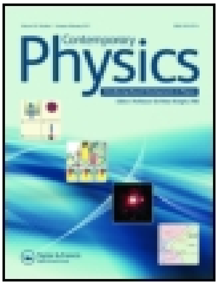

Contemporary Physics
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tcph20

On the psychology of scientific creativity

A. B. Migdal [a]

[a] L. D. Landau Institute of Theoretical Physics, USSR Academy of Sciences, Moscow

Available online: 20 Aug 2006

To cite this article: A. B. Migdal (1979): On the psychology of scientific creativity, Contemporary Physics, 20:2, 121-148

To link to this article: http://dx.doi.org/10.1080/00107517908219097

CONTEMP. PHYS., 1979, VOL. 20, NO. 2, 121–148

On the Psychology of Scientific Creativity

A. B. MIGDAL

L. D. Landau Institute of Theoretical Physics, USSR Academy of Sciences, Moscow.

ABSTRACT. This article discusses the motives that actuate the creative scientist, the dangers that beset him, and then, with numerous examples, discusses the mental processes involved in first-class research.

Introduction

Interest in the psychological side of scientific creativity has probably existed for as long as science itself. However, since this is not to be a literary essay, I shall leave out the inevitable anecdotes about absentmindedness, which begin with " Aristotle himself in his ' Organon ' . . . " and confine my remarks to the more recent past. In fact, during the last few years, interest in this question has suddenly blossomed and a large number of writings on the psychology and methodology of science have appeared. A new branch of knowledge has developed, the ' sociology of science '. How are we to account for this phenomenon? The answer is often to be found in the literature itself, where such explanations are given as: " Any nation today which is unable to appreciate the scientific mind is doomed ".

One of the distinctive features of the majority of these articles is their peculiar brand of specialization. They are, as a rule, written not by the people directly involved in science, but by experts on science and its structure, in the same way as essays on art are usually written not by actual artists, but by professional art critics.

This article is an exception. It reflects the thoughts and discussions not of experts on science, but of scientists themselves. It must therefore in no way be considered as a piece of research into the psychology and sociology of science: it is merely an attempt to share our own experience with the reader and to formulate the concepts which have accumulated in the course of work.

The questions I should like to consider first of all are: what makes a young man want to be a scientist? What personal qualities does the work demand?

1. Motives for scientific creativity

> Not for you are passion and goldlust,
> It is science that entices you.
>
> Passion may fade and love is betrayed
> But you cannot be deceived
> By the bewitching structure of the cockroach.
>
> N. OLENNIKOV, *Comic Verses*.

1.1. *Curiosity, self-expression, self-confirmation*

Let us concentrate on the motives which have a specific bearing on scientific creativity, omitting such factors as the social usefulness of the work, which have a similar influence in any field of human activity. (We shall not, of course, take into consideration such motives as the desire for a successful career or to attain material wellbeing.) Normally, these consist of a combination of the following elements. That most alien to the true scientific spirit is the desire for self-confirmation, the need to demonstrate to oneself and others that one can carry a task to its conclusion. Another element is the need for self-expression, that is, the search for a fuller expression of one's own individuality. But the most important motivating force is curiosity: the desire to know how Nature is composed. Where this force predominates, a man will take as much delight in a new discovery of someone else's as in his own. This was precisely the attitude to science of the brilliant Russian theoretical physicist, I. Ya. Pomyeranchuk, who, even on his deathbed, on regaining consciousness, asked about the latest developments in the theory of elementary particles and rejoiced in each new idea.

Such absolute purity of motive is very rarely found and is not a necessary condition. Usually all three motives are present in different proportions. Sometimes the need for self-expression is so strong that the pursuits of science alone are not enough to satisfy it. It is well known, for example, that Max Planck was an accomplished pianist, Einstein played the violin and Richard Feynman plays the bongos. Zeldovich the well-known Russian physicist writes poetry (" You are looking for explanations—know your atomic structure! ").

There are some people for whom the desire for self-confirmation is the chief incentive to creative work. Where this desire is not kept in check by impeccable conscientiousness, it almost inevitably turns into a search for positive results which leads to involuntary misinterpretation of the facts. How many talented minds have been lost to science through this fault!

Among non-scientists there is a widespread assumption that the scientist is guided in his work by the desire to make a discovery. But, on the contrary, this must not be his aim; his task is to make a deep and detailed study of the scientific field in which he is working. A discovery comes about only as a by-product of this study. This does not mean that scientists are by nature so disinterested that they do not wish to make a discovery. The wish is bound to be there, of course, but in the background: it not only should not determine, but should not even have the slightest influence on the way their research is conducted.

When I use discovery in this sense, I mean a significant leap forward in the understanding of Nature. Small discoveries which go unnoticed by the rest of the world are made constantly and it is these that make the day-to-day work of the scientist worth while. Any would-be scientist must have this ability to delight in every small step forward, every tiny discovery. It is important that newcomers to the profession should clearly understand the dangers of looking for self-confirmation in their work and of the search for positive results to which it so frequently gives rise and we shall therefore return to this subject more than once.

1.2. *The capacity for wonder; the concept of beauty in science*

For the moment, let us return to the researcher's curiosity. This is intimately bound up with his capacity for wonder, which is an essential quality in any field of creativity: no true poet, artist or scientist is without it. But whereas in art the direct, spontaneous reaction to what we see or hear is all-important, in science, wonder must be the result of thought applied to these sense-images, of the interpretation of accumulated knowledge. When the source of this wonder is clearly expressed, is it said to be a scientific paradox.

There are many well-known instances in which the scientist's wonder, formulated as a paradox, has led to a scientific revolution. Here are a few examples.

Gravity theory

Ever since Galileo, it has been known that all bodies fall with equal acceleration (if you do not take into account the resistance of the air). This means that the weight of the body, that is, the force with which the body is drawn towards the ground, is directly proportional to its mass. Also proportional to the mass is the force of inertia. Because these two forces depend in the same way on the mass, a man in a freely falling chamber will find himself in a state of weightlessness: the force of inertia exactly counterbalances the force of gravity.

We have grown so accustomed to this truth that we do not see anything remarkable about it. To Einstein, it was a marvel and he expressed his wonder in the form of a paradox: why should the *weight* of a body, regardless of its composition, and its *inertia*, both be in proportion to its mass? Does it not follow from this that there must be a deeper connection between the force of inertia and that of gravity? This idea was the starting point for one of the most remarkable of physical theories, Einstein's theory of gravity.

Why do the stars give so little light?

Let us assume that the stars are distributed more or less evenly over the Universe. Thus, the number of stars lying within a sphere with radius R, surrounding the Earth, would be proportional to R^3. The intensity of illumination received from each individual star follows the inverse-square law, and so the total effect should be proportional to $1/R^2$. Consequently, the full intensity of light from the stars inside the sphere would be proportional to R and, if the universe were infinite, then the brightness of the sky would only be limited by the negligible amount of light swallowed up in interstellar space. The sky should therefore shine " more brightly than a thousand Suns ". This paradox (Olbers' paradox) had been known for a long time but no one had been able to explain it, until it was finally resolved in Einstein's cosmology, which comprises an unprecedentedly bold application of the theory of gravity to the world as a whole. This theory states that the geometry of the world differs from the usual Euclidean geometry, that light rays do not travel in a straight line, that the Universe expands and only the light of a comparatively small number of stars reaches the Earth.

Let us take an analogy, rather inexact, as analogies tend to be. Instead of the three-dimensional world, let us look at a two-dimensional world on the surface of a sphere. In such a world, the light would shine along the shortest distance between two points. If the surface of the Universe were evenly covered with stars, then the number of stars as a whole and the number of stars lighting up some object would, of course, be finite. Now imagine (although even this is quite difficult) our three-dimensional world in the form of a sphere in four-dimensional space, i.e. imagine that it is enclosed within itself like a two-dimensional world on the surface of a three-dimensional sphere. Then the number of stars as a whole would be finite and the dimness of the night sky is explained. Now you can imagine another sort of geometry, with the number of stars as a whole infinite, but the number of stars shedding light on the Earth, finite. The question as to what the geometry of our world is, exactly, cannot be decided speculatively; it is a question for experiment to decide.

The leap forward in the understanding of our Universe represented by the advance from the flat space of Euclidean geometry to curved space is equivalent to the step forward that was made when men realized that the Earth was not flat, but round.

Quantum physics

A major paradox became evident at the end of the 19th century, when attempting to apply the laws of statistical physics to an unaccustomed subject: the standing electromagnetic waves which might arise in a box with reflecting walls. According to these laws, each independent mode of oscillation in the thermal equilibrium resulting from repeated radiation from and absorption into the walls should carry an energy kT, where T is the absolute temperature of the walls and k the Boltzmann constant. But the number of standing electromagnetic waves in the box is infinite. In fact, standing waves can be formed in the box if a whole number of half-waves is confined between the walls. The shorter the wavelength, the greater the number of possible directions for which this condition is fulfilled. Thus, the shorter the wavelength, the larger the number of possible oscillations and the greater the energy at that wavelength. Consequently, the electromagnetic field should absorb all the thermal energy of the walls, however much heat were applied to it. If each oscillation really possessed an energy kT, then, if we made a little hole in the box, we should have a source of incomparably bright light of very short wavelength. This paradox was given the dramatic name of ' the Rayleigh–Jeans catastrophe ', although in the experiment itself no catastrophe takes place.

To explain this contradiction, Max Planck suggested that electromagnetic oscillations change their energy in steps $\Delta E = \hbar\omega$, where $\hbar$ is a coefficient of proportionality and ω the angular frequency. If the minimum possible energy of oscillation $\hbar\omega$ is much greater than kT, then the oscillation will have a small intensity. According to the laws of statistical physics, the intensity of such an oscillation falls with the increase in the frequency, according to an exponential law. Thus, high-frequency oscillations make a small contribution to the thermal energy, and the paradox is resolved. This law is borne out with great exactitude in experiment, and the value of $\hbar$ can be determined.

This is how Planck's constant $\hbar$ first came into physics, specifying the possible discrete energy levels of electromagnetic waves of a given frequency, and how the concept of discontinuous processes came into being. The minimum unit of energy of an electromagentic oscillation was called a quantum. If the energy of the oscillation contains n units of $\hbar\omega$, then it is said that there are n quanta of angular frequency ω in the box.

Why do atoms radiate light not at all frequencies but only at discrete, precisely defined frequencies? If the electrons in an atom moved according to the laws of classical mechanics, they would give out light at all frequencies. Does not this mean that electrons in an atom, like electromagnetic oscillations, might possess not just any energy level, but only certain, strictly defined energies? By reflecting upon these and other problems, Niels Bohr was finally led to create a quantum model of the atom.

The universality of the conservation laws

Why is the law of conservation of energy equally valid for the most diverse processes: electromagnetic, mechanical, thermal, chemical, and so on ...? Surely this must mean that it is based on some deeper common property of the laws of Nature. Attempts to answer this question did not lead to a scientific revolution, but they did deepen out understanding of the world.

It was, in fact, found that the universality of the law of conservation of energy, like the other conservation laws (for example, the law of conservation of momentum), is connected with the general properties of space and time. It is possible to see that the law of conservation of energy is a consequence of time independence. Time independence means that an apparatus will work in exactly the same way at all times provided there is no change in the external conditions affecting the apparatus. To demonstrate the connection between the law of conservation of energy and time independence, we can argue that this law can be broken if time passes irregularly. Let the irregularity of time consist in the fact that gravity is not constant but varies from time to time. Then you would gain more energy than is spent: you need only pick up a weight during a period of weak gravity and let it fall in a period of strong gravity, thus turning a dynamo.

There are many similar examples in the physics of elementary particles. Frequently, a new particle has been discovered on resolving a paradox which arose from the need to explain some phenomenon. There is, for example, one paradox, as yet unresolved, which never ceases to amaze physicists: why is the charge on the proton always equal in magnitude to the charge on the electron, in spite of the fact that these two particles are totally dissimilar in all their properties? The explanation of this paradox will seriously influence the choice made between different possible theories describing the interaction of elementary particles with the electromagnetic field.

Super-thermal conductivity or superfluidity?

Yet another example is the discovery of superfluidity, made in the Institute of Physical Problems of the USSR Academy of Sciences in 1937 by Academician P. L. Kapitza. They were researching into the properties of liquid helium at very low temperatures. It was known that at temperatures of below 2·2 K ($-270{\cdot}8°$ C), liquid helium undergoes a further modification, becoming helium II, with completely different properties. It had already been found by the Dutch physicist Keesom in Leiden that helium II possesses a thermal conductivity a million times greater than that of copper, which, in itself, is very strange. Then it was found that helium II also had an abnormally low viscosity—a thousand times lower than that of water. And yet in fluids the microscopic mechanisms of thermal conductivity and viscosity are very similar and, where thermal conductivity is high, viscosity is nearly always high too. In fact, thermal conductivity is determined by the rate of transfer of kinetic energy from one layer to another, while viscosity is determined by the rate of transfer of momentum. The greater the one, the greater should be the other. But with the helium, the reverse was taking place.

After giving this paradox some thought, P. L. Kapitza came to the conclusion that there is no such thing as 'Super-thermal conductivity', but that the high thermal exchange rate observed by Keesom is caused by currents arising in the helium when it is in a state of *superfluidity*. In a state of superfluidity, the liquid helium can pass through a tube without any friction. Therefore, the very slightest unevenness in density, due to differences in temperature, is enough to cause currents, due to gravity, which carry the heat with them.

P. L. Kapitza was then obliged to carry out a dozen or so extremely delicate experiments, to convert this idea into an authentic truth.

The reader is no doubt aware that physicists are divided into experimentalists and theoreticians. It is very rare for one person to combine both. There is too great a difference in the nature of the specialist knowledge and the skills required by the two occupations. The first-class experimentalist, P. L. Kapitza, referred his experiments to the first-class theoretician, L. D. Landau. Theory and experiment stimulated one another. It was through this interaction that Landau gave birth to one of his very best theories, the theory of liquid helium II. This theory succeeded in making all the facts outlined in P. L. Kapitza's experiments quantifiable.

From these examples, it is easy to see the leading role that the capacity for wonder plays in science. Even more important, they give some idea of the beauty of science. Simple facts which do not, at first sight, have anything remarkable about them, when subjected to profound thought, can give rise to unexpected and important results. The dimness of the night sky can cause us to reappraise all our views on the geometry of the world. Phenomena as diverse as the law of conservation of energy and time independence can turn out to be intimately related. Laws that have been discovered when studying the movements of atoms in a heated gas are found to apply to the electromagnetic field, thus leading to the conclusion, quite alien to traditional mechanics, that the energy of an electromagnetic oscillation can change only in discrete units.

The logical interdependence of all the findings of science was expressed by the German mathematician, David Hilbert, like this: "You have only to accept that twice two are five, and I can prove to you that witches fly out of chimneys". The beauty of science lies in the shapeliness of its logical structure, its richness in interconnections. The concept of beauty can prove invaluable in checking results and in discovering new laws; it is the reflection in our consciousness of the harmony which exists in nature.

Anyone who chooses the scientific profession should be motivated above all by an appreciation of the beauty of science and by a sense of wonder. In his book, *Science and Method*, the prominent mathematician, Henri Poincaré, speaks of the beauty, "for whose sake the scientist undertakes such long and arduous tasks" and says, "I am thinking of that more profound beauty which consists in the harmony of the parts and is perceived only by the intellect. It is this that underlies and forms the basis for that play of visible beauties which delights our senses. Without its support, the beauty of fleeting impressions would be incomplete, like all that is vague and transient. Intellectual beauty, on the other hand, gives satisfaction in itself . . . ".

2. Underwater reefs

Niels Bohr once said, " A specialist is someone who knows some of the commonest mistakes in his field and is clever enough to avoid them ". Scientific research is fraught with psychological pitfalls. Let us try to analyse some of the most dangerous.

2.1. *Beneficial criticism*

In the very early stages of a piece of research, there is sometimes a real necessity to fan the faltering flame by finding arguments to confirm your point of view. But, as soon as the work has started to take shape, it can only do harm to keep reassuring oneself, and the greatest difficulty is to find opposing arguments; arguments ' pro ' seem to come forward quite of their own accord without any conscious effort.

The obligation to make a discovery quite often leads one to take refuge in reassuring arguments and even involuntarily to twist the facts. Here is one example of an instance in which a small lack of conscientiousness in dealing with experimental data snowballed to such an extent that a completely wrong conclusion was reached. The experiment was an attempt to investigate the energy spectrum of the alpha particles emitted from a certain nucleus. This spectrum consists of sharp peaks, and the distance between the values of the energy at these peaks gives the possible values of the excitation energy of the nucleus resulting from the alpha-decay process (' daughter ' nucleus). In the experiment, the alpha particles had fallen into groups with energies at equally-spaced intervals. This meant that the intervals between energy levels of the daughter nucleus were also equal. This result was quite unexpected and contradicted existing ideas about the structure of the nucleus.

The experimentalists asked the theoreticians to give an explanation. It was one of the rare cases where theoreticians can be proud to say that they were *not* able to construct a theory. On further experiment, the results were not repeated. It turned out that, when they had first started to measure, they happened to obtain curves with evenly spaced energies for the alpha particles. The experimentalists were so excited by this unusual result that, every time it was not repeated, they checked the voltage in the circuit and if the voltage differed from the standard, they threw out the results of their measurements. This check was only carried out when they obtained an unwanted result. Owing to the very large number of measurements that were made, this small discrimination led to almost exactly evenly spaced values for the energies of the alpha particles being produced. This happened in the laboratory of an experimental physicist who had won a high reputation for the conscientiousness of his work. However, on this occasion, he had lost control over the actions of his less experienced colleagues. There are no two kinds of conscientiousness: it is either irreproachable or it does not exist at all. It is like Voland's conversation with the barman in Bulgakov's book, *The Master and Margarita*, where he says, " there is only one sort of freshness—its first bloom is also its last ".

2.2. *The signs of the " Great Discovery "*

Efforts to make a discovery at all costs and to achieve a revolution in scientific thinking often break the bounds of their own possibilities and, at times, bring their authors to a sad end. It is well known that, in all physical institutes, some of the staff are always ' on duty ' writing answers to the authors of ' great discoveries '. All these works have the following things in common:

(1) They do not confine themselves to one question alone, but abrogate all the findings of contemporary science at once.

(2) The author has no specialist training in the subject in question.

(3) They never quote contemporary scientific literature, more often than not because the author is not familiar with it.

(4) The authors of such ' discoveries ' always claim that their work is the fruit of long years of labour, while it is patently obvious that if any time has been spent on it at all, it was not on the mathematical layout, nor on the experiments, nor even on the analysis of known facts.

(5) The author has not previously published any other smaller scale works.

These are the signs by which ' The Founder of a New System ' or ' The Inventor of a complete New Basis ' (the terms invented by Wolfgang Pauli to describe these crank scientists) can easily be recognized, regardless of the details they contain. An authentic revolutionary discovery directly concerns only a very narrow range of phenomena and is soundly based on the accepted findings of science in every other field. Modern science is so specialized that it demands a huge store of technical knowledge, much more than even a specialist training can give. It can only be gained through long, persistent and conscientious work in the field.

Unfortunately, ' discoveries ' of this kind sometimes obtain the support of people with scholarly titles and get published in the form of papers or books. These supporters of ' scientific sensations ', in spite of their degrees and qualifications, have as little to do with science as the authors of the discoveries themselves. Editors of publications of this kind who lack the necessary scientific qualifications would do well to be guided by this list of signs of the ' great discovery ' when judging the articles that come to their notice.

2.3. *Superstitions*

When the statistics of experimental data are not properly worked out, it inevitably leads to mistakes and the creation of superstitions.

At least once in every lifetime, incidents arise, which, it seems, can only be explained by telepathy. And yet, to this day, no-one has managed to produce a satisfactory proof that it exists. Even after years of research, there is not a single experiment which has achieved statistically convincing repeated results. It is therefore not merely pedantry to conclude that the existence of this phenomenon is, to say the least, doubtful. Arguments of personal experience and personal conviction count for nothing: remember that 100–200 years ago there were many people who claimed to have seen angels and devils. Even fairly recently, many people have communicated with spirits at spiritualist seances. On the other hand, the same scientific integrity we have already mentioned does not allow us to assert that telepathy does not exist. It cannot be proved that a phenomenon is absent, one can only state that no evidence has been found of its existence and that therefore the latter is improbable.

One of the most difficult aspects of scientific research is the transition from guesswork to authentic scientific fact. To reach the truth, the scientist must advance painfully, step by step, like a mountaineer scaling a sheer rock face.

This is why anyone who has anything to do with science was offended by the film, *Recollections of the Future*. This film is an example of the active propaganda of false science. With incomparable ease, it deduced from facts that might have had a thousand simple explanations that there were traces of astronauts from other planets having landed on the Earth. This is roughly the way they managed to twist the facts: if, in an ancient picture, a man was carrying a jug on his head, then he must be an astronaut; if he has not got a jug on his head, then it must have fallen off as his space-ship was landing. The authors never question why an astronaut from another planet should necessarily resemble our astronauts, why they should wear space suits like ours, etc., but this is not the worst. They did not understand, or pretended not to understand that there is a huge, insurmountable distance between a guess, even a plausible one, and an authentically proven truth, which systematically eliminates every possible explanation other than the one it puts forward.

There are several examples in the history of physics of ' superstitions ', that is, widespread delusions which have come to exist without proper foundation. Such was the idea that thermal energy was like a sort of fluid (caloric) running from a heated body to a cold one, or the nineteenth-century concept of a luminiferous ether, which filled all space.

In the twentieth century, delusions of this kind, if they arise at all, are shortlived. However, the words of the eighteenth century German physicist and philosopher, Lichtenberg, are still relevant today, " It is not the gross delusions, but the trifling untrue theories which obstruct the revelation of scientific truth ".

2.4. *Which should come first: understanding or research?*

There seems to be a vicious circle which it is impossible to break: you cannot accomplish a piece of scientific research without clear understanding, but clear understanding comes only at the end of the research.

This contradiction is one of the difficulties of scientific research. However, every piece of work that reaches completion must have overcome this paradox somehow. It does not generally happen all at once: as understanding grows, so the work can move forward a little further, which, in turn, brings new understanding.

Often, when starting on a piece of work, one leaves aside certain problems or queries which must be decided in due course but which, for the time being do not interfere with its progress. Occasionally, if you lose the list of queries among other papers, when it eventually turns up, you find that practically all the points that needed clarifying have resolved themselves while you have been working on the basic problem. Understanding in science is much like understanding a friend—it can only be achieved through long acquaintance.

The need to understand everything thoroughly before one starts to work is a very common cause of failure. However, some people are, by nature, incapable of feeling their way in the dark, of working without full understanding. This type of scientist is extraordinarily useful when it comes to assessing other people's work. It is difficult to appreciate just how much they contribute to the development of science: it is far more than one would think from studying their works, however valuable these might be. There was one very fine physicist who had this gift for deep understanding, the late Professor I. M. Shmushkevitch. Everyone who knew him tried to get his opinion on their work as soon as it was finished or even half-way through. The work was then said to have " passed Shmushkevitch ". His perusal was bound to bring all the doubtful or poorly thought out passages to light; and if the work had the good fortune to pass without comment, then this would mean that everything was in order.

This dislike for working in the dark sometimes manifests itself, more covertly, as a wish only to undertake ' authentic ' studies. Any work in which unsubstantiated but plausible assumptions are made is discarded as being ' unauthentic '. This quality tends to hamper the productivity even of first class physicists. Einstein, in his obituary to P. Ehrenfest, a most profound physicist, wrote, " His misfortune was that his critical faculties always got in the way of his creative powers ". Even such a remarkable physicist as W. Pauli was held back by this same weakness.

2.5. " *The service of the Muses abhors vanity* "

At the opposite extreme from the desire to understand everything before you begin, lies the impulse to ' jump the gun ', i.e. guess the result, leaving out the process of understanding altogether. We shall provisionally call this characteristic ' child prodigy-ism '. The instruction or self-instruction of the scientific researcher should begin with the elimination of all traces of ' child prodigy-ism '. It is interesting to note that L. D. Landau, who was distinguished as much by the impressive speed of his thought as by the impressive breadth of his grasp of all fields of physics never permitted himself even the slightest hint of this, but always did his utmost to bring the question in hand to a point of absolute clarity and extreme simplicity. He used to say jokingly, " I am the world's greatest genius at trivialization ".

Remarkable though it may seem, it is quite true to say that the more profound a scientific idea, the more it gains from simplification. In art it is quite the reverse: the finished work cannot be simplified—any attempt to simplify it destroys the essence. The words, " Boy, fill my cup with the heady bitterness of Falernia! ", when simplified, might give, " Bring me some more wine, boy! ". You can analyse the different elements which constitute the spell of a work of art, but the image that the work evokes cannot be broken down into its constituent parts, but must be taken as a whole. In science, anything can be broken down into constituent parts.

In order to understand a great work of art, you have to raise your consciousness to its level, whereas the achievements of science can be brought down to a lower level, ' laymanized '. The simplification of a scientific theory demands just about as much creative effort as its origination. This is why many of the profounder books of popular science, written by prominent scientists, give no less impetus to the development of science than the originals on which they are based. These books sometimes demand a great effort on the part of the reader, but, as against that, they do not avoid the difficulties and their simplifications do not descend to the level of vulgarization.

There is no room for haste and bustle in scientific work, but, on the other hand, lazy work not only takes up a lot of time, but wastes it too. However, this applies to any form of human activity.

There is yet another psychological characteristic which interferes with any type of creative work, and that is a belief in one's own infallibility. Of course, it is impossible to accomplish anything serious without a certain amount of confidence in one's own ability. But believing that one is infallible means that the scientist, having once set off in what seems to him the most probable direction, will stubbornly persevere, even when he comes up against a brick wall. One must follow the golden mean between self-confidence and doubt, hesitation and intransigence, flexibility and firmness.

2.6. *Unscientific questions*

Often, work is held up by the consideration of questions which are either unscientific or lie outside the boundaries of science. I do not mean questions so obviously unscientific as the sort of controversies that were carried on in the middle ages, ' how many angels could you fit on the point of a needle? ', etc., nor the many instances in which it is a point of terminology and not a reality which is under discussion. I am questioning the scientific value of such statements as, " There is another world which coexists with ours, but of which we are not aware because it does not interact with ours ". Clearly this statement lies outside the boundaries of science, since there is no way of ascertaining whether or not it is true. Here is another example of a question which contradicts the logic of science: " Is it permitted to doubt the laws of, for example, quantum mechanics? " Of course there is no truth which cannot be doubted, but one should not do this without sufficient grounds. Similarly, you do not lose faith in a well tried friend unless some circumstance arises to make you change your opinion of him. Without a certain respect for well-established laws, science would not be able to develop at all. Quantum mechanics and the theory of relativity are the most frequent targets for unscientific criticism, more often than not in the course of attempts to produce an alternative explanation for phenomena which are already predicted and explained by existing theories. So long as no experiments are proposed to demonstrate the correctness of the new theory or to disprove the old, then this view has no relation to science and, at best, can only have pedagogical value.

There is one universal criterion for distinguishing between scientific and unscientific questions. Any statement which cannot, even in principle, be tested, is unscientific. This criterion is called the principle of observability. It does not necessarily involve an actual test, only the theoretical possibility of carrying one out. It can be applied even to theories which do not describe our world and yet are logically tenable (as, for example, Lobachevsky's geometry). The theory is scientific if what follows from it can be imaginarily tested, using figurative experiments within the context of the imaginary world it describes, or, in short, if it leads to definite relationships between the quantities which enter into it.

One very good illustration is the concept of a God. If God is seen as a spiritual substance which has no influence on the laws of Nature, then his existence does not take the form of observable relations between different quantities and, consequently, such a God, according to the principle of observability, is an unscientific concept. On the other hand, if God is a material force, which has an influence upon the laws of Nature, then his existence is a question which science can decide. The scientist can only say that there are no experimental data which need to be explained by such a concept: all the known laws of Nature have been satisfactorily explained without introducing the concept of a force outside it.

Very often, queries which arise in the early stages of the work disappear or are cleared up by pronouncing the magic words, " formulate the question in terms of observable quantities ".

The principle of observability led to a brilliant finding at the beginning of the twentieth century. Before Einstein, the concept of the simultaneity of two events was understood intuitively. Einstein, basing his assumptions on the constancy of the speed of light, suggested a very simple method for testing the simultaneity of two events. Two flashes of light at points A and B can be considered simultaneous if the light from them simultaneously reaches a point lying in the centre between A and B. From this definition, it immediately follows that events which are simultaneous to a stationary observer are not simultaneous to an observer moving relative to points A and B. And in turn, it follows from this that the moving observer experiences time differently from the unmoving observer and therefore that time is a relative concept. Thus, in effect, the whole theory of relativity was arrived at through the consistent application of the principle of observability to the concepts of space and time.

Niels Bohr and Werner Heisenberg submitted such concepts as the position and momentum of particles to the test of observability. They found that in principle it was impossible to make exact measurements of position and momentum simultaneously. The more exactly you measure the position, the more indefinite become the readings for the momentum and vice versa. It was in this way that the uncertainty principle came into being. It was this relationship which formed the basis of a new mechanics, quantum mechanics, which took the place of traditional mechanics as applied to small objects.

2.7. *Ageing in scientists*

There is yet another danger which threatens the scientist and that is ' growing old '. I have put these words in inverted commas as I am not referring to literal old age. This type of ageing begins imperceptibly: it is very tempting to delegate routine work to younger colleagues, so as to leave oneself more time for important scientific business. Bit by bit, even calculations and some of the thinking are delegated. This cannot be done, just as you cannot maintain a relationship with a person you love through an intermediary. As soon as the scientist ceases to do his work for himself, make his own measurements if he is an experimentalist or do his own calculations if he is a theoretician, ' old age ' sets in, regardless of his years or qualifications. The capacity for wonder, the joy in each small step forward are lost, the desire to study disappears, conceit and self-importance appear and he starts to want only to solve problems of worldwide significance. Suddenly, he seems to produce works for publication faster than ever before; he acquires an inflated sense of the importance of his own opinions, a belief in their infallibility; he starts to think that it is enough to spend half an hour a week in the vicinity of each piece of apparatus in order to become the co-author of a work. Perhaps I should qualify this: in some instances, the opinion of a qualified and experienced person can have a decisive influence on the course of a piece of work. Sometimes the advice given can turn out to be so valuable that it does give the right to co-authorship. But, with this one exception, to take part in a large number of publications at once is usually a sign that one must be on one's guard, for often the scientist who does this not only does not command respect, but he lays himself open to ridicule. How do you explain this to the sufferer himself? Perhaps that unfortunate question that always appears on questionnaires, " how many scientific papers have you published? " should be dropped and replaced by, " what original results have you obtained? " " what problems have been solved as a result of your work? ", or, if it really is necessary to quantify, " how many references have been made to your works? ".

Besides the proliferation of scientific journals, unbridled writing creates an unhealthy atmosphere of cheap success which is quite alien to the purposes of science. One's sense of responsibility starts to be eroded; when writing an article, one ceases to weigh every word one writes in fear of making a misleading statement. One is reassured by the thought that, despite the errors, a mistaken theory can often point the way to a true one, etc. . . . Bit by bit, the scientific content gives way to observations of a general nature, the proportion of descriptive writing in the article increases and the number of formulae decreases. The ageing scientist tries to compensate for his lack of new ideas by making witty allusions. When, from time to time, he tries to return to true scientific writing, then his works, even those that seem to him original, have one thing in common: they are not new works at all but criticisms of the works of others. We can all think of examples of people who have ended up like this. This sort of activity is no substitute for the joy of true scientific research, and it nearly always gives rise to a deep, sometimes concealed feeling of dissatisfaction. Such is the price of neglecting scientific work.

On the other hand, the scientist who loves his work can go on producing original results to the end of his days, although there are many who think otherwise, I would maintain, wrongly. In V. Polinin's book on genetics, called " Mummy, Daddy and Me " he says, " Science is capricious, she loves the young . . . her preference is for the light-headed muddler, but she is obsessed with the soul of the rebel and the revolutionary ". Surely, it is self-evident that a man with these characteristics could never, at any age, accomplish anything even mildly useful in the interests of science, let alone anything remarkable.

It seems to me that success in science is not a matter of age, but of a certain sort of ability linked with a certain psychological disposition. These things do not improve with age, but they do not deteriorate either. However, this being so, how are we to deal with, what is undoubtedly true, that almost all important scientific discoveries have been made by young people? It is sometimes concluded from this statistical truth that it is only possible to do significant work in mathematics or theoretical physics up to the age of thirty. But here we are dealing with a very widespread source of fallacies, incorrect analysis of the statistical data. It would be easy to demonstrate the error of this assertion using examples, but instead, let us try to understand the reasons behind this statistical pattern. First of all, the statistics merely indicate a correlative (accompanying) connection between age and scientific success; it by no means follows from this that the connection is inevitable and springs from the nature of the work itself. Besides, the statistics are distorted by the fact that many scientists ' drop out ' for personal reasons that tend to (but do not necessarily) coincide with middle age: family responsibilities, illness, complacency. Scientific research is demanding work and there are many people who cannot stand the pace and change to lighter occupations.

One real and serious (though not insurmountable) difficulty lies in the fact that a scientist is obliged to change his system of views, his style of work and sometimes even his own mentality with each major discovery. This is sometimes easier for a younger man who is not overburdened with his own established ideas. However, the habitual flexibility of ideas which grows with experience in research can compensate for this beginner's advantage. In any case, the capacity to accept what is new tends to be a personal quality rather than one that is peculiar to a certain age-group. Therefore the age limit for scientific work cannot be established statistically, but is determined by the individual characteristics of the scientist.

But the chief cause of premature ageing, in my opinion, is that a scientist who has experienced success in early youth is often weakened by a desire to achieve further results of no less significance, and is thus deprived of that disinterested joy in his day-to-day work, that delight in each small discovery, without which science would come to a standstill. In short, he sooner or later treads the sad path of degeneration we have just described.

It seems to me that a clear understanding of the reasons for premature ageing can enable one to postpone the age limit for effective scientific work. But then I am not indifferent to this question and, in standing up for the ' class ' interests of my age-group, I may have been led astray. One thing is indisputable: when a man who is dedicated to science feels his imagination weakening, his creative powers starting to fail, owing to age or illness, when he can still work, but not so intesively, then there is only one dignified course of action for him to take—and that is to help his students and take pride in their work, in the same way as a sports coach who was once a champion himself takes pride in the records of the athletes he has trained.

3. How do you go about scientific research?

Throwing pebbles in the water,
Stay to watch the circles spreading,
Otherwise to do so would be
A stupid waste of time.

KOZMA PRUTKOV.

Is it possible to trace the origin of those mental leaps, unorthodox comparisons and sudden flashes of insight which, together, make up the creative process? How does one give one's imagination the right direction? What techniques are there to facilitate the search for a solution?

3.1. *The role of the subconscious*

In his book, *Science and Method*, Poincaré attempts to analyse the process of mathematical creativity. According to him, the creative process consists of an alternation of conscious and subconscious efforts. He cites some examples in which, after working on a problem for a long time to no effect, the work has been laid aside and then the solution has come up quite suddenly when out walking or riding on a bus. After that, it has been a matter of a few hours' conscious work to complete the research. The same is often found in theoretical physics and probably in many other fields. M. Zoshchenko, (the famous Russian writer) when he was unable to finish a story, would put it aside, saying, " Never mind; leave it to stew ". Sometimes the solution comes to one in a dream or, more frequently, in that state halfway between waking and sleeping which arises after intense work. I can remember how an explanation was arrived at for the electrons flying out of the atom during nuclear collisions. Qualitatively, everything was clear: on colliding with a nucleon (a neutron or a proton), the nucleus gathers speed for a short time, and the electrons with a smaller speed than that of the nucleus do not have time to fly on with it, but remain in the place where the collision occurred. But what was the quantitative solution? How do you get a formula giving the probability of any one of the electrons flying out? It was the subconscious that came up with the idea for a solution allegorically in a dream: a rider was galloping round in a circus arena, when, suddenly, she stopped. The flowers she had been holding in her hands were thrown into the audience. This picture seemed to suggest in symbolic terms that we had to go on to the system of coordinates at which the nucleus has come to rest after the collision, as in this position it is easier to describe the state of the electrons that are flying out. After this, all that was necessary was to translate this idea into the language of quantum mechanics.

Conscious efforts to solve a problem give the subconscious a frame of reference within which to look for an answer. The subconscious then selects from the stock of accumulated knowledge and especially the reserves of one's own personal experience the combination of concepts that might be of use. These are delivered to the conscious judgement and either remain there, if they turn out to be useful, or else disappear back into darkness. The main characteristic of subconscious work is that associations are uncontrolled. This makes it possible to come up with the most unexpected combinations of ideas.

Sometimes, during a sleepless night caused by worrying about work, you have the impression that you are witnessing this process and can stand aside to observe it, and then one's picture of the process becomes more detailed and, naturally, more subjective. Poincaré compared it to a group of molecules, set in motion by the preliminary work of the conscious mind, colliding and scattering and sometimes joining together to form new compounds. Another image for the subconscious mind is that of a gathering of friends and acquaintances, symbolizing the different concepts. They take an interest in one another and start to interact. You need to know which of them have already met before. You have to get some feeling for the atmosphere of this gathering, and this gives you the key to finding the missing ideas. Of course, this is merely an intuitive picture. According to the principle of observability, its details can only acquire a scientific value once it is demonstrated that they can be used as a basis for increasing the effectiveness of the subconscious process.

3.2. *Making the best use of imagination*

There are, indeed, methods of increasing the effectiveness of the subconscious process. For example, it is well known how important it is in the interests of a fruitful day's work, to have worked even a little the evening before. This seems to ' programme ' the subconscious, so that the following morning, on waking, you already have a clear idea of what needs to be done.

When one has reached a point of deadlock in solving a difficult problem, the best way out is consciously to go through every argument and calculation over and over again until one knows all the pros and cons off by heart and can do all the computations in one's head without any notes. By thoroughly preparing the way, you can make the job of the subconscious so much easier that very soon the solution comes up of its own accord.

It is also possible to regulate artificially the balance between conscious and subconscious work, analysis and intuition. To increase the amount of control, you can work with a partner who tends to make greater use of his critical faculties, while, to increase the role of intuition, you can work with someone who is more inclined to use his imagination.

Another way of increasing the role of intuition, is to forget any difficulties for the time being and let your thoughts wander freely around the subject. This method of ' free association ' is particularly useful to inventors, who need first and foremost to come up with the largest possible number of alternative solutions, before they come to tackle the technical difficulties of putting them into practice.

In order to develop in students the capacity to alternate between controlled and intuitive efforts, it is useful to give improvised lectures, in which the lecturer, with the participation of the students, tries to clarify a question which is new to them both, that is, he tries to show how he himself would go about solving the problem in question. From doing this it becomes clear how the course of the solution is dictated by the logic of the problem itself.

Another method which I have found useful in teaching theoretical physics is to work in the presence of the students. At first, the student only participates silently, trying to understand for himself the difficulties and surprises as they are encountered. But then, he really begins to take an active part in the work, putting forward questions and arguments; discussions start, and, finally, the moment comes when the student begins to think for himself, to define his own problems for solving himself. The fruitfulness of these studies lies not so much in the techniques learnt, as in the fact that the student, together with his guide, experiences the whole tortuous route, with all its ups and downs, from the original conception to the finished piece of work. This would probably be the best way to train mountaineers, if it were not against the safety regulations: for third-time climbers, together with their trainers to set off to climb a mountain of the fifth category of difficulty.

The relationship of a pupil to his teacher is much like that of a child to his parents: it starts off with unlimited respect or even adulation. This is the stage at which they learn the fastest, accepting all the advice their teacher can give them. Then comes the stage of sober assessment and a more critical relation to the teacher. Sometimes, after this, comes a period of coolness in the relationship, or even animosity. At this time, instruction is useless and it is wiser to keep one's distance for a while. As the years go by, animosity usually fades and gives way to respect once more; a mature love grows up, more tolerant of the teacher's faults.

Anyone who has ever done work which lies on the very boundary of the impossible or even beyond it, knows that there is only one way to go about it, and that is by persistent and continuous efforts, by solving auxiliary problems, approaching the problem from different angles, making a note of all the obstacles, stripping away all irrelevant ideas and bringing oneself to a state which might be called a state of ecstasy (or inspiration?), in which the conscious and subconscious merge, conscious thought continues even during sleep, while the subconscious works even during one's waking hours. This state of mind is dangerous because it verges on mental breakdown, on that state of mind which Chekhov described in *The Black Monk*. Einstein wrote that while he was working on the theory of relativity, he even started to have hallucinations.

To receive inspiration, you need a conjunction of several improbable events: the presence of a difficult problem, a feeling of excitement that reaches to the very core of one's being, a sense that you and you only are capable of solving the problem, a mastery of the necessary technique, sufficient experience in solving smaller problems of a similar nature, perfect health, to enable one to withstand a long period of sleeplessness or semi-sleeplessness and, finally, absolute freedom from outside worries. But the chief thing is to have the courage to believe in one's own results, however they might seem to differ from the generally accepted ones, so as not to be afraid of one's own conclusions and of following them through. How many remarkable studies have been laid aside through a lack of boldness!

3.3. *The style of research*

There is a deep affinity between the different forms of creativity. Descriptions of the creative process by artists and poets bear a close resemblance to Poincaré's description of mathematical creativity. There is even a similarity in the methods of executing a piece of work. This similarity was excellently expressed by I. Ya. Pomyeranchuk on his first visit to a sculptor's workshop. He said, " In art, as in science, you must know what you can ignore ". However, there is one essential difference between the truth as it is set down in a work of art and the truth which science strives after. The aim of science is to discover the objective laws of nature, and therefore the end result does not depend upon the personal qualities of the scientist. The aim of art is to know the world through the artists's mediation, to perceive the connection between Nature and the man who observes it. This aim is necessarily subjective, and, for this reason, a work of art always bears traces of the personality of its creator. But the objectivity of science disappears once we look not at the ends of science but at the methods by which they are attained, the methods of arriving at the truth, or the methodology. Every scientist has his own style of research, his own approach to solving the problems that confront him. The style and method of approach also determine the nature of the problems he will study. Here, the individuality of the scientist comes into play, in the same way as the individuality of an architect striving to achieve harmony within the framework of a practical problem.

In theoretical physics, this individuality of styles can be seen from the fact that there are some physicists who do not care which method they use to obtain their results, so long as it serves their purpose, while there are others (who, in my opinion, deserve more respect) who take pains over the methodology of their research and try to achieve their results not artificially, but through the methods most suited to the problem. This gives a deeper understanding, and, consequently, more reliable results. Some physicists are abstract theoreticians, solving problems that are not directly linked with experimental physics, while other theoretical physicists work in close contact with the experimentalists. To such theoreticians, the most important part of their work is the analysis of experiments, either already made or assumed. Some theoreticians, on the one hand, favour a strictly mathematical approach (which is, unfortunately, rarely feasible in theoretical physics), while others prefer the qualitative approach, whereby the results are first obtained from simplified models and, as far as possible, visually.

Among Russian physicists, the best example of a theoretician who strove to obtain his results through the method best suited to the problem was L. D. Landau. The recently deceased academician Vladimir Alexandrovitch Fok was one who always tried to formulate the question as exactly as possible. He obtained very important results in quantum theory by solving problems which could be given a mathematical formulation. The development of physics in several fields owes a great deal to the late Yacob Ilyich Frenkel, a remarkable physicist, to whom we are indebted for a large number of physical ideas which he originated but never attempted to follow through, confining himself to a qualitative study of the problem.

It is not surprising that the scientists who prefer, for example, a strictly formal, mathematical method of research attract, through their work, young people of the same disposition. Thus, a group of people, united by a common style of research and the common tasks which follow from this, is formed. This is how scientific schools come into being. Although the representatives of each school often think that their style is the only right one, the different tendencies, in actual fact, complement and stimulate each other. The truth cannot be affected by the method in which it is approached.

3.4. *Authentic and unauthentic work*

Is a scientist bound in some way by the style or school to which he subscribes or can these evolve with time? The nature of the tasks he chooses and the approach to them ought to change as a scientist becomes better qualified, as his work gains in technical accomplishment and as his experience increases. It is no good trying to tackle ill-defined, problematic work at the very beginning of one's career. One needs to gain experience and master the technical side of the work by solving problems which are not too complex. It is a significant fact that a successful piece of work, that is, one that has been thoroughly completed, is far more valuable as a learning experience than a dozen pieces of work that have been abandoned half-way through because they were too difficult. It is also essential to begin by tackling ' authentic ' problems, that is, problems that do not require unproven or undemonstrable assumptions to be made, but follow on from already existing results. A newcomer to scientific research cannot afford to produce incorrect work.

However, as his experience increases and the number of completed ' authentic ' studies grows, his attitude to ' unauthentic ' (' open-ended ') work should change. Should a serious scientist ever pride himself that he has never made a mistake? I am not, of course, referring here to trivial errors such as miscalculations, or using an unwashed chemical vessel. There are mistakes one should be ashamed of, in the same way as one would be ashamed of a fault in good manners. I am talking about the use of probable but unsubstantiated premises, the error of which is only made clear by subsequent developments in science. On the one hand, never to have produced an erroneous piece of work may be a sign of the conscientiousness and good intuition of the scientist, but, on the other hand, it may simply show that he lacks courage and initiative. A man who had never fallen would never be considered a good mountaineer or a good motorcyclist, because he could not have reached the limits of his own potential. It is precisely the unauthentic works which, if they are borne out by the subsequent developments of science, turn out to be the most interesting, because they substantiate the premise on which they were based. Conversely, totally authentic works which are the inevitable consequence of previously obtained results do not usually give much impetus to science.

This brings us to the relative significance of experiment and theory, a source of great controversy between theoretical and experimental physicists. The coincidence of theory and experiment is not the only, not even the chief argument in evaluating the theory. A good piece of theoretical work consists of a convincing conclusion drawn from previous scientific findings, which, in turn, were the result of an enormous number of repeatedly verified experiments. Where good theoretical work is not borne out by experiment, it means that one must revise the assumptions on which it is based and, as a rule, indicate that one is on the verge of some sort of discovery, great or small. Conversely, if an incorrect theory is supported by experiment, it does not make it any the more convincing. The quality of a theory is judged from its cogency and consistency. Convincingly constructed ' unauthentic ' theories can influence the development of science even in those cases where the assumptions on which they were based turn out to be wrong. I should like to take as an illustration of this an outstanding piece of work by the late academician, I. E. Tamm, which greatly influenced the physics of elementary particles. At that time (in 1934), Enrico Fermi's theory of beta-decay had only just appeared. This theory demonstrated the mechanism by which a neutron is converted into a proton, emitting an electron and a neutrino. It is this mechanism that sets in motion the radioactive conversion of one nucleus into another, with the emission of an electron and a neutrino. With this mechanism as his basis, Tamm constructed a theory of nuclear forces, that is, the forces that kept the nucleons (neutrons and protons) inside the nucleus. The basic idea of this theory was that one of the nucleons emits an electron and a neutrino (or a positron and a neutrino) and the other nucleon absorbs these particles. The subsequent development of science showed that the transfer of the electron and the neutrino plays very little part in nuclear forces. Nuclear forces do, indeed derive their energy from the fact that the nucleons, as in Tamm's theory, emit and absorb particles, but other particles that were discovered later. One of these is the π-meson. Thus, the initial premise of the theory was not confirmed. Nevertheless, the idea that nuclear forces have something to do with the emission and absorption of particles by the nucleon was not only right, but turned out to be fruitful.

Every scientist should, from time to time, ask himself the following question: why has so-and-so contributed more to science than I have, although my understanding and mathematical skill are in no way inferior to his? The answer is usually the same in every case. " He has the courage to go through with unauthentic pieces of work, whereas I expend all my efforts on authentic ones. "

3.5. *The new style of research of the second half of the twentieth century*

We have already dealt with changes in the style of work due to age, experience and qualifications. But the development of science itself can bring about far greater changes in style. During a lull in the development of science, it is best to get on with routine work, elaborating on existing results and preparing techniques for further research and perhaps further discoveries. But, during the period of turmoil which follows an important discovery, one should leave routine work aside and concentrate on achieving new results, even though these may be attained through coarser, less reliable methods.

During the second half of the twentieth century, a dramatic change has taken place in the style of theoretical research, although perhaps not all physicists have drawn the inevitable conclusions from this. Theoretical research has undergone a complete reorganization, which one might call the ' collective brain '. Supposing, as a result of the analysis of accumulated experiments or of some sort of experimental discovery, an important and complex problem arises, which is too difficult to be solved by one man. The following collective strategy is then adopted: a group of scientists (those who have a natural affinity for this sort of task) starts to work at generating ideas (any ideas, true or untrue, using the free association technique outlined above). They then try to produce a partial explanation of the phenomenon in terms of these ideas. These unfinished pieces of work are then published in the form of preprints over a period of ten to twenty days. Then, every two to three months, small conferences are held, at which the accumulated material is subjected to criticism by another group of physicists. As a result of the discussions, a preliminary selection is made, this time mainly by highly qualified physicists actively working in the relevant field. Basically, this group of physicists picks out the most sensible ideas, draws conclusions and indicates a direction for further theoretical and experimental research. About once a year, a large, summing-up conference is held. In this way, the initial ideas are like mutations, which either consolidate themselves or are quickly succeeded by others. The conferences are organized along the lines of ' natural selection ' and, owing to the spontaneous division of labour, an idea brought up by a young man just out of university can become the centre of attention of the whole conference. When it comes to the next problem, the same young man may find himself not among the ranks of those generating the ideas but among the physicists making the critical selection. By this method, some of the most important problems of the theory of elementary particles have been (and are still being) solved, to state them briefly without giving details: SU(3)—symmetry, quarks, dual models. Does the ' new 20th century style ' detract from the romanticism of research? On the contrary, although the role of the individual is diminished, it has created a new romanticism—the romanticism of collective work.

3.6. *The role of computers*

The invention of computers has brought about yet another substantial change in the style of theoretical physics.

In the past, a problem was considered solved if the solution could be expressed in terms of known (elementary) functions. This only happens extremely rarely and cases of this type are quickly exhausted. Later, it was considered enough to express the solution in terms of functions that had been specifically defined for a certain circle of problems (' special ' functions). However, even this did not satisfy the needs of science. Approximate methods were introduced. According to these, the solution is expressed in the form of a sum of an infinite series, each of the terms of which contains known functions. In order that these series can be used, it is necessary for the first few terms to give the result exactly enough (as mathematicians say, the series must ' be convergent '). For the terms of the series to decrease quickly, there has to be some sort of small parameter†, powers of which factorize the terms. Therefore the question which, until a short time ago, the theoretician was always asked was, " What is the small parameter of your problem? ". Usually this could be taken to mean, " Your theory is doubtful because it has no small parameter and it is not clear what is the contribution of the omitted terms ".

When solving a problem by computer, you do not need to use a small parameter. The solution is not expressed in terms of any sort of functions of the parameters of the problem (the analytical form of the solution), but it is

† The parameters are the set of quantities which define the conditions of a problem.

given in the form of a collection of numerical tables. The solution is not sought in analytical terms. With the appearance of computers, interest in the analytical form of the solution has sharply declined (but, as we shall see, it has not disappeared altogether!).

One extreme example of the computerized approach has been demonstrated by a brilliant exponent of the modern style, the American theoretical physicist Kenneth Wilson. He used the computer to solve a problem known as Kondo's problem, after the Japanese physicist of that name who made the first step in formulating the question. The problem consisted in finding an explanation for the abnormal behaviour at low temperatures of metals containing an admixture of atoms with magnetic moments. At very low temperatures, the magnetic susceptibility and electrical resistance first of all start to grow with the fall in temperature and then proceed to some constant limit. Theoretical research on this problem showed that, with a fall in temperature, the interaction of the electrons of the metal with the impurity atoms becomes so intense that the usual methods of examining them, which assumes only a small interaction, are quite inapplicable. A new approach had to be found which did not involve the use of a small parameter. Such methods were developing fast under the influence of problems which arose first in the theory of elementary particles and later in solid-state physics.

Nevertheless, attempts to solve the problem analytically proved fruitless. Wilson, after making a profound analysis of the problem, managed to formulate it in such a way that he made it possible to use computers, thus finding the magnetic susceptibility at a given temperature in only a few minutes' computer time. It is true that these ' few minutes ' were the fruit of long research into methods of simplifying the problem. Without these simplifications, the calculation would have been impossible, as it would have taken many hundred years of computer time. Thus, the problem had been removed from the agenda. Nevertheless, the physical understanding of Kondo's problem was an important step in the development of theoretical physics. It was this and not the explanation of the resistance or magnetic susceptibility at low temperatures which comprised the heuristic value of this problem.

Here we come to the question of how far it is possible to use computers in scientific research. Why should a theoretician who has already obtained a simple result by a reliable but complicated method necessarily go on to look for a simpler method, to have the result at his fingertips? He does it so that if, in solving another problem, he meets with a similar situation and the complicated method fails, he can then use the simpler one, which has evolved through his deeper understanding.

How was it that Wilson was able to create a calculational framework which cut down so dramatically on the computer's work? The answer is that Wilson had already been working on Kondo's problem for several years, trying to find an analytical solution; he had completed many other pieces of research in closely related fields and so was ready to solve this particular problem.

Let us look at another example. A large amount of research has been done to try to find an explanation for the properties of the nucleus, seeing it as a gas of neutrons and protons and using as a starting point the interaction of two nucleons which was found from nucleon–nucleon scattering in vacuum. This interaction is not small and in the problem there is no small parameter.

This is no longer an objection if one is using a computer. The calculation programme can be prepared in such a way that the answer will be presented numerically, despite the absence of a small parameter. Nevertheless, this gives rise to a large error: it does not take into account the possibility of new collective degrees of freedom appearing in the nucleus (π-meson condensation). The possibility of such unexpected phenomena must be taken into account when programming, and, for this to be done, a rough preliminary, analytical solution must be found.

The following conclusion seems to suggest itself: before one can use computers, the problem must be investigated from all angles by analytical methods. Analytical methods, "the old but trusty weapon", have not lost their significance.

3.7. *Common sense*

Usually, in scientific writings and especially in textbooks, the process of scientific discovery as a whole undergoes careful editing. This leaves it to be conjectured how any of the results quoted were obtained, what difficulties were encountered on the way and how they were overcome. And yet, what the science student would find most useful would be a detailed description of the mental trajectory followed, with its successes and failures, and of the attempts to tackle the problem from different angles. Moreover, to state the final results without specifying the methods by which they were obtained and the difficulties encountered only leaves the student with a sense of inadequacy, as it gives him the impression that, in order to be a scientist, you have to possess special intellectual qualities which differ completely from normal common sense and allow one suddenly to leap to unexpected conclusions.

In actual fact, we are all endowed with the same mind and one of the necessary tools of scientific research alongside intuition and imagination is that same common sense which a housewife uses when she makes wise purchases at the market. Fermi used to astonish his students by setting them the following question: how many piano makers are there in Chicago? He could judge from the way they went about obtaining figures how well they knew how to use their common sense.

The way to understand something, whatever it might be, even the most abstruse and complex of matters, is not through sudden, blinding intuition but by patient, persistent work. For this reason, despite the fact that conscious efforts alternate with unconscious and, it would seem, introduce an indefinite element of guesswork and intuition the results obtained in science are proportional to the amount of work done and the time that is spent on it.

It is through common sense that one can organize one's work and working methods in such a way that one need make only small intuitive leaps. A complex problem should be broken down into a number of much simpler problems. The movement towards the final result, like that of a mountain climber creeping towards the summit, is really no more than a process of overcoming comparatively small difficulties one by one, a step by step progression. How is this accomplished? First of all the problem is reduced to its simplest form, so that only the chief points remain. It is incomparably easier gradually to elaborate on a problem which is already solved than it is to solve a complex one as it stands. Once this is done, the possibility of solving the problem in limiting cases is investigated. Likewise, before attempting to find a quantitative solution, one must first obtain rough, qualitative results, which is much easier. And, finally, at every stage, one must try to disprove the results obtained, using all the relations discovered so far which the result obtained should comply with in limiting cases.

It is also essential to test the logical coherence of the results obtained. Does the result follow from your original premises? Does the result contradict any general principles which might be inadvertantly broken by conclusions drawn? Do the limits of applicability of the result coincide with those that will be established finally? Often the result is used in a wider context of assumptions than those which were made during the process by which it was obtained. Or, in mathematical terms, the result may sometimes be analytically extended beyond the limits of the assumptions made.

Was the result obtained too easily? There seems to exist something in the nature of a 'law of conservation of difficulty'. If, by approaching the problem in a particular way, the principal difficulties are solved, then, as a rule, it should be equally easy to clear them up using any other approach. Suppose that an ingenious circuitous technique has been thought up as a way of eliminating the difficulties: one should not be satisfied with results obtained in this way, but go on to investigate what it was that made the difficulties disappear. As a rule, once this investigation has been made, you can either successfully solve the problem by a more direct method, or else you find that the artificial (circuitous) solution was wrong.

Finally, does the result satisfy aesthetic requirements? Sometimes, these relate to the external appearance of the formula. If it contains large or implausibly small figures, then the expression looks ugly. If you see a formula with awkward figures you can suspect a mistake. And very often, on testing them, you find that these 'ugly' expressions are, indeed, correct. It looks ugly if the formula contains a large number of coefficients which cannot be calculated theoretically but must be determined by reference to the experiment. The feeling for 'beauty' is difficult to put across without using complicated examples. Sometimes it merely consists in the sheer simplicity of the expression, which is pleasing to the eye.

One way of telling whether a calculation is correct is the elimination of the complex intermediate expressions, which simplifies the final result and makes it look pleasing. As one physicist said, "Correct expressions have a tendency to simplify themselves". But much more important is not the external but the internal beauty of the results. An expression is beautiful if it shows the connection between heterogeneous phenomena in a simple form or discovers unexpected connections between them. One of the most beautiful formulae in theoretical physics is the formula of Einstein's theory of gravity, which relates the radius of the curvature of space to the density of the matter. Another remarkable example is that of Maxwell's equations, which contain in compact form all the possible information about every electrical and magnetic phenomenon. The test of beauty, although it is not infallible, plays a very important part both in searching for new laws of Nature and in checking results.

3.8. *The sequence of actions*

The following is, in my opinion, a good order in which to work in theoretical physics (and possibly in other fields as well)†.

You should start to tackle the problem before studying the literature. This early acquaintance with the problem without the preconceptions which can be formed from reading existing works on the subject, the first estimates of the order of values to be expected, the first explorations of different methods of solving it determine, to a large extent, the future course of the work. Next, one starts to take an active interest in the literature (the second stage of the work). Studying the literature for possible future use is never so effective as studying it for a specific purpose, from a specific point of view. After this or simultaneously with it, one starts to form an idea of the limits imposed on the possible result by the general principles of theoretical physics (for example, the conservation laws). The next stage is to try to find a rough, qualitative solution for different values of the parameters of the problem. Then, one should try to find a quantitative solution to the problem in limiting cases, that is, for values of the parameters which tend to make the problem as simple as possible. Here we come to what is possibly the most important and most difficult stage of the work, and that is to analyse and criticize the results obtained by all the methods I have already outlined. If all the work completed so far is found to be correct, then one can proceed to the last phase of the work—obtaining the quantitative result analytically or with the use of computers. At every stage, of course, the work should be submitted for criticism to all the people who have worked on the same or related problems. The work is finally completed on publication. The finished work, once it is ready to go to press, should be kept aside for a time and then reviewed once more. The time for which it is held back is at the author's discretion.

All this requires illustrating with examples. Unfortunately, we shall have to go back to school physics to find these illustrations.

3.9. *How to guess the solution*

Let us look at an example of how some aspects of a solution can be worked out before the apparatus for reaching a precise solution has been even prepared, to such an extent that the very equations on which the solution of the problem will be based can be found.

One of the most difficult unsolved problems of theoretical physics is that of the connection between gravitational and electromagnetic phenomena. If such a connection exists, then, by solving certain equations which have not yet been found, a dimensionless‡ figure will be obtained, giving the relation between the gravitational constant G and the quantities which are used in electrical

---

† A dimensionless quantity is a quantity which is not expressed in terms of a specific unit of measurement. Thus, for example, the length of a table has the dimensions of length and its expression in figures depends on whether it is being measured in centimetres or inches. The ratio of the length of the table to its width is a dimensionless quantity and does not depend on a chosen unit of measurement.

‡ Here I should like to mention (and recommend to the reader) a brilliant book by D. Polya, *How to Solve a Problem*, which suggests a good sequence of actions to follow when solving mathematical problems.

measurements, such as the speed of light c, the charge on the electron e and its mass m_e. If quantum phenomena have any importance, then the result might also include Planck's constant $\hbar$, which, as we have seen, describes the discontinuities of energy in elcetromagnetic waves. As we already know the dimensions of the quantities G, c, e, m_e and $\hbar$, it is easy to draw the conclusion that only two distinct dimensionless combinations can be composed from them:

$$\alpha = \frac{e^2}{\hbar c}; \quad \xi = \frac{\hbar c}{G m_e{}^2}$$

The first of these combinations is well known and is called the fine structure constant. A substitution of numerical values for symbols gives $\alpha = 1/137$; $\xi = 5 \times 10^{44}$. Can such a large figure as ξ be the solution to any sort of reasonable equation? Dimensionless quantities, when they appear in a formula which represents the solution to a problem in physics, are of the order of only a few units or a fraction. Therefore it is reasonable to expect that the value of ξ will come into the problem in such a form that the answer will be a number of about 1. So far we have used common sense. Now we must make a small intuitive leap. It is probable that the combination

$$\alpha \ln \xi \sim 1$$

will come into the theory: it is obvious that knowing this relation would make it easier to find a solution. And it is in precisely this form that the quantity ξ does come into current attempts to solve the problem of the connection between electrodynamics and gravity.

3.10. *Sometimes a mistake leaps to the eye*

What limits do the general theorems of theoretical physics impose upon the solution? One sunny day in winter, a large group of people were sunning themselves on the top of the Kokhta in Bakuriam. Some youngpeople were staring in delight and amazement at the bright blue sky. One of them said, "the blue of the sky can be attributed to the fact that, according to Rayleigh's law, the scattering of light is proportional to the third power of the frequency and a blue light, having a greater frequency, is scattered more strongly". This was too much for the physicist among them, who pointed out, "The scattering of light is a reversible process and cannot contain odd-numbered powers of the frequency, so Rayleigh's law must state the fourth and not the third power of this frequency. By accepting the scattering to be an odd function of the frequency, we are contradicting the law of reversibility of natural phenomena and thereby the whole of thermodynamics". This conversation raised the author's prestige, which had been suffering from his low qualification as a mountain climber.

There really does exist such a theorem: all the equations of physics and, consequently, all the natural phenomena described by them are time invariant, that is, they look the same whether you are looking at them from the past into the future of from the future into the past. It is from this that it follows that reversible quantities can only be even functions of a frequency.

3.11. *Logical analysis*

Let us study an example of an analysis of the logical structure of a proof.

We shall take as our example the theoretical proof that all bodies fall with equal speed. This proof was first set down by Galileo in his famous book, *Dialogue Concerning Two New Sciences* (1633). Galileo used the following arguments to dispute Aristotle's assertion that heavier bodies fall faster than lighter ones, which at that time was an act of considerable boldness. Let us assume that Aristotle was right and a heavier body does fall faster. If we were to attach two bodies together, a lighter one and a heavier one, the heavier body, falling faster, would accelerate the fall of the lighter one and the lighter one, falling more slowly, would slow down the fall of the heavier one. Therefore the two bodies fastened together would fall at a speed somewhere between those of the same two bodies falling seperately. However, it would be heavier than either of its parts and so should move not at an intermediate speed, but at a greater speed than that of its heavier part. This is a contradiction and therefore the initial assumption must be wrong.

The reverse of this assumption, that the light bodies fall faster than heavy ones, can also be reduced to a contradiction by an analogous argument. The same argument can be repeated, this time supposing that two identical bodies are attached together. These would not slow one another down or speed one another up, so they should move at the same speed as either of them separately. Thus, a body twice as large falls at the same speed. Consequently, all bodies fall with equal speed.

Let us examine these arguments more closely. At first sight, it seems that they comprise a strict, purely logical proof that all bodies fall with equal speed. But, on the other hand, we cannot accept this conclusion on the basis of a purely intellectual argument without using any sort of experimental data. Or, in more up-to-date language, this proof already suggests information that was obtained from Galileo's experiments throwing weights off the Leaning Tower of Pisa (lead shots of different masses reached the ground simultaneously) or from other similar experiments. So then, we do not understand the logical structure of this proof and are, consequently not wholly convinced by it.

Since the assumption that heavier bodies fall faster is logically tenable, we can be permitted to use Galileo's reasoning to establish by which facts his arguments contradict this hypothesis. In this case, the addition of the small body to the large one should not slow it down but speed it up, since, once they are attached, the body that is thus formed should fall with even greater speed. On the other hand, if the two bodies are tied together by a long, fine thread, then they will try to move as if they were not attached, that is, the heavier body will try to move faster and the smaller body will impede its fall. And yet, when attached in an ordinary manner the reverse should apply—the smaller body should speed up the larger. This would mean that the falling speed of an object depends upon whether its parts are loosely or tightly attached. Experiments on weighing have shown that the weight of an object is equal to the sum of the weights of the parts of which it is made up, regardles of how these parts are attached. Thus, the weight as a whole of a composite body does not depend on how its parts are attached, and yet the speed of its fall should. But this contradicts Galileo's experiments on motion on an inclined plane, from which it follows that, with a given mass, the falling speed is wholly determined by the force. So then, Galileo's proof is not absolutely logical: it does not make the fullest use of the experimental data that were available at the time.

In conclusion, I should like to try, as I have advocated, to reduce all I have said to its simplest form. The driving force behind scientific creativity should not be the desire to bring about a scientific revolution; nor should all one's efforts be directed towards achieving success, but one should be motivated by a love of knowledge, a capacity to wonder at and delight in each small success and, above all, a feeling for the beauty of science. It is important to develop impeccable conscientiousness and to learn to reduce the most complex question to extreme simplicity and clarity; to find a way out of many psychological contradictions; to be guided by intuition but not to put one's trust in it; to be conscious of all the difficulties, but to be able temporarily to divert one's attention from them; to believe in a result but at the same time to search patiently for a way to refute it; to find one's own style of working but to be able to change it with experience and with each new major discovery: in brief one must aim to understand everything, "The meaning, cause, foundation, roots and kernel...", as Pasternak put it. His poetical works begin with the words, "In everything I want to go to the heart of the matter, in work, in the search for a way, in the confusion of the heart". Let these lines serve as a watchword for anyone who would embark on a scientific career.