic (including endogenous opiates), and cholinergic, i.e. the very systems that the ANS uses to regulate homeostasis. These data well explain the mediator nature of the effector (motor and vegetative) manifestations of emotional reactions and allow us to include emotions in the context of organizing purposeful behavior and self-regulation of the body’s vital processes and even more broadly in the context of a higher order evolution – the biosphere.

From a physiological point of view emotion is the active state of specialized brain structures that induce changes in behavior in the direction of minimizing or maximizing any state. In general, an emotion resulting from higher neuropsychic activity is defined as a reflection by the brain of an actual need and an assessment of the likelihood of its satisfaction on the basis of genetic and previously acquired individual experience. Thus, emotion is a function of the need and information about the means and ways to satisfy it.

At the same time, for the organization of purposeful action an important step is not only the actualization of needs (activating the memory of the goals) but also the motivation – the physiological mechanism for activating the traces stored in memory (engrams) of those objects that must satisfy the existing need as well as an assessment of the probability of satisfying it. If the need has not become relevant and motivated or the probability of its satisfaction is very small, then there is no purposeful action. Due to the integration of the actualized and motivated need with a high probability of its satisfaction a positive emotion arises and the final stage of the organization of purposeful behavior is carried out. Since a positive emotion indicates the approach of satisfaction of need, and a negative one indicates distance from it the subject seeks to maximize (strengthen, prolong) the first state and minimize (weaken, interrupt, prevent) the second.

An important link in the development and consolidation of self-regulation skills is the formation of such a set and a hierarchy of needs that are most favorable for the realization of the goal and the realization of all the potential possibilities of the organism.

In particular, the will is of great importance for a need to overcome obstacles to the satisfaction of the dominant need that initiated the behavior. The independence of the will as a specific need is indicated by its ability to generate its own emotions in connection with overcoming or not overcoming the obstacle before the final goal is achieved.

Positive emotion creates additional reinforcement in developing the necessary skills and is a subconscious physician's assistant. It should be remembered that the subconscious is a kind of mental activity to which all that was conscious in certain conditions belongs, in particular well-automated and therefore ceased-to-be-realized skills. The essence of learning and follow-up training is to first teach the desired skills using positive emotions, and then to automate them through training to the subconscious level.

In the previous sections we introduced the reader to the basics of knowledge of the physiological mechanisms of adaptive, compensatory, and pathological reactions of the body. Even in animals many diseases are not determined by the congenital or acquired pathology of various organs but by the relationships that develop in the zoosocial group. For example, in one of the experimental groups of monkeys, the lead was seized by a male who became the leader thanks to the fact that he could loudly beat a stick on an empty iron barrel. After the barrel was taken away from him the group members ceased to recognize him as the leader and another young male took the lead, and the "overthrown king" developed a myocardial infarction from which he subsequently died.

According to the World Health Organization the number of psychosomatic diseases has increased many times lately. The reason for such a sharp jump is often seen in the characteristics of life of the population of industrialized countries in the negative consequences of the scientific and technological revolution. The idea has arisen of "informational neuroses" and even "informational pathology" of higher nervous activity of man. Recognizing the important role of these factors in the origin of chronic emotional stress, at the same time, it is difficult for us to agree with the view that the number of psychosomatic diseases is increasing as a direct consequence of the scientific and technological revolution. According to Boris Karvasarsky, the intensification of the production process as well as life itself, is not in itself pathogenic.

Neurasthenia due to overwork is an extremely rare occurrence. What is the cause of psychosomatic diseases of a person?

Even Pavlov tried to find this cause in the extreme stress of physiological reactions which is caused by the action of social conflicts. These conflicts (official, family, age, domestic) are of particular importance because when repeating they are fixed at the structural level determining pathological disorders in the interaction of brain structures. The results of experiments on animals with sequential or simultaneous damage to various brain structures suggest that the individual manifestations of the main varieties of human neuroses are determined by the interaction of the frontal cortex, temporal cortex and subcortical nuclei of the diencephalon and midbrain (the hypothalamus, an almond-shaped complex).

From the clinical experience of treating patients with a stable-changed state it is known that a group of pharmacological drugs with a therapeutic effect also has a normalizing emotiogenic effect. A single view of the mechanism of their action has not yet been developed but it is obvious that these compounds realize the effect on mental functions through neurotransmitter systems. Patients often have a competition of existing needs when the behavior is aimed at the preferential or even exceptional satisfaction of one of them. In these cases, the task of the doctor is to find ways and means to meet a particular need or to correctly switch the patient's aspiration to another replacing a need that can be satisfied in accessible ways. In this case it is important to evaluate and use the volitional qualities of the patient. In the psychological aspect, the will is an activity determined by the need to overcome obstacles independent and additional to the motive that organizes a behavioral act.

Unfulfilled fate, dramatic collisions and everyday turmoil sometimes stretching for years – these are the most typical situations that the doctor faces when talking to patients suffering from psychosomatic illness. We emphasize that contrary to the opinion of Freud, who saw the cause of these diseases almost exclusively in the disharmony of sexual relations the prevalence of sexual conflicts was observed only in 15% of cases in patients aged 19 to 50 years.

The clinic of neuroses does not practically meet with negative emotions arising on the basis of dissatisfaction with purely biological needs. As a rule, the emotional conflict of the neurotic leading to the development of somatic pathology is social in nature.

At present, it is generally accepted to define neuroses as mental diseases in the development of which an important role is played by the collision of emotionally saturated relations of a person with an unbearable life situation for him. The basis on which psychogenic somatic diseases are formed is the individual characteristics of the human nervous system and its interaction with other functional systems. Recall that the competition, at the same time important and often incompatible with other needs, is realized after the transformation of these needs into appropriate emotions, that is taking into account the probability (possibility) of their satisfaction in this particular situation. As noted, Alexander Vane neurosis is “a disease of unmet needs”.

Two factors are decisive for the emergence of neurosis and, after it, somatic psychogenic diseases: a situation of difficult choice, subjectively dependent on a person, and structural and functional features of his nervous system favoring a neurotic reaction. Neurosis does not occur if a situation does not arise that requires a backbreaking choice from a person, or if there is no previous (congenital or acquired) insufficiency of the corresponding functional brain systems. In a normal relationship between properly organized brain structures a person can always, replacing one need with another or choosing the most accessible way to satisfy his need, solve a particular situation without extreme emotional tension and breakdown in the disease. When there are pathological disorders of the interaction of brain structures manifestations occur of all known varieties of human neuroses. For the hysteria characteristic, for example, an overvalued idea, which occupies a dominant position in the life of the patient. The hysteric imposes on the environment his own version of the interpretation of external events. He is a pronounced “social egoist” who is fiercely demanding the attention of others, and in his actions, there is clearly a painfully hypertrophied social need “for oneself”. Hysteria or neurosis of obsessive states is given by the pathology of relationships in the system of the frontal cortex – the hypothalamus (the hypothalamic version of hysteria, or neurosis, in the case of a predominant defect of the frontal cortex).

Neurasthenia weakening volitional impulses is combined with acute sensitivity and irritability. Any unexpected event: a knock on the door, a phone call – can cause anxiety, palpitations, sweating, and muscle tremors. Leading to neurasthenia is violation of the functions of the temporal cortex, i.e. the almond-shaped complex which as a rule does not affect the higher intellectual functions, indicating a full-fledged activity of the structures of the new crust. Involvement of the anterior sections (frontal cortex) in the pathological process in combination with the impaired functioning of the amygdaloid complex leads to psycho-stenotic symptoms. A characteristic feature of psychasthenia is indecision, the inability to quickly make a decision and be guided by it (the pathology of the almond-shaped complex). This indecision is accompanied by mistrust, obsessive fears, and hypochondria. The last group of symptoms makes you think of a defect in the frontal regions of the cortex. The “biological egoism” of a psychasthenia focused on the smallest signs of its internal painful sensations differs from the “social egoism” of hysterics. Preoccupation with one's health, in which the whole world is obscured by the slightest signs of (sometimes non-existent!) diseases is nothing more than a hypertrophied biological need “for oneself” – the basis of hypochondria.

Another thing is guilt anxiety and despair at the thought that “I can’t do anything”. Here the chronic unmet need “for others” is already evident. Dissatisfaction with the needs of the development, advancement, and improvement of their life position causes depression or anguish. We emphasize that human needs are only partially realized and far from in accordance with their actual content. Only the signs of extreme emotional stress and the resulting compensatory and pathological reactions of the body are well manifested. In the later stages of the disease, structural signs of the neurogenic damage of various functional systems and their constituent tissues and organs are also highlighted.

Experimenters and clinicians have long known diseases associated with neurogenic dystrophies in the lung, gastrointestinal, cardiovascular, excretory and urinary systems: gastric ulcer and duodenal ulcer, coronary heart disease, myocardial infarction, trophic ulcers of the extremities and many others.

Pavlov said that in the case of a person it is necessary to find together with the patient or apart from him or even at his resistance, among the chaos of the vital relations those at once or slowly operating circumstances and conditions which can be connected with the right origin of a painful deviation.

Whether it is better perhaps, it is difficult to say, and we can only add that the primary task of the doctor is to find out the causes of the disease and arm the patient with the means to combat it. In addition to medication or the unconventional correction of impaired functions it is necessary to influence the patient's mind, show him the ways and means of satisfying this or that need or replacing it with another equivalent one in order to relieve emotional tension and thereby eliminate the pathogenic factor.

Conclusion

The problem of the evolution of science

There are more things in heaven
and earth, Horatio, than are dreamt
of in your philosophy...
William Shakespeare

In conclusion, let us dwell on the problem of the evolution of science itself. Of course, at first it would be necessary to consider the problem of how modern philosophy answers the question of the evolution of science, i.e. briefly touching on the history of positivism from Auguste Comte to Paul Feyerabend. The postpositivism of Karl Popper, the theory of Thomas Kuhn's paradigms and Michel Foucault's epistemes are extremely interesting in this respect but unfortunately the format of our book does not allow their inclusion. Therefore, we will assume that these questions are known to readers from the philosophy course.

For us, the greatest interest is in how scientists themselves are considering the evolution of science. The following material is an abbreviated and somewhat adapted version of a fragment of the article by Yuval Ne'eman (Ne'eman 1993, 70-90).

In the evolutionary theory of knowledge Popper and Donald Campbell formulated a kind of neo-Darwinian scheme that describes the evolution of science (Campbell 1974, 413-463).

If science is compared with a living organism then for its evolution there must exist its own "natural selection", its own "informational carriers", and its own "mutation mechanism". For the first category Popper used his idea of falsification. According to the principle of falsification no theory can be called a theory if it cannot be tested experimentally and if it does not turn out to be false in some of its predictions. Thus, in the evolutionary process of science falsification provides us with the necessary selection mechanism similar to Darwinian natural selection in biology. According to the third axiom of biology an appropriate random mutation mechanism is necessary. To find it Popper and Campbell introduced the hypothesis of blind variation according to which new ideas in science are born independently of the problem for which they exist. In addition, Popper and Campbell came to the conclusion that there is no a priori justification for any theory, and therefore the construction of a theory is nothing but a guessing, that is, a random operation. Guessing is based on intuition.

Later Ne'eman and Leonid Kantorovich proposed a modified theory of the mutational mechanism of science (Kantorovich and Ne'eman 1989, 505-529). In this theory there are three theses.

1. The key role in the mutational mechanism of science is played by luck. It is this phenomenon that is behind the main achievements in science either in theory or in the discovery of new phenomena or in the formulation of new problems that the researchers subsequently begin to deal with.

2. Luck must be applied to the revolutionary stages of science since in them the effect of chance is most pronounced. Sometimes the astounding results obtained “by chance” make the scientist himself such a stochastic component.

3. It is not the static structure of science that mutates but a dynamic research mechanism. Thus, there must be a constant search for “A” and then an accidental discovery of “B” may occur. This is analogous to the fact that for the fixation of a genetic mutation in DNA a process of reproduction is necessary, into which in turn error may also creep.

From our side we add that with such an approach science is likened to a living organism, in which the role of “information carriers” is fulfilled by the entire set of available scientific data, and the scientists and their research programs perform the role of SFWUs.

In the examples given by Ne'eman (Ne'eman 1993, 70-90) and also in some chapters of our book the following possible variants of discoveries were actually presented.

1. Research, but such research which is not a search for something concrete but nevertheless remains a search and which is completed as a result by the opening of “B”.

2. The search for “A” and the opening of “B” as a by-product.

3. The search for “A” and the opening of “B” instead of “A”. In some cases, “B” is “A” but for completely unexpected reasons.

New areas of science are particularly rich in discoveries which are based on luck. The initial database (the informational media of science) is usually limited here, and the theoretical knowledge with which it is necessary to coordinate unexpected results is less than that of the old theories. Therefore, mutations-luck in new sciences occurs much more often than in established doctrines. It is an analogue of the manifestation of a phenomenon that is well known in biological evolution and described by the model of the “punctuated equilibrium”. This is the model of Niles

Eldredge and Stephen Gould (Eldredge and Gould 1972; Gould and Eldredge 1993).

According to this theory, most of the evolutionary changes occur over short periods of time compared to much longer periods of evolutionary stability. Thus, evolution occurs much faster in small isolated populations where the chances of mutants for meeting and breeding are higher.

This is consistent with the fact that discovery occurs through randomness, i.e. fluctuation in small research teams since intelligence is a non-additive function. Then a new quality in a complex system arises through fluctuation and an increase in their number in the process of self-organization.

Let's stop on one more important moment. Man in his existence has apparently not undergone significant genetic changes (so far!). However, he has significantly evolved, self-organizing into communities.

It is noteworthy that until about the end of the XVII century technologies being a consequence of accumulated experience did not follow scientific advances but since the XVIII century new technologies are more and more the result of scientific progress. The emergence of new technologies directly affects the social evolution of mankind. Thus, first, some technologies and then science plus technologies become a kind of DNA (information carrier) of social evolution in which scientific and technological discoveries play the role of mutations.

Until recently it was believed that the fullness of knowledge and the necessary ability for intuitive judgment are fundamental conditions for the existence of homo sapiens. However, the development of artificial intelligence and especially self-learning systems shows that in the future, humanity will use computers not only as tools that enhance the capabilities of logical thinking but also possibly intuition. This is what defines the new paradigm of culture arising from the intellectual revolutionary end of the XX century and the beginning of the XXI century which frees mental activity from standardized routine work.

> The stream of Time, irresistible, ever moving, carries off and bears away all things that come to birth and plunges them into utter darkness, both deeds of no account and deeds which are mighty and worthy of commemoration; as the playwright says, it brings to light that which was unseen and shrouds from us that which was manifest. Nevertheless, the science of History is a great bulwark against this stream of Time; in a way it checks this irresistible flood, it holds in a tight grasp whatever it can seize floating on the surface and will not allow it to slip away into the depths of oblivion.

The Alexiad. By The Byzantine historian Anna Comnena. Translated by E. R. A. Sewter.

This thought, captured by the Byzantine princess Anna Comnena in "Alexiad" (XII century) stimulated the authors to write this "historical narration".

> Knowledge humbles the great, surprises the ordinary and inflates a little man.
>
> Lev Tolstoy.

BIBLIOGRAPHY

Arnold, Vladimir I. 1990. *Teoriya katastrof*. Moskva: Izdatel'stvo "Nauka."

Campbell, Donald. 1974. "Evolutionary epistemology". In *The Philosophy of Karl Popper*, edited by P. A. Schilpp vol. 1: 413-463. LaSalle, IL: The Open Court Publishing Co.

Eldredge, Niles, and S. J. Gould. 1972. "Punctuated equilibria: An alternative to phyletic gradualism." In: *Models in Paleobiology*, edited by T. J. Schopf, 82-115. San Francisco: Freeman.

Florenskiy, Pavel A. 1999. *Obratnaya perspektiva*. Soch. v 4-kh tt. T.3(1): 46-98. Moskva: Izdatel'stvo "Mysl."

Godik, Eduard E., and Yu.V. Gulyayev. 1990. "Fizicheskiye polya cheloveka i zhivotnykh." *V mire nauki* no. 5: 75-83.

Gould, Stephen J., and N. Eldredge. 1993. "Punctuated equilibrium comes of age." *Nature* 366, no. 18 (November): 223-227. https://doi.org/10.1038/366223a0.

Ignat'ev, Yuriy G. 2013. "Termodinamicheskoye ravnovesiye v uskorennoy vselennoy nedostizhimo?" *Prostranstvo, vrema i fundamental'nye vzaimodejstvia* no. 4: 28-55.

England, Jeremy L. 2015. "Dissipative adaptation in driven self-assembly." *Nature Nanotechnology* 10(11): 919-923. https://doi.org/10.1038/nnano.2015.250.

John, S. Chuang, O. Rivoire, and S. Leibler. 2009. "Simpson's Paradox in a Synthetic Microbial System." *Science* 323: 272-275.

Kantorovich, Aharon, and Y. Ne'eman. 1989. "Serendipity as a source of evolutionary progress in science." *Studies in History and Philosophy of Science Part* 20, Issue 4: 505-529. https://doi.org/10.1016/0039-3681(89)90021-6.

Khapachev, Yuriy P. et al. 1997. *Kontseptsii sovremennogo yestestvoznaniya*. 3-ye izd. pod red. Yu.P. Khapacheva. Nal'chik: Kabardino-Balkarskiy Universitet.

Khapachev, Yuriy P. 2000. "Fundamental'nyye konstanty khimii i biologii." *Rossiyskiy khimicheskiy zhurnal* (*Zhurnal Rossiyskogo khimicheskogo obshchestva im. D.I. Mendeleyeva*) 44, no. 3: 3-6.

Khapachev, Yuriy P., A.A. Dyshekov, and T.I. Oranova. 2013. *Sovremennaya yestestvenno-nauchnaya kartina mira. Kurs lektsiy*. Nal'chik: Kabardino-Balkarskiy Universitet.

Khapachev, Yuriy P. et al. 2017. "Sovremennaya yestestvenno-nauchnaya kartina mira. Kurs lektsiy. I-II chasti." *Aktual'nyye Voprosy Sovremennogo Yestestvoznaniya* 15: 4-127.

Khapachev, Yuriy P. et al. 2018. "Sovremennaya yestestvenno-nauchnaya kartina mira. Kurs lektsiy. III-V chasti." *Aktual'nyye Voprosy Sovremennogo Yestestvoznaniya* 16: 4-107.

Klimontovich, Yuriy L. 1996. "Relative ordering criteria in open systems." *Physics-Uspekhi* 39, issue 11: 1169-1179. http://dx.doi.org/10.1070/PU1996v039n11ABEH000181.

Kolmogoroff, Andrey N. 1933. "Sulla determinazione empirica di una legge di distribuzione." *Guornale dell' Instituto Italiano degli Attuari* 4, no 1: 83-91.

Kolmogorov, Andrey N. 1940. "Ob odnom novom podtverzhdenii zakonov Mendelja." *Doklady AN SSSR* 27, no 1: 38-42.

Kuznetsov, Vladimir I. 1973. *Dialektika razvitiya khimii*. Moskva: Izdatel'stvo "Nauka."

Landau, Lev D., and E.M. Lifshitz. 1976. *Mechanics*. Volume 1 of the *Course of Theoretical Physics* (3rd ed.). Butterworth-Heinemann. eBook ISBN: 9780080503479.

Lighthill, James. 1986. "The recognized failure of predictability in Newtonian dynamics." *Proceedings of the Royal Society* 407, no 1832: 35-50.

Markov, Aleksandr V. 2011. *Evolyutsiya cheloveka. Kniga pervaya. Obez'yany kosti i geny*. Moskva: Izdatel'stvo "Astrel'.", "Corpus."

Markov, Aleksandr V. 2014. *Rozhdeniye slozhnosti. Evolyutsionnaya biologiya segodnya. Neozhidannyye otkrytiya i novyye voprosy*. Moskva: Izdatel'stvo "Corpus."

Mednikov, Boris M. 1982. *Aksiomy biologii. Biologia axiomatica.* Moskva: Izdatel'stvo "Znaniye."

Ne'eman, Yuval. 1993. "Schastlivyy sluchay, nauka i obshchestvo. Evolyutsionnyy podkhod." *Put'* no. 4: 70-90.

Nicolis, Gregoire, and I. Prigogine. 1989. *Exploring Complexity: An Introduction*. New York: W. H. Freeman.

Popechitelev, Yevgeniy P. 1997. *Metody medico-biologicheskikh issledovanii. Sistemnyy aspekty. Uchebnoye posobiye*. Zhitomir: Izdatel'stvo "ZHITI."

Prigogine, Ilya, and I. Stengers. 1984. *Order Out of Chaos: Man's New Dialogue with Nature*. Bantam New Age Books.

Rudenko, Aleksandr P. 1969. *Teoriya samorazvitiya otkrytykh kataliticheskikh sistem*. Moskva: Izdatel'stvo "MGU."

Russell, Bertrand. 2008. *Mysticism and logic and other essays*. London: George Allen & Unwin Ltd., Ruskin House, Museum Street.

Simonov, Pavel V. 1986. *The Emotional Brain Physiology, Neuroanatomy, Psychology and Emotion*. New York and London: Plenum Press.

Simonov, Pavel V. 1991. *The Motivated Brain. A Neurophysiological. Analyses of Human Behavior*. Philadelphia: Gordon and Breach Science Publishers.

Tabor, Michael. 1989. *Chaos and Integrability in Nonlinear Dynamic. An Introduction*. New York/Chichester/Brisbane/Toronto/Singapore: John Wiley & Sons.

Zeki, Semir et al. 2014. "The experience of mathematical beauty and its neural correlates." *Front. Hum. Neurosci* 8, article 68: 1-12. https://doi.org/10.3389/fnhum.

Zhdanov, Yuriy A. 1977. "Uzlovoye ponyatiye sovremennoy teoreticheskoy khimii." *Voprosy filosofii* no. 1: 102-114.