Internalizing physics

Science and Technology Education Documents Series (STEDS)

No. 1 Glossary of Terms used in Science and Technology Education. 1981 (English)

No. 2 Methodologies for Relevant Skill Development in Biology Education. 1982 (English)

No. 3 Nutrition Education: Curriculum Planning and Selected Case Studies. 1982 (English) (Reprint in Nutrition Education Series No. 4)

No. 4 Technology Education as part of General Education. 1983 (English and French)

No. 5 Nutrition Education: Relevance and Future. 1982 (English) Reprint in Nutrition Education Series, No. 5)

No. 6 Chemistry Teaching and the Environment. 1983 (English)

No. 7 Encouraging Girls into Science and Technology Education: Some European Initiatives. (English)

No. 8 Genetically-Based Biological Technologies. 1984 (English)

No. 9 Biological Systemes, Energy Sources and Biology Teaching. 1984 (English)

No. 10 Ecology, Ecosystem Management and Biology Teaching. 1984 (Reprint 1986) (English)

No. 11 Agriculture and Biology Teaching. 1984 (English)

No. 12 Health Education and Biology Teaching. 1984 (English)

No. 13 The Training of Primary Science Educators - A Workshop Approach. 1985 (English)

No. 14 L'Économie sociale familiale dans le développement rural. 1984 (French)

No. 15 Human Development and Evolution and Biology Teaching. 1985 (English)

No. 16 Assessment: A Practical Guide to Improving the Quality and Scope of Assessment Instruments. 1986 (English)

No. 17 Practical Activities for Out-of-School Science and Technology Education. 1986 (English)

No. 18 The Social Relevance of Science and Technology Education. 1986 (English)

No. 19 The Teaching of Science and Technology in an Interdisciplinary Context. 1986 (English)

No. 20 Mathematics for All. 1986 (English, Spanish)

No. 21 Science and Mathematics in the General Secondary School in the Soviet Union. 1986 (English)

No. 22 Leisure, Values & Biology Teaching. 1987 (English and French)

No. 23 Use of Sea and its Organisms. 1987 (English)

No. 24 Innovations in Science and Mathematics Education in the Soviet Union. 1987 (English)

No. 25 Biology and Human Welfare. Case Studies in Teaching Applied Biology. 1988 (English)

No. 26 Sourcebook of Science Education Research in the Caribbean. 1988 (English)

No. 27 Pour un enseignement intégré de la science et de la technologie : trois modules. 1988 (French)

No. 28 Microbiological Techniques in School. 1988 (English)

No. 29 Games and Toys in the Teaching of Science and Technology. 1988 (English, French)

No. 30 Field Work in Ecology for Secondary Schools in Tropical Countries. 1988 (English, Arabic)

No. 31 Educational Materials Linking Technology Teaching with Science Education: Technology in Life. 1988 (English)

No. 32 Evaluation and Assessment in Mathematics Education. 1989 (English)

No. 33 Systems Thinking in Biology Education. 1989 (English)

No. 34 Base Physique de l'électronique dans l'enseignement secondaire ; module méthodologique. 1989 (French)

No. 35 Mathematics, Education and Society. 1989 (English)

No. 36 Bibliography in Integrated Science Teaching. 1990 (English)

No. 37 Educación Matemática en las Américas VII. 1990 (Spanish)

No. 38 The Teaching of Science and Technology in an Interdisciplinary Context. 1990 (English)

No. 39 Teaching Biotechnology in Schools. 1990 (English)

No. 40 Electronics Teacher's Guide. 1991 (English)

No. 41 Children, Health and Science. 1991 (English, French, Spanish)

No. 42 Reuniones del Primer Congreso Iberoamericano de Educación Matemática. 1992 (Spanish)

No. 43 Educación matemática en las Américas VIII. (Spanish)

No. 44 The Influence of Computers and Informatics on Mathematics and its Teaching. (English)

No. 45 Physics Examinations for University Entrance. (English)

No. 46 Education for Teaching Science and Mathematics in the Primary School. (English)

No. 47 Significant Influences on Children's Learning of Mathematics. (English)

1 2

3 4

Cover photos

1. Photo UNESCO/Paul Almasy
2. Photo UNATIONS
3. Photo UNESCO/D. Bahrman
4. Photo rights reserved

Science and Technology Education
Document Series No. 48

Internalizing physics

Making Physics Part of One's Life

Eleven Essays
of
Nobel Laureates

Edited
by
D.K. Nachtigall

Education Sector
UNESCO

The designations employed and the presentation of the material in this document do not imply the expression of any opinion whatsoever on the part of UNESCO concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.

Published in 1995 by the United Nations Educational, Scientific and Cultural Organization,
7, place de Fontenoy,
75352 Paris 07 SP

Printed by UNESCO

Printed in France

Contents

FOREWORD

Most students consider physics to be the most difficult, the most boring and the most disliked subject within the sciences.

This is particularly true for developing countries. The reason is that the formal approach and the symbolic representation of physics, practiced at the universities, tends to be adopted by most physics teachers. They teach as they were taught at the university. But the distance between teachers' style, reasoning and content delivery in the class-room on the one side, and daily experience, way of thinking and style of communication in the ordinary world on the other side, is very great. In order to bridge this gap we need more and better educated physics teachers all over the world.

The rich countries can solve this problem when insights of great physicists and research results of many physics educators are realized and put into practice. For developing countries the situation is more difficult. There are many more obstacles in the way of attempts to make physics a cultural enterprise.

First of all: physics is a relatively new occupation in these countries. An embedding context to physics is not yet established. The physics curricula adopted by the governments are bookish and of academic character. The content appearing in the books does not fit local needs and cultural self-consciousness. The context in which physics is presented appears meaningless, foreign, strange and esoteric. The methods of teaching are examination centered and rote learning and memorizing still have priority over understanding. The schools - mainly those in rural areas - are poorly equipped and the language of teaching physics is very often not the mother tongue. Most physics teachers are poorly trained, scarcely motivated, their salary and social prestige are very low and their teaching load is very burdensome.

However, as a visiting professor and as the nominee of the German Physical Society for physics teachers' education in developing countries I had the opportunity to meet senior physics teachers and trainers who really are involved and who try, with tremendous efforts, to improve the situation in their countries. They need and deserve help.

This help should be sensitive to their needs. And there are many needs. These are the most urgent requirements:

Physics teachers need

1. getting teaching aids such as books and equipment and modifying this material so that it is applicable to local environments and conditions,

2. learning to fabricate and to use experiental equipment made of low cost material available in the region,

3. becoming aware of the result of research in physics education, particularly concerning the role of common sense ideas and beliefs, of superstition, and of conceptions and misconceptions in their cultural and cross-cultural contexts in the process of teaching and learning physics,

4. becoming familiar with a set of flexible teaching methods adaptable to the actual situation in the class-room in place of authoritarian teacher-centered and linear didactical approaches,

5. understanding that physics is not a set of laws and definitions, but the permanent search for and application of methods of asking novel questions and trying to obtain partial and approximate answers,

6. overcoming the tendency of punitive authority inherently present in teacher education institutions which hinders and prevents the student-teachers to develop self-confidence.

The developing countries' drive towards economic self-reliance needs to overcome the low enrollment of students.

This is possible only when more good, that means competent and enthusiastic physics teachers are educated. The best way of doing this is just to let them get together with a good trainer or senior teacher who could give examples for effective and motivating physics classes. Watching good trainers does not mean copying them, but getting a guideline on how to come to a self-realization i.e. to increase one own's potential and ability as a teacher in the framework of an accepted educational philosophy. Such model teachers that I know - and there is fortunately a growing number of them - in most cases are already getting help as far as the first five items of my list of needs are concerned. What they still urgently need and what is so hard to provide is the development of their self-confidence, is moral support for their self-realization and is concrete evidence of justification of self-respect.

It's this point that induced me to ask Nobel Laureates to contribute to this 'book of encouragement'. They mark the top of the pyramid of the members of the community of physicists. At the bottom are the unknown physics teachers in the developing countries and among them the trainers and senior teachers. They are a fundament on which depends whether physics teaching and learning becomes something that can be implemented in the societies as an enterprise that is of great advantage to social and economic development, without destroying the cultural identity.

This book is dedicated to them. I consider it a unifying link between the top and the bottom of the pyramid. There are many ways in which this book will prove useful for the addressees. Here is just one way:

A. Einstein's remark on how thinking goes on and the aspect of a "conflict with the world of concepts already sufficiently fixed within us" is used as a basis of a teaching strategy.

N. Mott's plea for solar energy is much more than just interesting for developing countries.

A. L. Schawlow's advice to concentrate in physics on finding something "not obviously important", rather than "to wait for an earth-shaking inspiration", can encourage many unknown physicists.

L. M. Lederman states that "Being average now isn't decisive....Most scientists aren't brilliant", but "determination, doggedness and hard work are the characteristics that are highly valued...".

S. Chandrasekhar's article contains pieces of history of physics which show that physics is a human enterprise in which the whole scale of human affections, feelings and sensations, excitements and frustrations, joy and despair, can be experienced. The deepest desire of a physicist may be that he adds something to knowledge, helps others to add more,...and leaves behind some kind of memorial.

N. Bloembergen stresses the importance of the usage of simple physical phenomena in illustrating basic concepts of physics in quantitative reasoning. He gives advice about the application of the concept of "order of magnitude" or "powers of ten". This reads like a guide line of good physics education in the class-room.

The reader gets quite another type of insight when he/she studies J. Bardeen's article on semiconductor research and H. Rohrer's and G. Binnig's contribution on the scanning tunnel microscope. Both articles contain their Nobel lectures and the reader can witness how great physics is being done.

G. Binnig's discussion of the term 'creativity' and his warning to consider Man as being apart from nature rather than being part of the whole, as well as his claim that "man overvalues himself and underestimates the so-called inanimate nature...", part of which are computers which are also "somehow able to create something new", is certainly a challenging standpoint that should stimulate epistemological disputes among teachers and learners of physics.

A. Penzias' article on 'ideas' adds not only practical examples to the phenomenon 'creativity' but shows in a fascinating and exciting way the value of estimations, of 'inexact' answers, of the relevance of "orders of magnitudes" in daily life. Such insights should become common to all physics teachers.

The last article, a talk that R. P. Feynman gave before an audience of physics teachers, can be used as a thread through a textbook on physics education. How his father taught Richard Feynman to recognize patterns, how pi can be discovered, how "fiddling around" trains experimental skills, how important it is to find relations between formulas, how children can understand the concept of inertia etc., etc..., all these examples help teachers to improve their skills. When Feynman makes clear what the difference is between teaching a definition and teaching physics, then the reader will discover how bad so many textbooks are. And when he describes how one can find out whether a teacher has taught an idea or has taught only a definition, then I wish that not only school teachers but also university professors would take his advice to heart.

At the end of the book we find Feynman's very special message of encouragement:

"I am trying to inspire the teacher at the bottom to have some hope, and some self-confidence in common sense and natural intelligence. The experts who are leading you may be wrong."

Dortmund, September 1994

D. K. Nachtigall
EDITOR

I

The prize was awarded in 1922 to Albert Einstein for his services to Thoretical Physics, and especially for his discovery of the law of the photoeletric effect.

* * *

ALBERT EINSTEIN:

Autobiographical Notes.

What, precisely, is "thinking"? When, on the reception of sense impressions, memory pictures emerge, this is not yet "thinking". And when such pictures form sequences, each member of which calls forth another, this too is not yet "thinking". When, however, a certain picture turns up in many such sequences, then-precisely by such return-it becomes an organizing element for such sequences, in that it connects sequences in themselves unrelated to each other. Such an element becomes a tool, a concept. I think that the transition from free association or "dreaming" to thinking is characterized by the more or less preeminent role played by the "concept." It is by no means necessary that a concept be tied to a sensorily cognizable and reproducible sign (word); but when this is the case, then thinking becomes thereby capable of being communicated.

With what right - the reader will ask - does this man operate so carelessly and primitively with ideas in such a problematic realm without making even the least effort to prove anything? My defense: all our thinking is of this nature of free play with concepts; the justification for this play lies in the degree of comprehension of our sensations that we are able to achieve with its aid. The concept of "truth" can not yet be applied to such a structure; to my thinking this concept becomes applicable only when a far-reaching agreement (convention) concerning the elements and rules of the game is already at hand.

I have no doubt but that our thinking goes on for the most part without use of signs (words) and beyond that to a considerable degree unconsciously. For how, otherwise, should it happen that sometimes we "wonder" quite spontaneously about some experience? This "wondering" appears to occur when an experience comes into conflict with a world of concepts already sufficiently fixed within us. Whenever such a conflict is experienced sharply and intensively it reacts back upon our world of thought in a decisive way. The development of this world of thought is in a certain sense a continuous flight from "wonder."

A wonder of this kind I experienced as a child of four or five years when my father showed me a compass. That this needle behaved in such a determined way did not at all fit into the kind of occurrences that could find a place in the unconscious world of concepts (efficacy produced by direct "touch"). I can still remember - or at least believe I can remember - that this experience made a deep and lasting impression upon me. Something deeply hidden had to be behind things. What man sees before him from infancy causes no reaction of this kind; he is not surprised by the falling of bodies, by wind and rain, nor by the moon, nor by the fact that the moon does not fall down, nor by the differences between living and nonliving matter.

At the age of twelve I experienced a second wonder of a totally different nature - in a little book dealing with Euclidean plane geometry, which came into my hands at the beginning of a school year. Here were assertions, as for example the intersection of the three altitudes of a triangle at one point, that - though by no means evident-could nevertheless be proved with such certainty that any doubt appeared to be out of the question. This lucidity and certainty made an indescribable impression upon me. That the axioms had to be accepted unproved did not disturb me. In any case it was quite sufficient for me if I could base proofs on propositions whose validity appeared to me beyond doubt. For example, I remember that an uncle told me about the Pythagorean theorem before the holy geometry booklet had come into my hands. After much effort I succeeded in "proving" this theorem on the basis of the similarity of triangles; in doing so it seemed to me "evident" that the relations of the sides of the right-angled triangles would have to be completely determined by one of the acute angles. Only whatever did not in similar fashion seem to be "evident" appeared to me to be in need of any proof at all. Also, the objects with which geometry is concerned seemed to be of no different type from the objects of sensory perception, "which can be seen and touched." This primitive conception, which probably also lies at the bottom of the well-known Kantian inquiry concerning the possibility of "synthetic judgments *a priori*," rests obviously upon the fact that the relation of geometrical concepts to objects of direct experience (rigid rod, finite interval, etc.) was unconsciously present.

If thus it appeared that it was possible to achieve certain knowledge of the objects of experience by means of pure thinking, this "wonder" rested upon an error. Nevertheless, for anyone who experiences it for the first time, it is marvelous enough that man is capable at all of reaching such a degree of certainty and purity in pure thinking as the Greeks showed us for the first time to be possible in geometry.

Now that I have allowed myself to be carried away sufficiently to interrupt my barely started obituary, I shall not hesitate to state here in a few sentences my epistemological credo, although in what precedes something has already incidentally been said about this. This credo actually evolved only much later and very slowly and does not correspond to the point of view I held in younger years.

I see on the one side the totality of sense experiences and, on the other, the totality of the concepts and propositions that are laid down in books. The relations between the concepts and propositions among themselves are of a logical nature, and the business of logical thinking is strictly limited to the achievement of the connection between concepts and propositions among themselves according to firmly laid down rules, which are the concern of logic. The concepts and propositions get "meaning," or "content," only through their connection with sense experiences. The connection of the latter with the former is purely intuitive, not itself of a logical nature. The degree of certainty with which this connection, or intuitive linkage, can be undertaken, and nothing else, differentiates empty fantasy from scientific "truth." The system of concepts is a creation of man, together with the rules of syntax, which constitute the structure of the conceptual systems. Although the conceptual systems are logically entirely arbitrary, they are restricted by the aim of permitting the most nearly possible certain (intuitive) and complete coordination with the totality of sense experiences; secondly they aim at the greatest possible sparsity of their logically independent elements (basic concepts and axioms), i.e., their undefined concepts and underived (postulated) propositions.

A proposition is correct if, within a logical system, it is deduced according to the accepted logical rules. A system has truth-content according to the certainty and completeness of its possibility of coordination with the totality of experience. A correct proposition borrows its "truth" from the truth-content of the system to which it belongs.

A remark as to the historical development. Hume saw clearly that certain concepts, as for example that of causality, cannot be deduced from the material of experience by logical methods. Kant, thoroughly convinced of the indispensability of certain concepts, took them - just as they are selected - to be the necessary premises of any kind of thinking and distinguished them from concepts of empirical origin. I am convinced, however, that this distinction is erroneous or, at any rate, that it does not do justice to the problem in a natural way. All concepts, even those closest to experience, are from the point of view of logic freely chosen posits, just as is the concept of causality, which was the point of departure for this inquiry in the first place.

And now back to the obituary. At the age of twelve through sixteen I familiarized myself with the elements of mathematics, including the principles of differential and integral calculus. In doing so I had the good fortune of encountering books that were not too particular regarding logical rigor, but that permitted the principal ideas to stand out clearly. This occupation was, on the whole, truly fascinating; there were peaks whose impression could easily compete with that of elementary geometry - the basic idea of analytical geometry, the infinite series, the concepts of derivative and integral. I also had the good fortune of getting to know the essential results and methods of the entire field of the natural sciences in an excellent popular exposition, which limited itself almost throughout to qualitative aspects (Bernstein's *Popular Books on Natural Science*, a work of five or six volumes), a work that I read with breathless attention. I had also already studied some theoretical physics when, at the age of seventeen, I entered the Polytechnic Institute of Zürich as a student of mathematics and physics.

There I had excellent teachers (for example, Hurwitz, Minkowski), so that I should have been able to obtain a mathematical training in depth. I worked most of the time in the physical laboratory, however, fascinated by the direct contact with experience. The balance of the time I used, in the main, in order to study at home the works of Kirchhoff, Helmholtz, Hertz, etc. The fact that I neglected mathematics to a certain extent had its cause not merely in my stronger interest in the natural sciences than in mathematics but also in the following peculiar experience. I saw that mathematics was split up into numerous specialties, each of which could easily absorb the short lifetime granted to us. Consequently, I saw myself in the position of Buridan's ass, which was unable to decide upon any particular bundle of hay. Presumably this was because my intuition was not strong enough in the field of mathematics to differentiate clearly the fundamentally important, that which is really basic, from the rest of the more or less dispensable erudition. Also, my interest in the study of nature was no doubt stronger; and it was not clear to me as a young student that access to a more profound knowledge of the basic principles of physics depends on the most intricate mathematical methods. This dawned upon me only gradually after years of independent scientific work. True enough, physics also was divided into separate fields, each of which was capable of devouring a short lifetime of work without having satisfied the hunger for deeper knowledge. The mass of insufficiently connected experimental data was overwhelming here also. In this field, however, I soon learned to scent out that which might lead to fundamentals and to turn aside from everything else, from the multitude of things that clutter up the mind and divert it from the essentials. The hitch in this was, of course, that one had to cram all this stuff into one's mind for the examinations, whether one liked it or not. This coercion had such a deterring effect (upon me) that, after I had passed the final examination, I found the consideration of any scientific problems distasteful to me for an entire year. Yet I must say that in Switzerland we had to suffer far less under such coercion, which smothers every truly scientific impulse, than is the case in many another locality. There were altogether only two examinations; aside from these, one could just about do as one pleased. This was especially the case if one had a friend, as did I, who attended the lectures regularly and who worked over their content conscientiously. This gave one freedom in the choice of pursuits until a few months before the examination, a freedom I enjoyed to a great extent, and I have gladly taken into the bargain the resulting guilty conscience as by far the lesser evil. It is, in fact, nothing short of a miracle that the modern methods of instruction have not yet entirely strangled the holy curiosity of inquiry; for this delicate little plant, aside from stimulation, stands mainly in need of freedom; without this it goes to wrack and ruin without fail. It is a very grave mistake to think that the enjoyment of seeing and searching can be promoted by means of coercion and a sense of duty. To the contrary, I believe that it would be possible to rob even a healthy beast of prey of its voraciousness if it were possible, with the aid of a whip, to force the beast to take food continuously even when not hungry, especially if the food handed out under such coercion were to be selected accordingly.

II

In 1977 the prize was awarded to Sir Neville F. Mott for his fundamental theoretical investigations of the electronic structure of magnetic and disordered systems.

* * *

Can we really use solar energy?

by Nevill Mott.

Our universe is thought by the astronomers to be 10 to 20 billion years old, and our earth some four and a half billion. For more than half this time some kind of life has existed on the earth, but modern man, whom we have called homo sapiens, has been here only 100,000 years. But for him this may be only a beginning. The sun will not go on giving out light and heat for ever, but scientists estimate that it will last for about five billion more years and so man, or some more intelligent creatures which might evolve and take his place, could be here for very much longer than man, or indeed any living creature, has been here already. But it is only in this century that it has become clear that we are leaving to our descendants a planet in which it will be much more difficult for them to live than it is for us, because we are rapidly using up the fossil fuels, in the form of coal and oil, which are the residues of the forests which grew using the sunlight millions of years ago. It is not of course an immediate problem. Worldwide, oil will be expensive in a hundred years, coal in a few hundred, and while most of us worry about our children and grandchildren, we tend to think that future generations can look after themselves. They will have clever scientists, and we think perhaps too confidently that clever scientists will find a way to do without coal and oil.

There are only two ways to do this. One is to develop nuclear power, with all its dangers - and even our supplies of uranium are limited. Of course, if controlled nuclear fusion works, the dangers are less and the supply of fuel greater, but this is a big "if". The other way is to use the heat of the sun. The sun's energy is the source of all life. It allowed the forests to grow that we now burn as coal and oil; it causes the winds that work our windmills, and lifts up the water to fill the reservoirs which we use for hydroelectric power. For the future, we cannot grow new tropical forests for fuel; it would take far too long. In the end, if we do not depend on nuclear fusion, we shall have to generate electricity directly from the sun's radiation. There is no other way. Of course, we know how to do it; what we have not yet learned is how to do it cheaply enough to compete with coal, oil or nuclear generated electricity or how to store the electricity during the nights. The technique is now limited to powering spacecraft and to small scale generators of one kind or another. What we do not have is a major solar power station.

I do not believe we should defer the work on this until it becomes essential. One reason is, as everyone knows, it is becoming highly desirable to limit the production of carbon dioxide by the burning of fossil fuels because of the so-called greenhouse effect, the warming of the atmosphere and with it the melting of the ice-caps and the flooding of much low lying land now densely inhabited. Also the industrialists of the poorer parts of the world will demand a massive increase in power supply.

The technology of producing electricity from sunlight is called photovoltaics. It is part of silicon technology; light falls on the surface of a silicon crystal treated in a certain way (a p-n junction) and electricity is generated. But, to generate a useful amount of electricity, sunlight must be used falling on a large area - several square kilometers for a useful power station, and silicon crystals are expensive. A most promising way to do this is to use thin layers of silicon in a non-crystalline form, which can be deposited quite cheaply. That this can be done, and that the silicon can be prepared (doped) so as to act as a p-n junction was discovered only fifteen years ago by Walter Spear and Peter LeComber in Dundee, Scotland; this was not expected by the theoreticians, because a group in Leningrad under Boris Kolomiets, a decade earlier, had shown that some other non-crystalline materials, that is to say kinds of glass, could not be doped, and the present author has given a theory to explain why this was so.

This discovery set off world-wide research work into these materials, both because of their scientific interest and their various technical applications. For work on these, Japanese firms have been particularly successful, as has also Stanford Ovshinsky's firm "Energy

Conversion Devices" in Detroit, which for 40 years has specialised in applications of non-crystalline materials.

It will be realised that, if in the future we are to depend more and more on solar energy, the problem of storage has to be solved, and also that large areas of land will have to be covered with solar cells. An agency in the Federal Republic of Germany, in co-operation with the government of Saudi Arabia, has embarked on a project which may point to the solution of both. Saudi Arabia has oil, but it will not last for ever. Also it has sun, and plenty of semi-desert land. The idea is to install a large solar generator there, and use it to decompose water into hydrogen and oxygen. Research will be needed on how to do this on a large scale. Hydrogen is the perfect non-poluting fuel; one asks if it could be exported in tankers to the indus-trialised countries and burned in power stations. It is however highly explosive; few of us would like to drive a hydrogen fueled car. More probably a hydrocarbon would be used - although this when burned would be a source of carbon dioxide.

All this is for the future. I describe it to show, for those thinking of a career in science, that there are projects in the future that will serve mankind's absolutely vital needs.

III

The prize was awarded in 1981 to Arthur L. Schawlow for his contribution to the development of laser spectroscopy.

* * *

Discovering science

by Arthur L. Schawlow.

There are many good reasons for studying science. It tells us much about the way the world works, from the tiniest biological cell or even a single atom to the astronomical vastness of distant galaxies. It even gives us some measure of ability to predict the future. A spacecraft launched today can be predicted with confidence to pass close to the planet Jupiter on a certain date seven years later, and it will be there exactly when expected. By knowing and understanding science, you can not only tell how things work, but how they can be made to work if they are put together in different ways. This is the basis of engineering and increasingly also of invention.

But one of the great joys of science is to be able to discover new things and add to the store of knowledge. Science is cumulative, unlike the arts. That is, each little new piece of knowledge can be added to and used with what is already known. It is not necessary to be an Einstein or a Newton to make worthwhile additions to scientific knowledge. You do need a solid grounding in the basics of some branch of science and mathematics. Initially it should not be too specialized, both because many things will change over the course of one's working lifetime, and because you cannot always predict what will be useful. From time to time you will need to dig more deeply into a specialized area as needed. It is also good to change your field of interest from time to time, to gain a fresh perspective.

You can never know everything about even a limited area of sciences, and you don't need to. **All you need to discover something new is to recognize something that is not known.** That is something you can often find out by reading a few recent publications in that field, looking for the gaps.

Of course everyone would like his or her discoveries to be important. It is often no harder to do something important than to do something trivial. But what is an important discovery? It is one that other people can use, that opens up new possibilities for them. In some ways finding a worthwhile problem to investigate is one of the most difficult aspects of scientific research. You really cannot know in advance everything that your research may uncover, and even more you cannot know all the needs of everyone in the various fields of science and technology. Even beyond that, you are never able to tell how someone in the future may take what you have done and add to it to produce something major. Thus it is often more rewarding to concentrate on finding something new, even if not obviously important, than to wait for an earth - shaking inspiration.

In order to make a good choice of problems on which to work, it is important to develop good scientific taste, perhaps best done by working with or under someone who has already a good record of scientific discoveries. That is why graduate education in the sciences is primarily an apprenticeship, where the student does research with some guidance from a professor. If you have that opportunity, ry to work with the best researcher you can find.

Quite often good scientific questions can be found as extensions of classic problems, which have been fruitful over the years. In physics, one of these is the quests for higher temperature superconductors. Another is the search for ways of generating shorter wavelengths of electromagnetic radiation, which has led from radio waves to microwaves and then on to lasers. These classic fields sometimes stall for a time, waiting for some fresh approach, but there is usually room for further excitement in them.

When you do find some research results, it is important to check as well as you can to make sure that it is both new and true. Absolute certainty is elusive, but as a scientist you have a responsibility to be careful about the accuracy of your results and to give proper credit to the discoveries of previous workers that have led up to what you have done. At this stage, you must tell the world, or at least the part of it interested in your subject, about your findings. This is most importantly done by writing articles for publication in the scientific journals. Papers submitted to a journal will be sent to at least one, and often several, referees who often will find things to criticize in your manuscript. Such criticisms can be helpful, for referees are chosen to be somewhat expert in the area of your research, but sometimes they just fail to understand what you are trying to convey. For this reason, and for the sake of attracting readers after your work is published, it is important to learn to write clearly. You need a reasonable command of the language, but more importantly you need to have clearly in your own mind what you are trying to convey.

Scientific research can result in finding new facts about the way things behave in nature or new laws which govern their behavior. But it can also be directed toward finding new ways of putting things together to make things behave in a new way, and so providing a useful tool for science or technology. A good example of this is the laser. Since ancient times, men have dreamed of having some device which would create powerful beams of light which could be projected over great distances. Usually they were thought of as weapons which could incinerate an enemy in a flash. For instance, the Martian invaders in H.G. Wells' **The War of the Worlds** were armed with swords of heat. In the 193Os comic strip, Buck Rogers had his disintegrator pistol. However, nobody knew how to make anything like that sort of a device until the scientific principles of the way light is generated were known and understood.

In 1957, when Charles Townes and I started to study the possibility of using atoms or molecules to generate light in a controlled, coherent way the principles of quantum mechanics had been established for thirty years. It was known that most light was emitted spontaneously by excited atoms which had momentarily stored some energy. It was also known, since Albert Einstein had predicted theoretically in 1917, that excited atoms could be stimulated or forced to give up their stored energy to a light wave of the same color as they would emit spontaneously. Although Einstein was very much interested in inventions (he had worked as an examiner in the Swiss patent office), he did not invent the laser. If he thought about it at all, he might have been discouraged by the fact that stimulated emission is much smaller than ordinary absorption at any temperature, no matter how high. However, by 1957 we knew that it was sometimes possible to get away from thermal equilibrium and have a preponderance of atoms in one excited state. This had been shown a few years earlier by Townes' invention of the MASER (Microwave Amplification by Stimulated Emission of Radiation), which generated microwaves of wavelength around one centimeter by stimulating emission from excited ammonia molecules. Some of the emitted microwaves were stored in a metal box cavity resonator, and served to stimulate other excited molecules as they arrived. Thus the emissions of the molecules were synchronized to produce a steady, coherent wave.

It had been known for nearly a hundred years that light is electromagnetic waves like radio waves, but with wavelength more than ten thousand times shorter. For such tiny waves, it would not be practical to make a metal box resonator like that used for the radio waves. We were led, then to invent a structure consisting of two small mirrors facing each other at the ends of a long, pencil-like column of excited atoms. This is a good resonator for light waves that travel along the axis of the column, straight back and forth between the mirrors. If one of the mirrors is partly transparent, some of this wave comes through it to produce a highly directional output beam. This is the structure that is now used for essentially all lasers. We did not set out to find a device to generate a beam of light, but that was the natural result of the search for a suitable resonator.

We realized that there were many different substances that could be used in a laser and many different ways of exciting them. The substance chosen would determine the wavelength or color of the light produced by the laser. The same factors would influence the power output. Some lasers might give only barely detectable output beams, while others might be powerful enough to cut, drill or weld. Fortunately, we did not have to invent a laser for a particular application, but merely to show that lasers could be made to work. As it turned out, the first lasers either gave fairly high power for a short time, less than a thousandth of a second, or weak continuous beams.

Immediately some people started to call the laser a solution in search of a problem. Indeed there were potential applications of which the builders of the early lasers were not aware. For instance, one of the very first applications of lasers was for surgery on the retina of the eye, to prevent retinal detachment which can lead to blindness. Ophthalmologists had been using bright light sources such as the sun or xenon arc lamps for this purpose, but neither Townes nor I had ever heard of a detached retina. Ruby lasers were also quickly put to work drilling holes in diamonds for wire-drawing dyes. But the early, primitive lasers would not do most of the obvious jobs that people saw for light rays, such as the ancient dream of all-destroying death rays, which even now remains out of reach. Many more inventions, discoveries and technical advances were needed before lasers could reach their present stage of widespread applications.

Some of these things may seem remote to a student in high school or beginning college. But, as Sylvanus Thomson remarks in his wonderful little book, Calculus Made Easy, **"What one fool can do, another can."** By far, the most scientific discoveries are made, not by great geniuses but by ordinary mortals who happen to enjoy science enough to feel impelled to work at it.

IV

The prize was awarded in 1988 to Leon M. Lederman for the neutrino beam method and the demonstration of the doublet structure of the leptons through the discovery of the muon neutrino.

* * *

Low pay and long hours

by Leon M. Lederman

Last summer I gave a lecture entitled "Low Pay and Long Hours" to an audience of aspiring college students. My sermon had to do with the joys of the life of science. About a week later, a thoughtful letter arrived from a young undergraduate who had attended the lecture. His remarks carried the following thrust:

Dear Dr. Lederman:

I have worked hard and performed reasonably well in my academic studies, but I have yet to show real promise in any area, and despite all my efforts, I seem to be stuck in the crowd of average students. I ask myself why should I bother to work hard in graduate school and then in an academic or a governmental research career, only to discover at best one or two things that anyone else who went through the same motions might discover? Instead, with only a bachelor's degree, I can get a high-paying job with 9-to-5 hours as an actuary.

I have to admit that the possibility of becoming an actuary seems somewhat dissatisfying, because I want to pursue a career that actively promotes the welfare of humanity, and I believe that for me science will provide the best way to accomplish this goal. However, I am very discouraged by the fact that my best only seems equal to what is average, and at times I really wonder why I should bother to pursue a career in science. In your talk, you suggested that the rewards for participating in science are those moments when you make a discovery and realize that you know something that no one else knows. Yet if my past is any indication of the future, I can expect such moments in my career to occur infrequently. Indeed, it seems to me that only the people who have done well and won awards in the past succeed and win awards in the future.

(Incidentally, it seems to me that the only people who say one should not care about winning awards are those who either know they will win awards or know they do not have a chance to win awards. It is people like me - who know what greatness is and can just taste it, but cannot seem to achieve it - who care about awards. Also, when our society gives awards, it seems to focus on actual accomplishments rather than on the hard work that led to those accomplishments. It is the general lack of appreciation for people who work hard but do not succeed that tends to discourage me.)

I would like to conclude by asking you two sets of questions. My first set of questions concerns you: What motivated you to pursue a career in science? When did you recognize that you were talented in science and that you excelled in relation to your classmates or colleagues? Also, did you know that you were "Nobel Prize stuff" when you did your research 40 years ago that earned you the 1988 Nobel Prize in Physics? Perhaps most important, what has kept you motivated throughout your long and productive career?

My second set of questions concerns the "rest of us" - those aspiring young students who ,despite their efforts, have yet to distinguish themselves from the crowd of other average students: Why should we bother to pursue careers in science? What are our prospects for success, both in terms of making great scientific discoveries and in terms of pulling ourselves above the crowd? Is hard work an adequate for natural talent, or must one work hard <u>and</u> possess natural brilliance to succeed? Finally, how can we keep ourselves motivated throughout our careers, especially during the long pauses between our successes?

Sincerely, A Young Undergrad

Dear Young Undergrad,

I'm not sure I can offer clarifying guidance on so complex and subtle a series of issues. But I can tell you my own experience. In high school I was B-to-B+ student. I graduated from City College, a tough (free) college in New York, *cum laude* - that is, with a B+ average. I had a passion for science, but I knew that I was far below the class leaders in both high school and college. But they were my best friends and the ones I enjoyed being with over all others. Three years in the US Army during World War II gave me time to think, and so I started graduate school in physics with this idea: If I can do well enough to associate, inconspicuously, with my genius friends, that will result in a good enough life. My Depression-years upbringing also fashioned a fatalistic attitude toward money. In City College we used to say: "I'm going to be unemployed in chemistry. What are you planning to be unemployed in?"

Today any trained scientist or engineer who is average (B) is assured employment at reasonable wages. But what I think you must know is yourselves: What do you want out of life? If you can imagine waking up in the morning and not being able to wait to get to work; if staying up for 30 hours is for you a sign of passion and not of desire for overtime pay; if you seek real joy in the workplace, whether you're there 40 or 70 hours a week (it will still be a major occupier of your time) - if all of these are true of you, then you still have to ask: Are these "joys" worth the extra $20 000 a year you'll give up when you give up actuary work? What will the better-paying job do for your life?

I don't think you *need* the great rewards of the superscientist. Teamwork is more often than not essential. Much of the pleasure of science is a kind of voyeurism; you have to learn to take joy in other's achievements. If you struggle hard through the drudgery of the academic process and win through, then you are a scientist! Instantly, you are part of an awesome set of traditions and masters: Newton, Faraday, Einstein, Fermi.... Think of how you will describe your daily work to your children when you come home at night.

To summarize:

- Being average *now* isn't decisive. Find out about yourself. Do you dream? Do you ever have ideas, even wrong ones? Do you enjoy the scientific process, even as an observer?

- Aiming higher than you believe plausible *is* worthwhile. You can retreat later. As far as I know, we are only given one shot at the whole living process.

- Ask yourself lots of hard questions. Try to be as hard-nosed and skeptical of your own motivations as you can. What really gives you pleasure? What is really worthwhile on this planet? Why did you decide to do such and such last week? What has driven you in the past? And so on.

<u>To answer your specific questions:</u>

It was probably five years after my PhD when I began to realize that I was fairly competent. By year 10, I realized to my surprise that I was as productive as those best friends who brought me into physics, even though they understood much more than I did.

A good experiment like our neutrino research led to the pleasure of giving talks but much more obsessively it led to the next experiment.

The continuing drive? The science itself! The extra added ego boost of success. At low points (many!) it was a job, but there were my associates, students, teachers, pals worldwide, who gave me support.

I've already more or less addressed the second set of questions. Hard work - yes, it really accounts for a lot of success. Most scientists aren't brilliant. Some are even very slow. Being solid is important - that means really knowing what you have to know even if it takes a long time. Many "brilliant" guys are superficial. Determination doggedness and hard work are the characteristics that are highly valued in a group. Imagination puts the icing on the cake.

I hope some of this is useful. Good luck!

Sincerely, Leon M. Lederman

V

The prize was awarded in 1983 to Subrahmanyan Chandrasekhar for his theoretical studies of the physical process of importance to the structure and evolution of the stars.

* * *

The pursuit of science: Its motivations

by S. Chandrasekhar.

I

"The Pursuit of Science: Its Motivations" is a difficult subject because of the variety and the range of the motives of the individual scientists; they are as varied as the tastes, the temperaments, and the attitudes of the scientists themselves. Besides, their motivations are subject to substantial changes during the lifetimes of the scientists; indeed, it is difficult to discern a common denominator.

I shall restrict myself to reflections on the lives and the accomplishments of some of the great scientists of the past. Reflecting on the motives and the attitudes of great men is beset with grave semantic difficulties of communication: the words and phrases that language allows have overtones of criticism or judgment. Indeed, when speaking about others, it is well to heed Turgenev's admonition, through his character Insarov, in *On the Eve*.

We are speaking of other people; why
bring in yourself?

To set my account in its proper perspective, I shall begin with a conversation between Majorana and Fermi in the mid-1920s when both were also in their middle twenties. The conversation was reported to me by one who was present on the occasion:

Majorana: There are scientists who "happen" only once in every 500 years, like Archimedes or Newton. And there are scientists who happen only once or twice in a century, like Einstein or Bohr.

Fermi: But where do I come in, Majorana?

Majorana: Be reasonable, Enrico! I am not talking about you or me. I am talking about Einstein and Bohr.

II

For a discussion of the motivations which impel one to pursue the goals of science, no example is better than that of Johannes Kepler. Kepler's uniqueness derives from the position he occupies at the great crossroads where science shed its enveloping dogmas and the pathway was prepared for Newton. Kepler, in his inquiries, asked questions that none before him, including Copernicus, had asked. Kepler's laws differ qualitatively from earlier assumptions about planetary orbits: the assertion that planetary orbits "are ellipses" in no way resembles the kind of improvements that his predecessors had sought. In his analysis of the motions of the planets, Kepler was not preoccupied with geometrical questions; he asked, instead, questions such as: "What is the origin of planetary motions?" "If the sun is at the center of the solar system, as it is in the Copernican scheme, should not that fact be discernible in the motions and in the orbits of the planets themselves?" These are questions in physics - not in some preconceived geometrical framework.

While Kepler's approach to the problem of planetary motions was radically different from that of anyone before him, his work is preeminent for the manner in which he extracted general laws from a careful examination of observations. His examination was long and it was arduous: it took him twenty and more years of constant and persistent effort, but he never lost sight of his goal. For him, it was a search for the Holy Grail in a very literal sense.

From the outset Kepler realized that a careful study of the orbit of Mars would provide the key to planetary motions because its orbit departs from a circle the most: it had defeated Copernicus; and further that an analysis of the accurate observations of Tycho Brahe was an essential prerequisite. As Kepler wrote:

> Let all keep silence and hark to Tycho who has devoted thirty-five years to his observations.... For Tycho alone do I wait; he shall explain to me the order and arrangement of the orbits.(1)

Tycho possesses the best observations, and thus so-to-speak the material for the building of the new edifice.(2)

> ... I believe it was an act of Divine Providence that I arrived just at the time when Longomontanus was occupied with Mars. For Mars alone nables us to penetrate the secrets of astronomy which otherwise would remain forever hidden from us...(3)

Indeed, Kepler went to extraordinary lengths to acquire the observations of Tycho which he so badly needed. It is not an exaggeration to say that he committed larceny, for, as he confessed: "I confess that when Tycho died, I quickly took advantage of the absence, or lack of circumspection, of the heirs, by taking the observations under my care, or perhaps usurping them."(4) And as he explained: "The cause of this quarrel lies in the suspicious nature and bad manners of the Brahe family, but on the other hand also in my own passionate and mocking character. It must be admitted that Tengnagel had important reasons for suspecting me. I was in possession of the observations and refused to hand them over to the heirs."(5)

With Tycho's observations thus acquired, Kepler constantly asked himself: "If the sun is indeed the origin and the source of planetary motions, then how does this fact manifest itself in the motions of the planets themselves?" Noticing that Mars moved a little faster when nearest the sun than when farthest away, and "remembering Archimedes", he determined the area described by the radius vector joining the sun to the instantaneous position of Mars, as we follow it in its orbit. As Kepler wrote:

> Since I was aware that there exists an infinite number of points on the orbit and accordingly an infinite number of distances [from the sun] the idea occurred to me that the sum of these distances is contained in the *area* of the orbit. For I

> remembered that in the same manner Archimedes too divided the area of a circle into an infinite number of triangles.(6)

This was how Kepler discovered in July 1603 his law of areas, the second of his three great laws in Newton's enumeration that has been adopted ever since. The establishment of this result took Kepler some five years; for, already prior to the publication of his *Mysterium Cosmographicum* in 1596, he had sought for such a law in connection with his association of the five regular solids with the existence of the six planets known in his time.

The law of areas determined the variation of the speed along its orbit, but it did not determine the shape of the orbit. A year before he had arrived at his final statement of the law of areas, Kepler had in fact discarded circular orbits for the planets, for in October of 1602 he had written: "The conclusion is quite simply that the planet's path is not a circle - it curves inward on both sides and outward again at opposite ends. Such a curve is called an oval. The orbit is not a circle, but an oval figure."(7)

Even after concluding that the orbit of Mars is an "oval," it took Kepler an additional three years to establish that the orbit was in fact an ellipse. When that was established, he wrote:

> Why should I mince my words? The truth of Nature, which I had rejected and chased away, returned by stealth through the back door, disguising itself to be accepted. That is to say, I laid [the original equation] aside, and fell back on ellipses, believing that this was a quite different hypothesis, whereas the two, as I shall prove in the next chapter, are one and the same.... I thought and searched, until I went nearly mad, for a reason, why the planet preferred an elliptical orbit [to mine]....Ah, what a foolish bird I have been!(8)

Finally, in 1608, his *Astronomia Nova* was published. As Arthur Koestler wrote:

> It was a beautifully printed volume in folio, of which only a few copies survive. The Emperor [Rudolph] claimed the whole edition as his property and forbade Kepler to sell or give away any copy of it "without our foreknowledge and consent." But since his salary was in arrears, Kepler felt at liberty to do as he liked, and sold the whole edition to the printers. Thus the story of the *New Astronomy* begins and ends with acts of larceny, committed *ad majorem Dei gloriam*.(9)

Ten more years elapsed before Kepler discovered his third law: that the squares of the periods of revolution of any two planets is in the ratio of the cubes of their mean distances from the sun. The law is stated in his *Harmonice Mundi* completed in 1618. Here is how Kepler describes his discovery:

> On 8 March of this present year 1618, if precise dates are wanted, [the solution] turned up in my head. But I had an unlucky hand and when I tested it by computations I rejected it as false. In the end it came back to me on 15 May, and in a new attack conquered the darkness of my mind; it agreed so perfectly with the data which my seventeen years of labour on Tycho's observations had yielded, that I thought at first I was dreaming.(10)

Thus ended Kepler's long and arduous search for his Holy Grail.

In his first book, *Mysterium Cosmographicum*, Kepler exclaimed: "Oh! that we could live to see the day when both sets of figures agree with each other"(11) Twenty-two years later, after he had discovered his third law and his poignant cry had been answered, he added the following footnote to this exclamation in a reprinting of *Mysterium Cosmographicum*: "We have lived to see this day after 22 years and rejoice in it, at least I did; I trust that Maestlin and many other men will share in my joy!"(12)

III

In his novel, *The Redemption of Tycho Brahe*, Max Brod - the Czech writer who is also known for publishing, posthumously, the works of Franz Kafka - portrays and contrasts the characters of Tycho Brahe and Kepler. While Brod's novel is grossly inaccurate historically, yet his idea of what a scientist like Kepler might have been is worth quoting:

> Kepler now inspired him [Tycho] with a feeling of awe. The tranquility with which he applied himself to his labours and entirely ignored the warblings of flatterers was to Tycho almost superhuman. There was something incomprehensible in its absence of emotion, like a breath from a distant region office...(13)

Is the tranquility and the absence of emotion, which Brod attributes to his imagined Kepler, ever attained by a practicing scientist?(14)

IV

The most remarkable aspect of Kepler's pursuit of science is the constancy with which he applied himself to his chosen quest. His "was a character superior in singleness," to use

Shelley's phrase. But does the example of Kepler provide any assurance of success for a similar constancy in others? I shall consider two examples.

First, the example of Albert Michelson. His main preoccupation throughout his life was to measure the velocity of light with increasing precision. His interest came about almost by accident, when the commander of the United States Naval Academy asked him - he was then an instructor at the Academy - to prepare some lecture-demonstrations of the velocity of light. That was in 1878, and it led to Michelson's first determination of the velocity of light in 1880. On 7 May 1931, two days before he died and fifty years later, he dictated the opening sentences of a paper, posthumously published, which gave the results of his last measurement. Michelson's efforts resulted in an improvement in our knowledge of the velocity of fight from 1 part in 3,000 to 1 part in 30,000, that is, by a factor of 10. But by 1973 the accuracy had been improved to 1 part in 10^{10}, a measurement that made obsolete, beforehand, all future measurements.

Were Michelson's efforts over fifty years in vain? Leaving that question aside, one must record that, during his long career, Michelson made great discoveries derived from his delight in "light waves and their uses." Thus, his development of interferometry, leading to the first direct determination of the diameter of a star, is breathtaking. And who does not know the Michelson-Morley experiment which, through Einstein's formulation of the special and the general theory of relativity, changed irrevocably our understanding of the nature of space and time? It is a curious fact that Michelson himself was never happy with the outcome of his experiment. Indeed, it is recorded that when Einstein visited Michelson in April 1931, Mrs. Michelson felt it necessary to warn Einstein in a whisper when he arrived: "Please don't get him started on the subject of the ether." (15)

A second example is Eddington, who devoted the last sixteen years of his life to developing his "fundamental theory." Of this prodigious effort, he said a year before he died: "At no time during the past 16 years have I felt any doubt about the correctness of my theory." (16) Yet, his efforts have left no trace on subsequent developments.

Is it wise then to pursue science with a single objective and with a singleness of purpose?

V

While Kepler provides the supreme example of sustained scientific effort leading to great and fundamental discoveries, there are instances in which great thoughts have seemingly occurred spontaneously. Thus, Dirac has written that his work on Poisson brackets, and on his relativistic wave equation of the electron, were consequences of ideas "which had just come out of the blue. I could not very well say just how it had occurred to me. And I felt that work of this kind was a rather 'undeserved success'".(17)

Dirac's recollection, that the ideas underlying his work on Poisson brackets and his relativistic wave equation of the electron came to him "out of the blue," is an example of what is apparently not a unique phenomenon: those who have made great discoveries seem to remember and cherish the occasions on which they made them. Thus, Einstein has recorded that: "When in 1907 I was working on a comprehensive paper on the special theory of relativity... there occurred to me the happiest thought of my life. . . that '*for an observer falling freely from the roof of a house there exists* - at least in his immediate surroundings - no gravitational field'."(18) This "happy thought" was, of course, later enshrined in his principle of equivalence that is at the base of his general theory of relativity.

A recollection in a similar vein is that of Fermi. I once had the occasion to ask Fermi, referring to Hadamard's perceptive *Essay on the Psychology of Invention in the Mathematical Field*, what the psychology of invention in the realm of physics might be. Fermi responded by narrating the occasion of his discovery of the effect of slow neutrons on induced radioactivity. This is what he said:

> I will tell you how I came to make the discovery which I suppose is the most important one I have made. We were working very hard on the neutron-induced radioactivity and the results we were obtaining made no sense. One day, as I came to the laboratory, it occurred to me that I should examine the effect of placing a piece of lead before the incident neutrons. Instead of my usual custom, I took great pains to have the piece of lead precisely machined. I was clearly dissatisfied with something; I tried every excuse to postpone putting the piece of lead in its place. When finally, with some reluctance, I was going to put it in place, I said to myself: "No, I do not want this piece of lead here; what I want is a piece of paraffin." It was just like that, with no advance warning, no conscious prior reasoning. I immediately took some odd piece of paraffin and placed it where the piece of lead was to have been.(19)

Perhaps the most moving statement in this general context is that of Heisenberg relating the moment when the laws of quantum mechanics came to a sharp focus in his mind.

> ... one evening I reached the point where I was ready to determine the individual terms in the energy table, or, as we put it today, in the energy matrix, by what would now be considered an extremely clumsy series of calculations. When the first terms seemed to accord with the energy principle, I became rather excited, and I began to make countless arithmetical errors. As a result, it was almost three o'clock in the morning before the final result of my computations lay before me. The energy principle had held for all terms, and I could no longer doubt the mathematical consistency and coherence of the kind of quantum mechanics to which my calculations pointed. At first, I was deeply alarmed. I had the feeling that, through the surface of atomic phenomena, I was looking at a strangely beautiful interior, and felt almost giddy at the thought that I now had to probe this wealth of mathematical structure nature had so generously spread out before me. I was far too excited to sleep, and so, as a new day dawned, I made for the southern tip of the island, where I had been longing to climb a rock jutting out into the sea. I now did so without too much trouble, and waited for the sun to rise.(20)

There is no difficulty for any of us in sharing in Heisenberg's exhilaration of that supreme moment. We all know of the difficulties and paradoxes that beset the "old" Bohr-Sommerfeld quantum theory of the time; and we also know of Heisenberg's long puzzlement with Sommerfeld, Bohr, and Pauli over these difficulties and paradoxes. He had already published at that time his paper with Kramers on the dispersion theory - a theory which in many ways was the precursor to the developments that were to follow.

But what is our reaction to Heisenberg's account of his ideas on the theory of elementary particles that he developed some thirty years later, after his tragic experiences during the war and his disappointments and frustrations of the post-war years? Mrs. Heisenberg, in her book on her husband, has written: "One moonlight night we walked all over the Hainberg Mountain, and he was completely enthralled by the visions he had, trying to explain his newest discovery to me. He talked about the miracle of symmetry as the original archetype of creation, about harmony, about the beauty of simplicity, and its inner truth."(21) She quotes from one of Heisenberg's letters to her sister at this time:

> In fact, the last few weeks were full of excitement for me. And perhaps I can best illustrate what I have experienced through the analogy that I have attempted an as yet unknown ascent to the fundamental peak of atomic theory, with great efforts during the past five years. And now, with the peak directly ahead of me,

> the whole terrain of interrelationships in atomic theory is suddenly and clearly spread out before my eyes. That these interrelationships display, in all their mathematical abstraction, an incredible degree of simplicity, is a gift we can only accept humbly. Not even Plato could have believed them to be so beautiful. For these interrelationships cannot be invented; they have been there since the creation of the world.(22)

You will notice the remarkable similarity in language and in phraseology with the description of his discovery of the basic rules of quantum mechanics some thirty years earlier. But do we share in his second vision in the same way? In the earlier case, his ideas won immediate acceptance. In contrast, his ideas on particle physics were rejected and repudiated even by his long-time critic and friend, Pauli. But it is moving to read what Mrs. Heisenberg writes toward the end of her biography:

> With smiling certainty, he once said to me: "I was lucky enough to look over the good Lord's shoulder while He was at work." That was enough for him, more than enough! It gave him great joy, and the strength to meet the hostilities and misunderstandings he was subjected to in the world time and again with equanimity, and not to be led astray. (23)

VI

A different aspect of the effect a great discovery can have on its author is provided by Hideki Yukawa, in his autobiography, *The Traveler*, written when Yukawa was past fifty. One would normally have expected that an autobiography entitled *The Traveler* by one whose life, at least as seen from the outside, had been rich and fruitful, would be an account of an entire life. But Yukawa's account of his "travels" ends with the publication of his paper of 1934 describing his great discovery with the sombre note: "I do not want to write beyond this point, because those days when I studied relentlessly are nostalgic to me; and on the other hand, I am sad when I think how I have become increasingly preoccupied with matters other than study."(24)

While all of us can share in the joy of the discoveries of the great men of science, we may be puzzled by what those many, very many, less perceptive and less fortunate are to cherish and remember. Are they, like Vladimir and Estragon in Samuel Beckett's play, destined to wait for Godot? Or are they to console themselves with Milton's thought "they also serve who only stand and wait"?

VII

I now turn to the role of approbation and approval in one's pursuit of science. Wordsworth's example of Newton "voyaging through strange seas of thought alone" is not one that any of us can follow. I have referred to Eddington's lonely efforts in pursuing his fundamental theory. In spite of the confidence he expressed in the correctness of his theory, Eddington must have been deeply frustrated by the neglect of his work by contemporaries. This frustration is evident in his plaintive letter to Dingle written a few months before he died:

I am continually trying to find out why people find the procedure obscure. But I would point out that even Einstein was considered obscure, and hundreds of people have thought it necessary to explain him. I cannot seriously believe that I ever attain the obscurity that Dirac does. But in the case of Einstein and Dirac people have thought it worthwhile to penetrate the obscurity. I believe they will understand me all right when they realize they have got to do so - and when it becomes the fashion "to explain Eddington".(25)

The lack of approval by one's contemporaries can have tragic consequences when it is expressed in the form of sharp and violent criticism. Thus, Ludwig Boltzmann, greatly depressed by the violence of the attacks directed against his ideas by Ostwald and Mach, committed suicide, "a martyr to his ideas," as his grandson Flamm has written. And Georg Cantor, the originator of the modern theory of sets of points and of the orders of infinity, lost his mind because of the hatred and the animosity against him and his ideas by his teacher Leopold Kronecker: he was confined to a mental hospital for many years at the end of his life.

VIII

A case very different from the ones I have considered so far is that of Rutherford.

Consider his record. In 1897 he analyzed radioactive radiations into Â-particles, ß-rays, and G-rays in a nomenclature he then introduced. In 1902 he formulated the laws of radioactive disintegration - the first time a physical law was formulated in terms of probability and not certainty, and a forerunner of the probability interpretation of quantum mechanics that was to become universal some twenty-five years later. Between 1905 and 1907 he formulated, with Soddy, the laws of radioactive displacement and identified the Â-particle as the nucleus of the helium atom; and, with Boltwood, he initiated the determination of the ages of rocks and minerals by their radioactivity. In 1909-10, there were the experiments of

Geiger and Marsden, the discovery of the large-angle scattering of Â-rays, and Rutherford's formulation of the law of scattering and the nuclear model of the atom. Then in 1917 he effected the first laboratory transformation of atoms: that of nitrogen-14 into oxygen-17 and a proton by Â-ray bombardment. In the 192Os he was associated with the clarification of the relationship between the Â-ray and the G-ray spectra. And 1932 - the *annus mirabilis*, as R. H. Fowler called it - saw the discovery of the artificial disintegration of Lithium-7 into two Â-particles by Cockcroft and Walton, of positrons in cosmic-ray showers by Blackett, and of the neutron by Chadwick - all of them in Rutherford's Cavendish Laboratory at Cambridge. In the following year, with Oliphant, Rutherford himself discovered hydrogen-3 and helium-3.

Rutherford's attitude to his own discoveries is illustrated by his response to a remark of someone present at the moment of one of his great discoveries: "Rutherford, you are always on the crest of the wave." Rutherford responded: "I made the wave, didn't I?" Somehow from Rutherford's vantage point everything he said seems right, even including his remark, "I do not let my boys waste their time", when asked if he encouraged his students to study relativity. Rutherford was a happy warrior if ever there was one.

IX

So far, I have tried to illustrate facets of the pursuit of science by drawing on incidents in the lives of some great men of science. I return now to some more general matters, and start with an example. When Michelson was asked, towards the end of his life, why he had devoted such a large fraction of his time to the measurement of the velocity of light, he is said to have replied, "It was so much fun." There is no denying that "fun" does play a role in the pursuit of science. But the word "fun" suggests a lack of seriousness. Indeed, The *Concise Oxford Dictionary* gives to "fun" the meaning "drollery." We can be certain that Michelson did not have that meaning in mind when he described his life's main interest as "fun." What, then, is the precise meaning we are to attach to "fun" in the context in which Michelson used it? More generally, what is the role of pleasure and enjoyment?

While "pleasure" and "enjoyment" are often used to characterize one's efforts in science, failures, frustrations, and disappointments are equally, if not the more, common ingredients of scientific experience. Overcoming difficulties, undoubtedly, contributes to one's final enjoyment of success. Is failure, then, a purely negative aspect of the pursuit of science?

A remark by Dirac describing the rapid development of physics following the founding of the principles of quantum mechanics in the middle and the late 1920s is apposite in this connection:

It was a good description to say that it was a game, a very interesting game one could play. Whenever one solved one of the little problems, one could write a paper about it. It was very easy in those days for any second-rate physicist to do first-rate work. There has not been such a glorious time since then.(26)

Consider in the context of those remarks, J.J. Thomson's assessment of Lord Rayleigh in his memorial address given in Westminster Abbey:

> There are some great men of science whose charm consists in having said the first word on a subject, in having introduced some new idea which has proved fruitful; there are others whose charm consists perhaps in having said the last word on the subject, and who have reduced the subject to logical consistency and clearness. I think by temperament Lord Rayleigh belonged to the second group.(27)

This assessment by J.J. Thomson has sometimes been described as double-edged. But could one not conclude, instead, that Rayleigh by temperament chose to address himself to difficult problems and was not content to play the kind of games that Dirac describes in his characterization of the "glorious time'" in physics as a time "when second-rate physicists could do first-rate work"?

This last question concerning Rayleigh's temperament raises the further question: after a scientist has reached maturity, what are the reasons for his continued pursuit of science? To what extent are they personal? To what extent are aesthetic criteria, like the perception of order and pattern, form and substance, relevant? Are such aesthetic and personal criteria exclusive? Has a sense of obligation a role? I do not mean obligation with the common meaning of obligation to one's students, one's colleagues, and one's community. I mean, rather, obligation to science itself. And what, indeed, is the content of obligation in the pursuit of science *for* science?

Let me finally turn to a different aspect. G. H. Hardy concludes *A Mathematician's Apology* with the following statement:

> The case for my life, then, or for that of anyone else who has been a mathematician in the same sense in which I have been one, is this: that I have

> added something to knowledge, and helped others to add more: and that these somethings have a value which differs in degree only, and not in kind, from that of the creations of the great mathematicians, or of any of the other artists, great or small, who have left some kind of memorial behind them.(28)

Hardy's statement referred to mathematicians; but it is equally applicable to all scientists. I want to draw your attention particularly to his reference to wanting to leave behind some kind of memorial, that is, something that posterity may judge. To what extent, then, is the judgment of posterity - which one can never know - a conscious motivation in the pursuit of science?

X

The pursuit of science has often been compared to the scaling of mountains, high and not so high. But who amongst us can hope, even in imagination, to scale the Everest and reach its summit when the sky is blue and the air is still, and in the stillness of the air survey the entire Himalayan range in the dazzling white of the snow stretching to infinity? None of us can hope for a comparable vision of nature and of the universe around us. But there is nothing mean or lowly in standing in the valley below and awaiting the sun to rise over Kinchinjunga.

Notes

1. Letter to Maestlin, 16-26 February 1599, in *Johannes Kepler gesammelte Werke*, ed. W. von Dyck and M. Caspar (Munich, 1938), 13:289 (hereinafter cited as *Gesammelte Werke*). Quoted in Arthur Koestler, The Sleepwalkers(London: Hutchinson, 1959), p. 278.
2. Letter to Herwart, 12 July 1600, *Gesammelte Werke* 14:218 (Koestler, p. 104).
3. Kepler, *Astronomia nova*, in *Gesammelte Werke*, vol. 3, dedication (Koestler, p. 325).
4. Letter to Heyden, October 1605, *Gesammelte Werke* 15:231 (Koestler, p. 345).
5. Letter to D. Fabricius, 1 October 1602, Gesammelte Werke 15: 17 (Koestler, p.345).
6. Kepler, *Astronomia nova*, chap. 40 (Koestler, p.327).
7. Kepler, *Astronomia nova*, chap. 44 (Koestler, p. 329).
8. *Gesammelte Werke* 15:314 (Koestler, p. 333).
9. Koestler, p.340.
10. *Gesammelte Werke* 16:373 (Koestler, pp. 394-95).
11. *Gesammelte Werke*, vol. 1, chap. 21 (Koestler, p. 260).
12. Ibid., note 7.
13. Max Brod, *The Redemption of Tycho Brahe* (New York: Knopf, 1928), p. 157.
14. When he wrote *The Redemption of Tycho Brahe*, Max Brod was a member of the small circle in Prague that included Einstein and Franz Kafka. Brod's portrayal of Kepler is said to have been influenced by his association with Einstein. Thus, Walter Nernst is reported to have said to Einstein: "You are this man Kepler." See Philip Frank, *Einstein: His Life and Times* (New York: Knopf, 1947), p. 85.
15. Dorothy Michelson Livingston, The Master of Light: *A Biography of Albert A. Michelson* (Chicago: University of Chicago Press, 1974), p. 334.
16. A. Eddington, Dublin Institute of Advanced Studies A, 1943, p. 1.
17. P. A. M. Dirac,"Recollections of an Exciting Era." in *History of Twentieth Century Physics*, Proceedings of the International School of Physics,"Enrico Fermi" (New York: Academic Press, 1977), pp. 137-38.
18. A. Pais, *Subtle is the Lord* (New York: Oxford University Press, 1982), p. 178.
19. S. Chandrasekhar, *Enrico Fermi: Collected Papers*, 2 vols. (Chicago: University of Chicago Press, 1962), 2:926-27.
20. W. Heisenberg, *Physics and Beyond: Encounters and Conversations* (New York: Harper & Row, 1971), p. 61.
21. E. Heisenberg, *Inner Exile*, trans. S. Cappellari and C. Morris (Boston: Birkhäuser, 1984), p. 143.
22. Ibid., pp. 143-44.
23. Ibid., p. 157.
24. H. Yukawa, *Tobibito (The Traveler)* (Singapore: World Scientific Publishing, 1982), p. 207.
25. J. G. Growther, *British Scientists of the Twentieth Century* (London: Routledge & Kegan Paul, 1952), p. 194.
26. P. A. M. Dirac, *Directions in Physics* (New York: Wiley, 1978), p. 7.
27 R. J. Strutt, 4th Baron Rayleigh, *Life of John William Strutt, Third Baron Rayleigh*, O.M., F.R.S. (Madison: University of Wisconsin Press, 1968), p. 310.
28. G. H. Hardy, *A Mathematician's Apology* (Cambridge: Cambridge University Press, 1967), p. 151.

VI

The prize was awarded in 1981 to Nicolaas Bloembergen for his contribution to the development of laser spectroscopy.

* * *

Physics in our daily lives and physics as an intellectual adventure

by N. Bloembergen.

The world around us is full of physical phenomena accessible to direct observation without the need for expensive equipment. There is the motion of the sun and moon, the planets and the stars. There are the spectral colors of the rainbow, the flow of water, the form of droplets on glass and other surfaces, the reflection of light by a mirror or by a puddle of water, the apparent bending of a wooden stick, when it is poked at a slanting angle into a canal, the motion of a swing or a merry-go-round, or the formation of waves.

We all know that a rock can be broken into two parts, that this process can be repeated again and again, until we have grains of sand. What happens if the grain of sand is split again? How far can these divisions be pursued? A Greek philosopher by the name of Democritus, posed these questions about twenty-five hundred years ago, and proposed the concept of the atom, a particle that could not be divided further. It took the efforts of untold numbers of scientists to answer and refine the seemingly simple question raised by Democritus. Only during the last few years have scientists been able to keep one isolated atom in place and observe it for a long time. Therefore, atoms are indeed very real. We also know that such atoms can be subdivided, or ionized. Each atom is really a planetary system, in which the nucleus functions as the sun, and circling electrons play the role of planets. Large accelerating machines have enabled scientists to probe even deeper, obtaining information about fundamental particles inside the nucleus. Thus, a simple question led to very large and exciting excursions into the structure of matter by thousands of scientists over many centuries. And this exploration is still going on in research laboratories around the world.

Let us, however, return to simpler questions. Already, as a boy of six I was required to wear eyeglasses. Eight years later, during a school lesson in optics, I was very interested to learn how images are formed by lenses. Listening to a radio was at first a mystery and so was television, but one quickly takes such technological developments for granted. A curious and inquisitive person will ask questions: "How does it work?", "Why do I need an antenna?", or "What happens if I put the radio in a paper box and what happens if I put the radio in a metal box?" If one is really interested in answers to these questions, it becomes quickly apparent that a study of electricity and magnetism is a great challenge.

What attracted me to science, and in particular to physics, was the challenge of seemingly simple questions. I have spent a large part of my life studying electromagnetic properties of matter, and I am still learning.

When we were taught in school about barometric pressure and the change of melting and boiling temperatures with pressure, I understood why it takes longer to boil an egg on a mountaintop than in the valley. The question pops up: "Does it take longer to fry an egg on a mountaintop?"

What I found most fascinating as I progressed in asking questions, is that mathematics can be so helpful, and is in fact indispensable, in describing the variety of physical phenomena. The motions of the planets are described by elegant equations put forth by Newton. The equations of mechanics also describe the motion of balls, arrows, bombs, spaceships, Maxwell's equations describe the behavior of electromagnetic waves, radio, radar and light. The equations of quantum mechanics describe the motion of electrons in atoms, molecules and metals. The correspondence between mathematical equations and physical phenomena is so striking, that it is almost "uncanny". It is a source for continuing fascination and challenge for the professional scientist, but precisely this connection between mathematics and physical facts may deter many schoolchildren from learning physics.

Simple physical phenomena can, however, be helpful in illustrating basic concepts in quantitative reasoning. For example, the idea of proportionality and the meaning of graphs can be elucidated by observing the motion of balls, or falling objects, or by weighing objects by balance scales. The concept of "order of magnitude" or "powers of ten" is also very important, and can be communicated by starting with the size of a finger tip, then proceeding to the size of the arm, the human body, a house, a village, a city, a country, a continent, the earth, and extrapolating to the distance of the moon, the sun, the stars and the galaxies. To proceed in the other direction of the length scale, one could consider the size of a pinprick, the size of a microbial object under a microscope, and extrapolate to the size of molecules, atoms, and beyond to nuclei and electrons.

Go the same gamut of powers of ten in time. Begin with the oscillation of a grandfather clock, about a second. Increase the time to a minute, an hour, a day, a year, a century, geological periods, the age of the earth and the age of the universe. Proceed in the opposite direction to smaller time intervals. The period of a sound oscillation, ultrasonics or the vibration of the quartz crystal in a digital watch, down to oscillation periods of electrons in atoms. The concept of logarithmic, or exponential, scales should acquire some meaning from such discussions.

I must confess, that I never found physics easy. It is probably this intellectual challenge to probe a little deeper into the secrets of the structure of matter, that motivated my professional career. For me, the relationship between rather recondite research, involving complicated experimental apparatus and theoretical equations, and technological applications serving all mankind has probably been the most gratifying experience.

My research in 1946-1947 for the Ph.D. degree, carried out under the guidance of Professor E.M. Purcell, who shared the Nobel Prize for Physics in 1952 with F. Bloch, was concerned with the measurement of relaxation times T_1 and T_2 of the nuclear spins of protons in water and other fluids. We certainly had no inkling at that time that refinements of those techniques would eventually lead to magnetic resonance scanning of the human body. This is a major medical advance that permits the observation of blood flow, tumors, etc. in the human brain and other organs. The times T_1 and T_2 are fundamental in this important application. My subsequent work on masers, lasers and optics has also been relevant to far reaching technological developments. The interaction of focused laser beams with materials leads to operations of drilling and welding in various industries, including the manufacture of auto and jet engines. The field of surgery is revolutionized by using laser beams as a scalpel. Many surgical procedures including delicate operations on the eye, the vocal cords, and many other organs, are now routinely carried out with lasers.

The use of optical fibers, in combination with semiconductor lasers, has also revolutionized the field of communications. Optical fiber cables have been laid under the Atlantic and

Pacific ocean. They can carry the information of 40,000 telephone conversations simultaneously. It is clear that improved communications via satellite or via optical fibers provides the means to reach remote areas in Third World countries, and people living in these areas are brought into closer contact with each other and with other nations.

It is for these reasons that I intend to remain interested in the interaction of laser beams with matter, even after having reached the age of retirement.

VII

The Prize was awarded in 1956 to John Bardeen for his research on semiconductors and for the discovery of the transistor effect.

* * *

Semiconductor research leading to the point contact transistor

by J. Bardeen.

I. Introduction.

In this lecture we shall attempt to describe the ideas and experiments which lead to the discovery of the transistor effect as embodied in the point-contact transistor. ... . As we shall see, the discovery was but a step along the road of semiconductor research to which a great many people in different countries have contributed. It was dependent both on the sound theoretical foundation largely built up during the thirties and on improvement and purification of materials, particularly of germanium and silicon, in the forties. About half of the lecture will be devoted to an outline of concepts concerning electrical conduction in semiconductors and rectification at metal-semiconductor contacts as they were known at the start of our research program.

The discovery of the transistor effect occurred in the course of a fundamental research program on semiconductors initiated at the Bell Telephone Laboratories in early 1946. Semiconductors was one of several areas selected under a broad program of solid state research, of which S. O. Morgan and W. Shockley were co-heads. In the initial semiconductor group, under the general direction of Shockley, were W. H. Brattain, concerned mainly with surface properties and rectification, G. L. Pearson, concerned with bulk properties, and the writer, interested in theoretical aspects of both. Later a physical chemist, R. B. Gibney and a circuit expert, H. R. Moore joined the group and made important contributions, particularly to chemical and instrumentation problems, respectively.

It is interesting to note that although Brattain and Pearson had had considerable experience in the field prior to the war, none of us had worked on semiconductors during the war years. We were able to take advantage of the important advances made in that period in connection with the development of silicon and germanium detectors and at the same time have a fresh look at the problems. Considerable help was obtained from other groups in the Laboratories which were concerned more directly with war-time developments. Particular mention should be made of J. H. Scaff, H. C. Theuerer and R. S. Ohl.

The general aim of the program was to obtain as complete an understanding as possible of semiconductor phenomena, not in empirical terms, but on the basis of atomic theory. A sound theoretical foundation was available from work done during the thirties:

1) Wilson's quantum mechanical theory (1), based on the energy band model, and describing conduction in terms of excess electrons and holes. It is fundamental to all subsequent developments. The theory shows how the concentration of carriers depends on the temperature and on impurities.

2) Frenkel's theories of certain photoconductive phenomena (2) (change of contact potential with illumination and the photomagneto electric effect) in which general equations were introduced which describe current flow when non-equilibrium concentrations of both holes and conduction electrons are present. He recognized that flow may occur by diffusion in a concentration gradient as well as by an electric field.

3) Independent and parallel developments of theories of contact rectification by Mott (3), Schottky (4) and Davydov (5). The most complete mathematical theories were worked out by Schottky and his co-worker, Spenke. Of great importance for our research program was the development during and since the war of methods of purification and control of the

electrical properties of germanium and silicon. These materials were chosen for most of our work because they are well suited to fundamental investigations with the desired close coordination of theory and experiment. Depending on the nature of the chemical impurities present, they can be made to conduct by either excess electrons or holes.

Largely because of commercial importance in rectifiers, most experimental work in the thirties was done on copper oxide (Cu_2O) and selenium. Both have complex structures and conductivities which are difficult to control. While the theories provided a good qualitative understanding of many semiconductor phenomena, they had not been subjected to really convincing quantitative checks. In some cases, particularly in rectification, discrepancies between experiment and theory were quite large. It was not certain whether the difficulties were caused by something missing in the theories or by the fact that the materials used to check the theories were far from ideal.

In the U. S. A., research on germanium and silicon was carried out during the war by a number of university, government and industrial laboratories in connection with the development of point contact or "cats whisker" detectors for radar. Particular mention should be made of the study of germanium by a group at Purdue University working under the direction of K. Lark-Horovitz and of silicon by a group at the Bell Telephone Laboratories. The latter study was initiated by R. S. Ohl before the war and carried out subsequently by him and by a group under J. H. Scaff. By 1946 it was possible to produce relatively pure polycrystalline materials and to control the electrical properties by introducing appropriate amounts of donor and acceptor impurities. Some of the earliest work (1915) on the electrical properties of germanium and silicon was done in Sweden by Prof. C. Benedicks. Aside from intrinsic scientific interest, an important reason for choosing semiconductors as a promising field in which to work was the many and increasing applications in electronic devices, which, in 1945, included diodes, varistors and thermistors. There had long been the hope of making a triode, or an amplifying device with a semiconductor. Two possibilities had been suggested. One followed from the analogy between a metal semiconductor rectifying contact and a vacuum tube diode. If one could somehow insert a grid in the space-charge layer at the contact, one should be able to control the flow of electrons across the contact. A major practical difficulty is that the width of the space-charge layer is typically only about 10^{-4} cm. That the principle is a sound one was demonstrated by Hilsch and Pohl (6), who built a triode in an alkali-halide crystal in which the width of the space-charge layer was of the order of one centimeter. Because amplification was limited to frequencies of less than one cycle per second, it was not practical for electronic applications.

The second suggestion was to control the conductance of a thin film or slab of semiconductor by application of a transverse electric field (called the field effect). In a simple form, the slab forms one plate of a parallel plate condenser, the control electrode being the other plate. When a voltage is applied across the condenser, charges are induced in the slab. If the induced charges are mobile carriers, the conductance should change with changes of voltage on the control electrode. This form was suggested by Shockley; his calculations indicated that, with suitable geometry and materials, the effect should be large enough to produce amplification of an a. c. signal (7).

Point contact and junction transistors operate on a different principle than either of these two suggestions, one not anticipated at the start of the program. The transistor principle, in which both electrons and holes play a role, was discovered in the course of a basic research program on surface properties.

Shockley`s field effect proposal, although initially unsuccessful, had an important bearing on directing the research program toward a study of surface phenomena and surface states. Several tests which Shockley carried out at various times with J. R. Haynes, H. J. McSkimin, W. A. Yager and R. S. Ohl, using evaporated films of germanium and silicon, all gave negative results. In analyzing the reasons for this failure, it was suggested (8) that there were states for electrons localized at the surface, and that a large fraction of the induced charge was immobilized in these states. Surface states also accounted for a number of hitherto puzzling features of germanium and silicon pointcontact diodes.

In addition to the possibility of practical applications, research on surface properties appeared quite promising from the viewpoint of fundamental science. Although surface states had been predicted as a theoretical possibility, little was known about them from experiment. The decision was made, therefore, to stress research in this area. The study of surfaces initiated at that time (1946) has been continued at the Bell Laboratories and is now being carried out by many other groups as well (9).

It is interesting to note that the field effect, originally suggested for possible value for a device, has been an extremely fruitful tool for the fundamental investigation of surface states. Further, with improvements in semiconductor technology, it is now possible to make electronic amplifiers with high gain which operate on the field effect principle.

Before discussing the research program, we shall give first some general background material on conduction in semiconductors and metal-semiconductor rectifying contacts.

II. Nature of Conduction in Semiconductors.

An electronic semiconductor is typically a valence crystal whose conductivity depends markedly on temperature and on the presence of minute amounts of foreign impurities. The ideal crystal at the absolute zero is an insulator. When the valence bonds are completely occupied and there are no extra electrons in the crystal, there is no possibility for current to flow. Charges can be transferred only when imperfections are present in the electronic structure, and these can be of two types: Excess electrons which do not fit into the valence bonds can move through crystal, and holes, places from which electrons are missing in the bonds, which also behave as mobile carriers. While the excess electrons have the normal negative electronic charge, -e, holes have a positive charge, +e. It is a case of two negatives making a positive; a missing negative charge is a positive defect in the electron structure.

The bulk of a semiconductor is electrically neutral; there are as many positive charges as negative. In an intrinsic semiconductor, in which current carriers are created by thermal excitation, there are approximately equal numbers of excess electrons and holes. Conductivity in an extrinsic semiconductor results from impurity ions in the lattice. In n-type material, the negative charge of the excess electrons is balanced by a net positive space charge of impurity ions. In p-type, the positive charge of the holes is balanced by negatively charged impurities. Foreign atoms which can become positively charged on introduction to the lattice are called donors; atoms which become negatively ionized are called acceptors. Thus donors make a semiconductor n-type, acceptors p-type. When both donors and acceptors are present, the conductivity type depends on which is in excess. Mobile carriers then balance the net space charge of the impurity ions. Terminology used is listed in the table below:

Table I

Designation of conductivity type		Majority carrier	Dominan impurity ion
n-type	execss	electron; n/cm^3	donor
p-type	defect	hole; p/cm^3	acceptor

These ideas can be illustrated quite simply for silicon and germanium, which, like carbon, have a valence of four and lie below carbon in the periodic table. Both crystallize in the diamond structure in which each atom is surrounded tetrahedrally by four others with which it forms bonds. Carbon in the form of a diamond is normally an insulator; the bond structure is complete and there are no excess electrons. If ultraviolet light falls on diamond, electrons can be ejected from the bond positions by the photoelectric effect. Excess electrons and holes so formed can conduct electricity; the crystal becomes photoconductive.

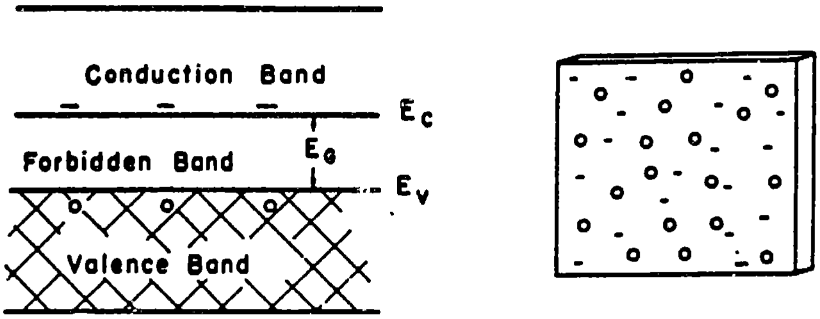

Fig 1. Energy level diagram of an intrinsic semiconductor. There is a random distribution of electrons and holes in equal numbers.

The energy required to free an electron from a bond position so that it and the hole left behind can move through the crystal is much less in silicon and germanium than for diamond. Appreciable numbers are released by thermal excitations at high temperatures; this gives intrinsic conductivity.

Impurity atoms in germanium and silicon with more than four valence electrons are usually donors, those with less than four acceptors. For example, group V elements are donors, group III elements acceptors. When an arsenic atom, a group V element, substitutes for germanium in the crystal, only four of its valence electrons are required to form the bonds. The fifth is only weakly held by the forces of coulomb attraction, greatly reduced by the high dielectric constant of the crystal. The energy required to free the extra electron is so small that the arsenic atoms are completely ionized at room temperature. Gallium, a typical group III acceptor has only three valence electrons. In order to fill the four bonds, Ga picks up another electron and enters the crystal in the form of a negative ion, Ga^-. The charge is balanced by a free hole.

While some of the general notions of excess and defect conductivity, donors and acceptors, go back earlier, Wilson (1) was the first to formalize an adequate mathematical theory in terms of the band picture of solids. The band picture itself, first applied to metals, is a consequence of an application of quantum mechanics to the motion of electrons in the periodic potential field of a crystal lattice. Energy levels of electrons in atoms are discrete. When the atoms are combined to form a crystal, the allowed levels form continuous bands. When a band is completely occupied, the net current of all of the electrons in the band is zero. Metals have incompletely filled bands. In insulators and semiconductors, there is an energy gap between the highest filled band and the next higher allowed band of levels, normally unoccupied.

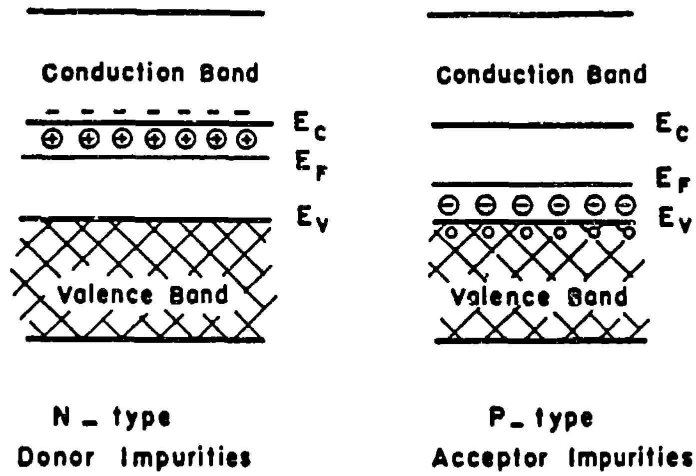

Fig. 2. Energy level diagrams for n- and p-type semiconductors.

The relations are most simply illustrated in terms of an energy level diagram of the crystal. In Fg. 1 is shown a schematic energy level diagram of an intrinsic semiconductor. Electrons taking part in the chemical bonds form a continuous band of levels called the valence band. Above these is an energy gap in which there are no allowed levels in the ideal crystal, and then another continuous band of levels called the conduction band. The energy gap, E_G, is the energy required to free an electron from the valence bonds. Excess, or conduction, electrons have energies in the lower part of the conduction band. The very lowest state in this band, E_C, corresponds to an electron at rest, the higher states to electrons moving through the crystal with additional energy of motion. Holes correspond to states near the top of the valence band, E_V, from which electrons are missing. In an intrinsic semiconductor, electrons and holes are created in equal numbers by thermal excitation of electrons from the valence to the conduction band, and they are distributed at random through the crystal.

In an n-type semiconductor, as illustrated in Fig. 2a, there is a large number of electrons in the conduction band and very few holes in the valence band. Energy levels corresponding to electrons localized around group V donor impurity atoms are typically in the forbidden gap and a little below the conduction band. This means that only a small energy is required to ionize the donor and place the electron removed in the conduction band. The charge of the electrons in the conduction band is compensated by the positive space charge of the donor ions. Levels of group III acceptors (Fig. 2b) are a little above the valence band. When occupied by thermal excitation of electrons from the valence band, they become negatively charged. The space charge of the holes so created is compensated by that of the negative acceptor ions.

Occupancy of the levels is given by the position of the Fermi level, E_F. The probability, f, that a level of energy E is occupied by an electron is given by the Fermi-Dirac function:

$$f = \frac{1}{1+\exp\left[\frac{E-E_F}{kT}\right]}$$

The energy gap in a semiconductor is usually large compared with thermal energy, kT (≈ .025 eV at room temperature), so that for levels well above E_F one can use the approximation

$$f \approx \exp\left[-\frac{E-E_F}{kT}\right] .$$

For levels below E_F, it is often more convenient to give the probability

$$f_P = 1 - f = \frac{1}{1+\exp\left[\frac{E_F-E}{kT}\right]}$$

that the level is unoccupied, or "occupied by a hole". Again, for levels well below E_F,

$$f_P \approx \exp\left[-\frac{E_F-E}{kT}\right] .$$

The expressions for the total electron and hole concentrations (number per unit volume), designated by the symbols *n* and *p* respectively, are of the form

$$n = N_C \exp\left[-\frac{E_C - E_F}{kT}\right] ,$$

$$p = N_V \exp\left[-\frac{E_F - E_V}{kT}\right] ,$$

where N_C and N_V vary slowly with temperature compared with the exponential factors. Note that the product $n{\cdot}p$ is independent of the position of the Fermi level and depends only on the temperature:

$$n \cdot p = n_i^2 = N_C \cdot N_V \exp\left[-\frac{E_C - E_V}{kT}\right] , \ or$$

$$n \cdot p = n_i^2 = N_C \cdot N_V \exp\left[-\frac{E_G}{kT}\right].$$

Here n_i is the concentration in an intrinsic semiconductor for which $n = p$.

In an n-type semiconductor, the Fermi level is above the middle of the gap, so that $n \approx p$. The value of n is fixed by the concentration of donor ions, $N_d{}^+$, so that there is electric neutrality:

$$n - p = N_d{}^+.$$

The minority carrier concentration, p, increases rapidly with temperature and eventually a temperature will be reached above which n and p are both large compared with $N_d{}^+$ and the conduction is essentially intrinsic. Correspondingly in a p-type semiconductor, in which there are acceptor ions, $p \approx n$, and the Fermi level is below the center of the gap.

The Fermi level is equivalent to the chemical potential of the electrons. If two conductors are electrically connected together so that electrons can be transferred, the relative electrostatic potentials will be adjusted so that the Fermi levels of the two are the same. If the n- and p-type materials of Fig. 2 are connected, a small number of electrons will be transferred from the n-type to the p-type. This will charge the p-type negatively with respect to the n-type and raise the electrostatic potential energy of the electrons accordingly. Electron transfer will take place until the energy levels of the p-type material are raised relative to those of the n-type by the amount required to make the Fermi levels coincide.

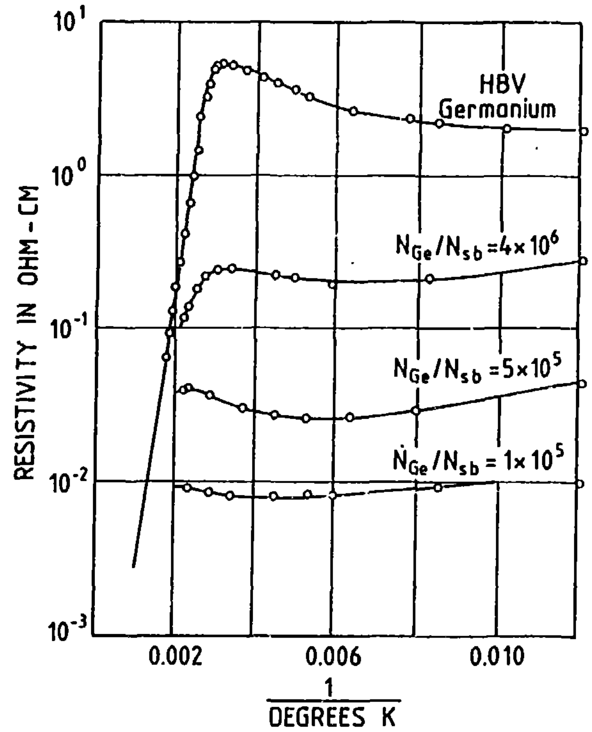

Fig. 3. Conductivity versus 1/T for germanium with antimony added as a donor impurity.

The amount of impurity required to make significant changes in the conductivity of germanium or silicon is very small. There is given in Fig. 3 a plot, on a log scale, of the resistivity versus 1/T for specimens of germanium with varying amounts of antimony, a donor impurity. This plot is based on some measurements made by Pearson several years ago (11). The purest specimens available at that time had a room temperature resistivity of about 10 - 20 ohm cm, corresponding to about one donor atom in 10^8 germanium atoms. This material (H. B. V.) is of the sort which was used to make germanium diodes which withstand a high voltage in the reverse direction (High Back Voltage) and also used in the first transistors. The purest material available now corresponds to about one donor or acceptor in 10^{10}. The resistivity drops, as illustrated, with increasing antimony concentration; as little as one part in 10^7 makes a big difference. All specimens approach the intrinsic line corresponding to pure germanium at high temperatures.

Conduction electrons and holes are highly mobile, and may move through the crystal for distances of hundreds or thousands of the interatomic distance, before being scattered by thermal motion or by impurities or other imperfections. This is to be understood in terms of the wave property of the electron; a wave can travel through a perfect periodic structure without attenuation. In treating acceleration in electric or magnetic fields, the wave aspect can often be disregarded, and electrons and holes thought of as classical particles with an effective mass of the same order, but differing from the ordinary electron mass. The effective mass is often anisotropic, and different for different directions of motion in the crystal. This same effective mass picture can be used to estimate the thermal motion of the gas of electrons and holes. Average thermal velocities at room temperature are of the order of 10^7 cm/sec. Scattering can be described in terms of a mean free path for the electrons and holes. In relatively pure crystals at ordinary temperatures, scattering occurs mainly by interaction with the thermal vibrations of the atoms of the crystal. In less pure crystals, or in relatively pure crystals at low temperatures, the mean free path may be determined by scattering by impurity atoms. Because of the scattering, the carriers are not uniformly accelerated by an electric field, but attain an average drift velocity proportional to the field. Ordinarily the drift velocity is much smaller than the average thermal velocity. Drift velocities may be expressed in terms of the mobilities μ_n and μ_p of the electrons and holes respectively.[1)]

In an electric field E,

$$(V_d)_n = - \mu_n E$$
$$(V_d)_p = + \mu_p E$$

Because of their negative charge, conduction electrons drift oppositely to the field. Values for pure germanium at room temperature are μ_n = 3800 cm^2/volt sec; μ_p = 1800 cm^2/vclt sec. This means that holes attain a drift velocity of 18oo cm/sec in a field of one volt/cm.

Expressions for the conductivity are:

n-type:	σ_n	$= ne\mu_n$
p-type:	σ_p	$= pe\mu_p$
intrinsic:	σ	$= ne\mu_n + np\mu_p$

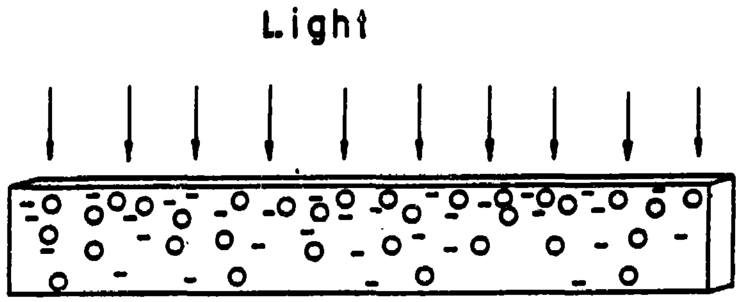

Fig.4. Schematic diagram of diffusive flow of electrons and holes created near the surface by absorption of light

[1)]A subscribt n (referring to negativ charge) is used for conduction electrons, p (positive) for holes.

It is not possible to determine n and μ_n separately from measurements of the conductivity alone. There are several methods to determine the mobility; one which has been widely used is to measure the Hall coefficient in addition to the conductivity. As part of the research program at the Bell Laboratories, Pearson made Hall and resistivity measurements over a wide range of temperatures of silicon containing varying amounts of boron (a group III acceptor) and of phosphorus (a group V donor). Analysis of the data (1o) gave additional confirmation of the theory we have outlined. Similar measurements on germanium were made about the same time by Lark-Horovitz and co-workers, and more recently more complete measurements on both materials have been made by other groups. The result of a large amount of experimental and theoretic work has been to confirm the Wilson model in quantitative detail.

Carriers move not only under the influence of an electric field, but also by diffusion; the diffusion current is proportional to the concentration gradient. Expressions for the particle current densities of holes and electrons, respectively, are

$$j_p = p\mu_p E - D_p \operatorname{grad} p$$
$$j_n = n\mu_n E - D_n \operatorname{grad} n$$

Einstein has shown that mobilities and diffusion coefficients are related:

$$\mu = \frac{e}{kT} \cdot D ,$$

where k is Boltzmann's constant. Diffusion and conduction currents both play an important role in the transistor.

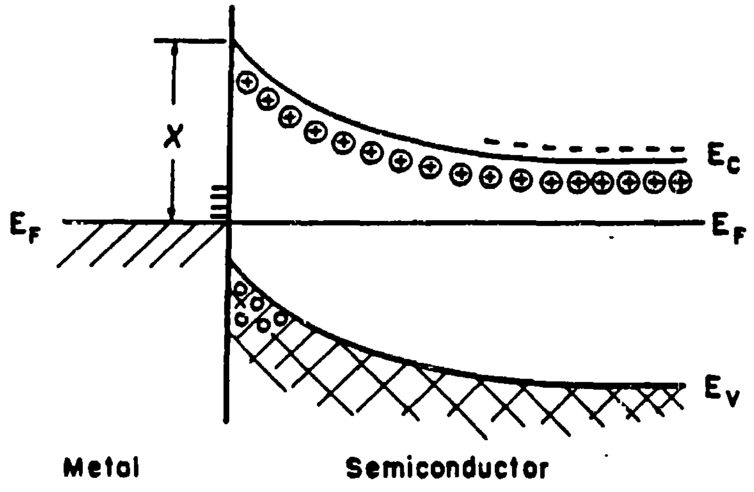

Fig. 5. Equilibrium energy level diagram for a metal-semiconductor rectifying contact along a line perpendicular to the interface. Variations in the energy bands of the semiconductor result from changes in electrostatic potential due to the layer of uncompensated space-charge. The overal change in potential from the surface to the interior is such as to bring the Fermi level in the interior of the semiconductor into coincidance with that of the metal. In this example, there is an inversion from n-type conductance in the bulk to p-type at the surface.

The diffusion term was first considered by Wagner in his theory of oxidation of metals. The equations were worked out more completely by Frenkel (2) in an analysis of the diffusive flow which occurs when light is absorbed near one face of a slab, as shown schematically in Fig. 4. The light quanta raise electrons from the valence to the conduction bands, creating conduction electrons and holes in equal numbers. These diffuse toward the interior of the slab. Because of recombination of conduction electron and hole pairs, the concentration drops as the diffusion occurs. Frenkel gave the general equations of flow when electrons and holes are present in attempting to account for the Dember effect (change in contact potential with light) and the photomagneto-electric (PME) effect. The latter is a voltage analogous to a Hall voltage observed between the ends of a slab in a transverse magnetic field (perpendicular to the paper in the diagram). The Dember voltage was presumed to result from a difference of mobility, and thus of diffusion coefficient, between electrons and holes. Electrical neutrality requires that the concentrations and thus the concentration gradients be the same. Further, under steady state conditions the flow of electrons to the interior must equal the flow of holes, so that there is no net electrical current. However, if D_n is greater than D_p, the diffusive flow of electrons would be greater than that of holes. What happens is that an electric field, E, is introduced which aids holes and retards the electrons so as to equalize the flows. The integral of E gives a voltage difference between the surface and the interior and thus a change in contact potential. As we will mention later, much larger changes in contact potential with light may come from surface barrier effects.

III. Contact Rectifiers.

In order to understand how a point-contact transistor operates, it is necessary to know some of the features of a rectifying contact between a metal and semiconductor. Common examples are copper oxide and selenium rectifiers and germanium and silicon point-contact diodes which pass current much more readily for one direction of applied voltage than the opposite. We shall follow Schottky's picture (4), and use as an illustration a contact to an n-type semiconductor. Similar arguments apply to p-type rectifiers with appropriate changes of sign of the potentials and charges. It is most convenient to make use of an energy level diagram in which the changes in energy bands resulting from changes in electrostatic potential are plotted along a line perpendicular to the contact are shown, as in Fig. 5. Rectification results from the potential energy barrier at the interface which impedes the flow of electrons across the contact.

The Fermi level of the metal is close to the highest of the normally occupied levels of the conduction band. Because of the nature of the metal-semiconductor interface layers, a relatively large energy, χ , perhaps of the order of o.5 eV, is required to take an electron from the Fermi level of the metal and place it in the conduction band in the semiconductor. In the interior of the semiconductor, which is electrically neutral, the position of the Fermi level relative to the energy bands is determined by the concentration of conduction electrons, and thus of donors. In equilibrium, with no voltage applied, the Fermi levels of the metal and semiconductor must be the same. This is accomplished by a region of space charge adjacent to the metal in which there is a variation of electrostatic potential, and thus of potential energy of the electron, as illustrated.

In the bulk of the semiconductor there is a balance between conduction electrons and positive donors. In the barrier region which is one of high potential energy for electrons, there are few electrons in the conduction band. The uncompensated space charge of the donors is balanced by a negative charge at the immediate interface. It is these charges, in turn, which produce the potential barrier. The width of the space charge region is typically of the order of 10^{-5} to 10^{-4} cm.

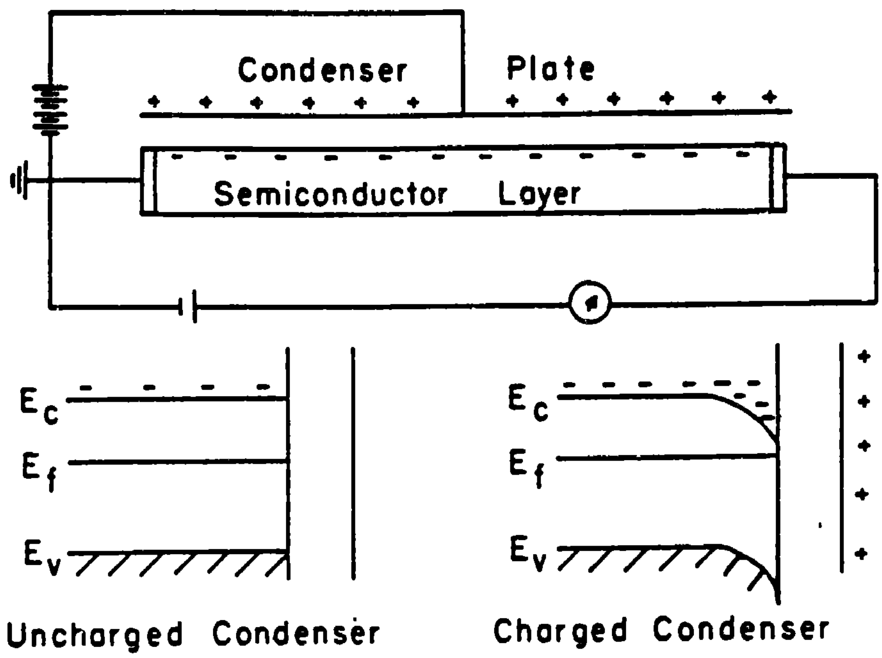

Fig. 6. Schematic diagram of a field effect experiment for an n-type semiconductor with no surface states

When a voltage is applied, most of the drop occurs across the barrier layer. The direction of easy flow is that in which the semiconductor is negative relative to the metal. The bands are raised, the barrier becomes narrower, and electrons can flow more easily from the semiconductor to the metal. In the high resistance direction, the semiconductor is positive, the bands are lowered relative to the metal, and the barrier is broadened. The current of electrons flowing from the metal is limited by the energy barrier, χ , which must be surmounted by thermal excitation.

If χ is sufficiently large, the Fermi level at the interface may be close to the valence band, implying an inversion from n-type conductivity in the bulk to p-type near the contact. The region of hole conduction is called, following Schottky, an inversion layer. An appreciable part of the current flow to the contact may then consist of minority carriers, an this case holes. An important result of the research program at the Bell Laboratories after the war was to point out the significance of minority carrier flow.

IV. Experiments on Surface States.

We have mentioned in the introduction that the negative result of the field effect experiment was an important factor in suggesting the existence of surface states on germanium and silicon and directing the research program toward a study of surface properties. As is shown in Fig. 6, the experiment consists of making a thin film or slab one plate of a parallel plate condenser and then measuring the change in conductance of the slab with changes in voltage applied across the condenser. The hypothetical case illustrated is an n-type semiconductor with no surface states. When the field plate is positive, the negative charge induced on the semiconductor consists of added electrons in the conduction band. The amount of induced charge can be determined from the applied voltage and measured capacity of the sytem. If the mobility is known, the expected change in conductance can be calculated readily.

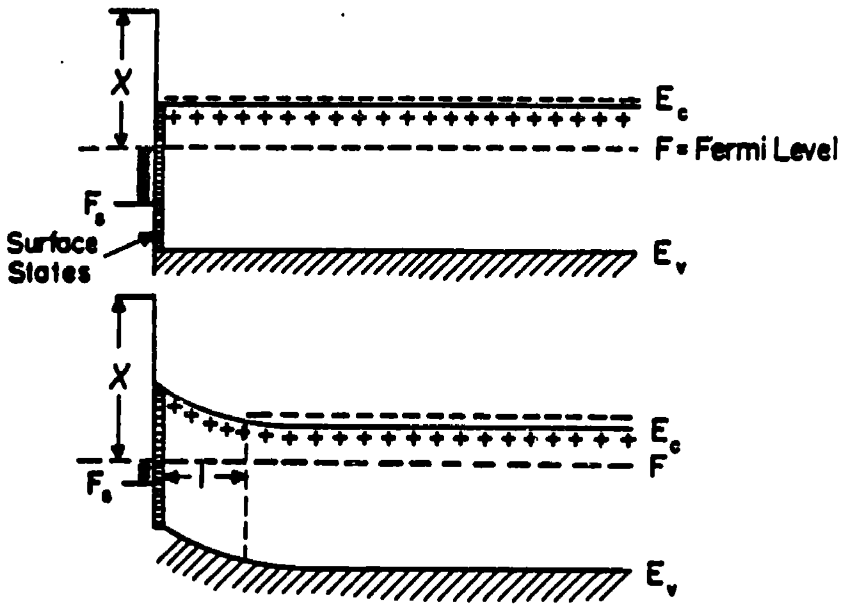

Fig. 7. Formation of a space-charger barrier layer at the free surface of a semiconductor.

When experiments were performed on evaporated films of germanium and silicon, negative results were obtained; in some cases the predicted effect was more than one thousand times the experimental limit of detection. Analysis indicated that a large part of the discrepancy, perhaps a factor of 50 to 100, came from the very low mobility of electrons in the films as compared with bulk material. The remaining was attributed to shielding by surface states.

It was predicted that if surface states exist, a barrier layer of type found at a metal contact might be found at the free surface of a semiconductor. The formation of such a layer is illustrated schematically in Fig. 7. Occupancy of the surface levels is determined by the position of the Fermi level at the surface. In the illustration, it is presumed that the distribution of surface states is such that the states themselves would be electrically neutral if the Fermi level crossed at the position F_S relative to the bands. If there is no surface barrier, so that the Fermi level crosses the surface above F_S, there are excess electrons and a net negative charge in the surface states. When the surface as whole is neutral, a barrier layer is formed such that the positive charge in the layer is compensated by the negative surface states charge. If the density of surface states is reasonably high, sufficient negative charge is obtained with the Fermi level crossing only slightly above F_S.

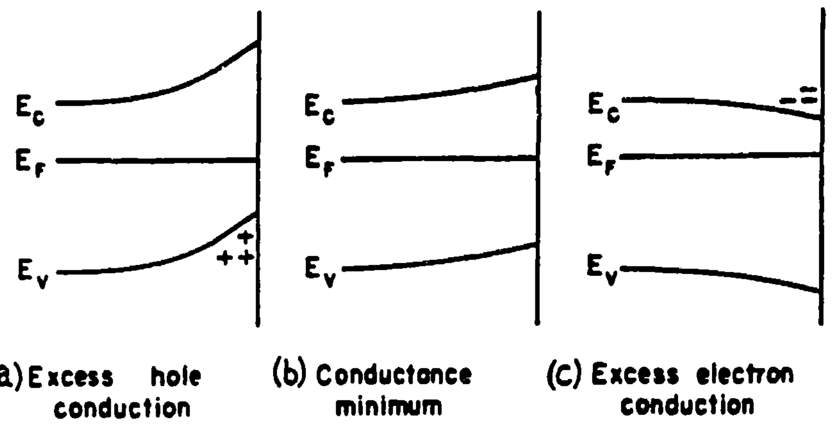

Fig. 8. Types of barrier layers which may exist at the free surface of an n-type semiconductor: (a) excess conductance from an inversion layer of p-type conductivity, (b) near the minimum surface conductance, (c) excess conductance from an accumulation layer of electrons.

Types of barriers which may exist at the surface of an n-type semiconductor are illustrated in Fig. 8. On the left (a) the energy bands are raised at the surface so as to bring the valence band close to the Fermi level. An inversion layer of opposite conductivity type is formed, and there is excess conductance from mobile holes in the layer. Negative charge on the surface proper is balanced by the charge of holes and of fixed donor ions in the barrier region. In (b) the bands are raised at the surface, but not enough to form a barrier layer. The surface resistance is near a maximum. In (c), the bands bend down so as to form an accumulation layer of excess electron conductance near the surface. The charge on the surface proper is now positive, and is balanced by the negative charge of the excess electrons in the layer.

The postulated existence of surface states and surface barrier layers on the free surface of germanium and silicon accounted for several properties of germanium and silicon which had hitherto been puzzling (8). These included (1) lack of dependence of rectifier characteristics on the work function of the metal contact, (2) current voltage characteristics of a contact made with two pieces of germanium, and (3) the fact that there was found little or no contact potential difference between n- and p-type germanium and between n- and p-type silicon.

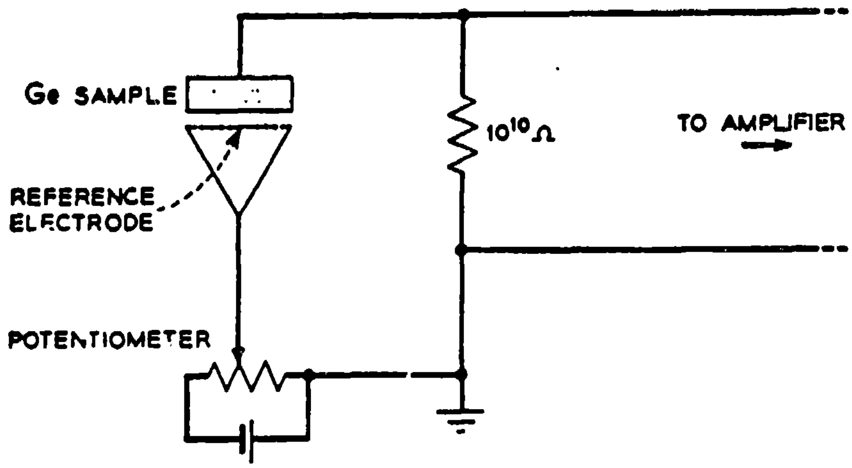

Fig. 9. Schematic diagram of apparatus used by Brattain to measure contact potential and change of contact potential with light.

While available evidence for surface states was fairly convincing, it was all of an indirect nature. Further, none of the effects gave any evidence about the height of the surface barrier and of the distribution of surface states. A number of further experiments which might yield more concrete evidence about the surface barrier was suggested by Shockley, Brattain and myself. Shockley predicted that a difference in contact potential would be found between n- and p-type specimens with large impurity concentration. A systematic study of Brattain and Shockley (12) using silicon specimens with varying amounts of donor and acceptor impurities showed that this was true, and an estimate was obtained for the density of surface states. Another experiment which indicated the presence of a surface barrier was a measurement of the change in contact potential with illumination of the surface. This is just the Dember effect, which Frenkel had attempted to account for by the difference in mobilities of the electrons and holes generated by the light and diffusing to the interior. It was found (13) that the change is usually much larger and often of the opposite sign than predicted by Frenkel's theory, which did not take into account a surface barrier.

Some rather difficult experiments which at the time gave negative results much later have been carried out successfully by improved techniques, ...

Apparatus used by Brattain to measure contact potential and change in contact potential with illumination is shown in Fig. 9. The reference electrode, generally platinum, is in the form of a screen so that light can pass through it. By vibrating the electrode, the contact potential itself can be measured by the Kelvin method. If light chopped at an appropriate frequency falls on the surface and the electrode is held fixed, the change with illumination can be measured from the alternating voltage developed across the condenser. In the course of the study, Brattain tried several ambient atmospheres and different temperatures. He observed a large effect when a liquid dielectric filled the space between the electrode and semiconductor surface. He and Gibney then introduced electrolytes, and observed effects attributed to large changes in the surface barrier with voltage applied across the electrolyte. Evidently ions piling up at the surface created a very large field which penetrated through the surface states.

V. Experiments on Inversion Layers.

Use of an electrolyte provided a method for changing the surface barrier, so that it should be possible to observe a field effect in a suitable arrangement. We did not want to use an evaporated film because of the poor structure and low mobility. With the techniques available at the time, it would have been difficult to prepare a slab from bulk material sufficiently thin to observe a sizable effect. It was suggested that one could get the effect of a thin film in bulk material by observing directly the flow in an inversion layer of oppo-site conductivity type near the surface. Earlier work of Ohl and Scaff indi-cated that one could get an inversion layer of n-type conductivity on p-type silicon by suitably oxidizing the surface. If a point contact is made which rectifies to the p-type base, it would be expected to make low resistance contact to the inversion layer.

The arrangement which Brattain and I used in the initial tests is shown in Fig. 10. The point - contact was surrounded by but insulated from a drop of electrolyte. An electrode in the electrolyte could be used to apply a strong field at the semiconductor surface in the vicinity of the contact. The reverse, or high resistance direction is that in which point is positive relative to the block. Part of the reverse current consists of electrons flowing through the n-type inversion layer to the contact. It was found that the magnitude of this current could be changed by applying a voltage on the electrolyte probe, and thus, by the field effect, changing the conductance of the inversion layer. Since under static conditions only a very small current flowed through the electrolyte, the set-up could be used as an amplifier. In the initial tests, current and power amplification, but not voltage amplification, was observed. As predicted from the expected decrease in number of electrons in the inversion layer, a

·negative voltage applied to the probe was found to decrease the current flowing in the reverse direction to the contact. It was next decided to try a similar arrangement with a block of n-type germanium. Although we had no prior knowledge of a p-type inversion layer on the surface, the experiments showed definitely that a large part of the reserve current consisted of holes flowing in an inversion layer near the surface. A positive change in voltage on the probe decreased the reverse current. Considerable voltage as well as current and power amplification was observed.

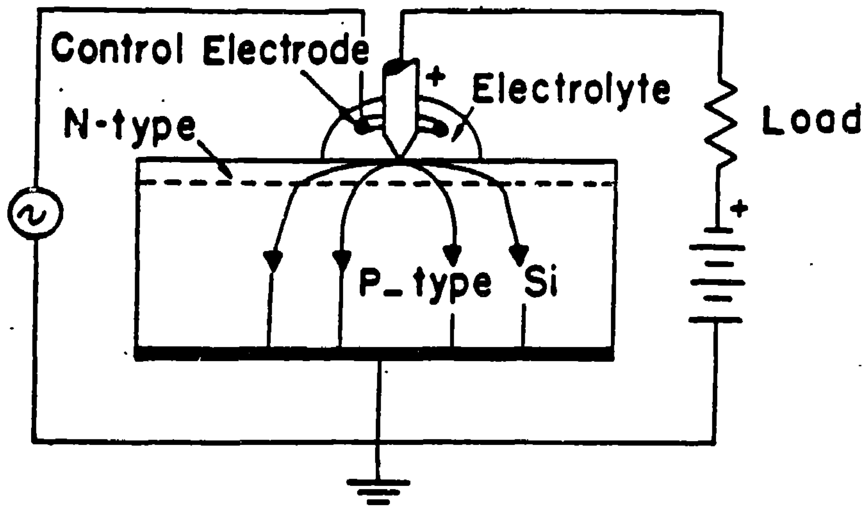

Fig 10. Diagram of experiment used to observe effect of the field produced by an electrolyte on an inversion layer of n-type conductance at the surface of a p-type silicon block. Negative potential applied to the probe in the electrolyte decreases the number of electrons in the inversion layer and thus the current of electrons flowing to the point contact biased in the reverse direction. Arrows indicate the convential direction of current flow; electrons move in the opposite direction.

Because of the long time constants of the electrolyte used, amplification was obtained only at very low frequencies. We next tried to replace the electrolyte by a metal control electrode insulated from the surface by either a thin oxide layer or by a rectifying contact. A surface was prepared by Gibney by anodizing the surface and then evaporating several gold spots on it. Although none made the desired high resistance contact to the block, we decided to see what effects would be obtained. A point contact was placed very close to one of the spots and biased in the reverse direction (See Fig. 11). A small effect on the reverse current was observed when the spot was biased positively, but of opposite direction to that observed with the electrolyte. An increase in positive bias increased rather than decreased the reverse current to the point contact. The effect was large enough to give some voltage, but no power amplification. This experiment suggested that holes were flowing into the germanium surface from the gold spot, and that the holes introduced in this way flowed into the point contact to enhance the reverse current. This was the first indication of the transistor effect.

It was estimated that power amplification could be obtained if the metal contacts were spaced at distances of the order of .oo5 cm. In the first attempt, which was successful, contacts were made by evaporating gold on a wedge, and then separating the gold at the point of the wedge with a razor blade to make two closely spaced contacts. After further experimentation, it appeared that the easiest way to achieve the desired close separation was to use two appropriately shaped point contacts placed very close together. Success was achieved in the first trials; the point contact transistor was born (14).

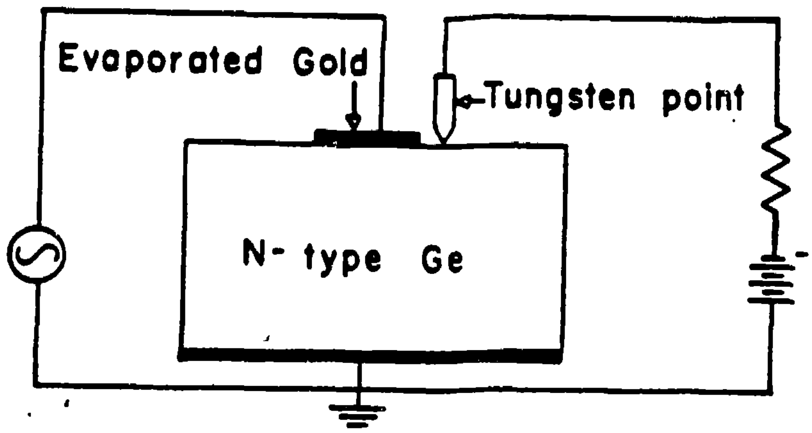

Fig. 11. Diagram of experiment in which the transistor effect was first observed. Positive voltage applied to the gold spot introduced holes into the n-type germanium block which flowed to the point contact biased in the reverse direction. It was found that an increase in positive voltage increased the reverse current. When connected across a high impedance, the change in voltage of the point contact was larger than the change at the gold spot, both measured relative to the base electrode.

It was evident from the experiments that a large part of both the forward and reverse currents from a germanium point contact is carried by minority carriers, in this case holes. If this fact had been recognized earlier, the transistor might have come sooner.

Operation of a point contact transistor is illustrated in Fig. 12. When operated as an amplifier, one contact, the emitter, is biased with a d-c voltage in the forward direction, the second, the collector, in the negative or high resistance direction. A third contact, the base electrode, makes a low resistance contact to the block. A large part of the forward current consists of holes flowing into the block. Current from the collector consists in part of electrons flowing from the contact and in part of holes flowing toward the contact. The collector current produces an electric field in the block which is in such a direction as to attract holes introduced at the emitter. A large part of the emitter current, introduced at low impedance flows in the collector circuit. Biased in the reverse direction, the collector has high impedance and can be matched to a high impedance load. There is thus a large voltage amplification of an input signal. It is found (14) that there is some current amplification as well, giving an overall power gain of 20 db or more. An increase in hole current at the collector effects the barrier there in such a way as to enhance the current of electrons flowing from the contact.

The collector current must be sufficiently large to provide an electric field to attract the holes from the emitter. The optimum impedance of the collector is considerably less than that of a good germanium diode in the reverse direction. In the first experiments, it was attempted to achieve this by treating the surface so as to produce a large inversion layer of p-type conductivity on the surface. In this case, a large fraction of the hole current may flow in the inversion layer. Later, it was found that better results could be obtained by electrically forming the collector by passing large current pulses through it. In this case the surface treatment is less critical, and most of the emitter current flows through the bulk. Studies of the nature of the forward and reverse currents to a point contact to germanium were made by making probe measurements of the variation of potential in the vicinity of the contact (15). These measurements showed a large increase in conductivity when the contact was biased in the forward direction and in some cases evidence for a conducting inversion layer near the surface when biased in the reverse direction.

Before it was established whether the useful emitter current was confined to an inversion layer or could flow through the bulk, Shockley (16) proposed a radically different design for a transistor based on the latter possibility. This is the junction transistor design in which added minority carriers from this emitter diffuse through a thin base layer to the collector. Independently of this suggestion, Shive (17) made a point contact transistor in which the emitter and collector were on opposite faces of a thin slab of germanium. This showed definitely that injected minority carriers could flow for small distances through bulk material. While transistors can be made to operate either way, designs which make use of flow through bulk material have been most successful junction transistors have superseded point contact transistors for most applications.

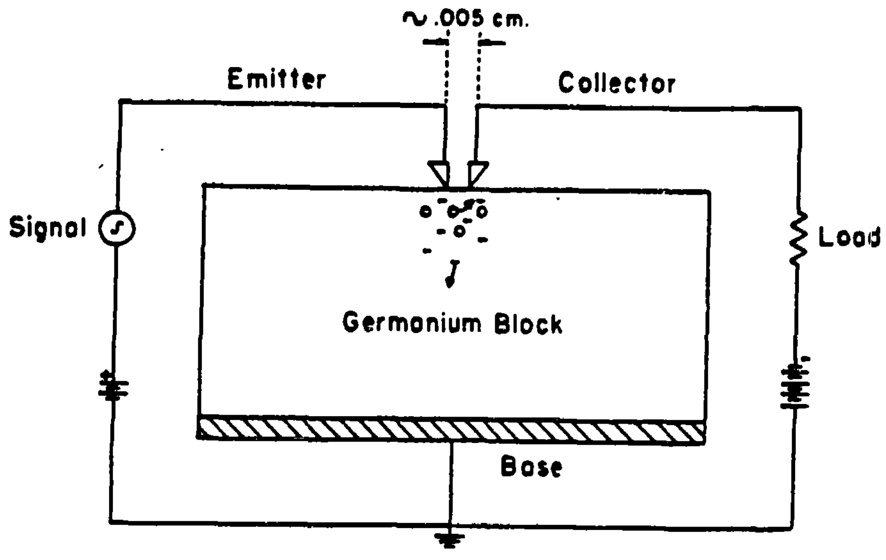

Fig. 12. Schematic diagram of point contact.

Following the discovery of the transistor effect, a large part of research at the Bell Laboratories was devoted to a study of flow on injected minority carriers in bulk material. Much of this research was instigated by Shockley, ...

Research on surface properties of germanium and silicon, suspended for some time after 1948 because of the pressure of other work, was resumed later on by Brattain and others, and is now a flourishing field of activity with implications to a number of scientific fields other than semiconductors such as adsorption, catalysis, and photoconductivity. ... .

It is evident that many years of research by a great many people both before and after the discovery of the transistor effect has been required to bring our knowledge of semiconductors to its present development. We were fortunate enough to be involved at a particularly opportune time and to add another small step in the control of nature for the benefit of mankind. In addition to my colleagues and to others mentioned in the lecture, I would like to express my deep gratitude to Drs. M. J. Kelley and Ralph Bown for the inspired leadership of Laboratories when this work was done.

REFERENCES

1. A. H. Wilson, Proc. Roy. Soc. (London) A 133, 458 (1931); A 134, 277 (1931); A 136, 487 (1932).
2. J. Frenkel, Physik Z. Sowjetunion 8, 185 (1935).
3. N. F. Mott, Proc. Roy. Soc. (London) A 171, 27 (1939).
4. W. Schottky, Z. Physik 113, 367 (1939); 118, 539 (1942).
5. B. Davydov, J. Tech. Phys. USSR, 5, 87 (1938).
6. R. Hilsch and R. W. Pohl, Z. Physik.
7. Amplifiers based on the field effect principle had been suggested earlier in the patent literature (R. Lillienfeld and others), but apparently were not successful. Shockley's contribution was to show that it should be possible according to existing semiconductor theory to make such a device. An early successful experiment is that of W. Shockley and G. L. Pearson Phys. Rev. 74, 232 (1948).
8. J. Bardeen, Phys. Rev. 71, 717 (1947).
9. A review is given in the lecture of Dr. Brattain (Les Prix Nobel, Stockholm, 1957).
10. G. L. Pearson and J. Bardeen. Phys. Rev. 75, 865 (1948).
11. See K. Lark-Horovitz, Elec. Eng. 68, 1047 (1949).
12. W. H. Brattain and W. Shockley, Phys. Rev. 72, 345 (1947).
13. W. H. Brattain, Phys Rev. 17, 345 (1947).
14. J. Bardeen and W. H. Brattain, Phys. Rev. 70, 230 (1948); 75, 1208 (1949).
15. W. H. Brattain and J. Bardeen, Phys. Rev. 74, 231 (1948).
16. W. Shockley, Electrons and Holes in Semiconductors, Van Nostrand, New York (1950), p. 86.
17. John N. Shive, Phys. Rev. 75, 689 (1949).

VIII

The prize was awarded in 1986 to Heinrich Rohrer for his design of scanning tunneling microscope.

* * *

Scanning tunneling microscopy - from birth to adolescence

by Heinrich Rohrer and Gerd Binnig (see p. 80).

We present here the historic development of Scanning Tunneling Microscopy; the physical and technical aspects have already been covered in a few recent reviews and two conference proceedings [1] and many others are expected to follow in the near future. A technical summary is given by the sequence of figures which stands alone. Our narrative is by no means a recommendation of how research should be done, it simply reflects what we thought, how we acted and what we felt. However, it would certainly be gratifying if it encouraged a more relaxed attitude towards doing science.

Perhaps we were fortunate in having common training in superconductivity, a field which radiates beauty and elegance. For scanning tunneling microscopy, we brought along some experience in tunneling [2] and angstroms [3], but none in microscopy or surface science. This probably gave us the courage and light-heartedness to start something which should "not have worked in principle" as we were so often told.

"After having worked a couple of years in the area of phase transitions and critical phenomena, and many, many years with magnetic fields, I was ready for a change. Tunneling, in one form or another had intrigued me for quite some time. Years back, I had become interested in an idea of John Slonczewski to read magnetic bubbles with tunneling; on another occasion, I had been involved for a short time with tunneling between very small metallic grains in bistable resistors, and later I watched my colleagues struggle with tolerance problems in the fabrication of Josephson junctions. So the local study of growth and electrical properties of thin insulating layers appeared to me an interesting problem, and I was given the opportunity to hire a new research staff member, Gerd Binnig, who found it interesting, too, and accepted the offer. Incidentally, Gerd and I would have missed each other, had it not been for K. Alex Müller, then head of Physics, who made the first contacts [1] ."

The original idea then was not to build a microscope but rather to perform spectroscopy locally on an area less than 100 Å in diameter. "On a house-hunting expedition, three months before my actual start at IBM, Heini Rohrer discussed with me in more detail his thoughts on inhomogeneities on surfaces, especially those of thin oxide layers grown on metal surfaces. Our discussion revolved around the idea of how to study these films locally, but we realized that an appropriate tool was lacking. We were also puzzling over whether arranging tunneling contacts in a specific manner would give more insight on the subject. As a result of that discussion, and quite out of the blue at the LT15 Conference in Grenoble - still some weeks before I actually started at IBM - an old dream of mine stirred at the back of my mind, namely, that of vacuum tunneling. I did not learn until several years later that I had shared this dream with many other scientists, who like myself, were working on tunneling spectroscopy. Strangely enough none of us had ever talked about it, although the idea was old in principle." Actually it was 20 years old, dating back to the very beginning of tunneling spectroscopy [4]. Apparently, it had mostly remained an idea and only shortly after we had started, did Seymour Keller, then a member of the IBM Research Division's Technical Review Board and an early advocate of tunneling as a new research area in our Laboratory, draw our attention to W.A. Thompson's attempting vacuum tunneling with a positionable tip [5] .

We became very excited about this experimental challenge and the opening up of new possibilities. Astonishingly, it took us a couple of weeks to realize that not only would we have a local spectroscopic probe, but that scanning would deliver spectroscopic and even topographic images, i.e., a new type of microscope. The operating mode mostly resembled that of stylus profilometry [6], but instead of scanning a tip in mechanical contact over a surface, a small gap of a few angstroms between tip and sample is maintained and controlled by the tunnel current flowing between them. Roughly two years later and shortly before getting our first images, we learned about a paper by R. Young *et al.* [7] where they described a type of field-emission microscope they called "topografiner". It had much in common with our basic principle of operating the STM, except that the tip had to be rather far away from the surface, thus on high voltage producing a field-emission current rather than a tunneling current and resulting in a lateral resolution roughly that of an optical microscope. They suggested to improve the resolution by using sharper field-emission tips, even attempted vacuum tunneling, and discussed some of its exciting prospects in spectroscopy. Had they, even if only in their minds, combined vacuum tunneling with scanning, and estimated that resolution they would probably have ended up with the new concept, Scanning Tunneling Microscopy. They came closer than anyone else.

Mid-January 1979, we submitted our first patent disclosure on STM. Eric Courtens, then deputy manager of physics at the IBM Rüschlikon Laboratory, pushed the disclosure to a patent application with "thousands of future STM's". He was the first believer in our cause. Shortly afterwards, following an in-house seminar on our STM ideas, Hans-Jörg Scheel became the third. For the technical realization of our project, we were fortunate in securing the craftsmanship of Christoph Gerber. "Since his joining IBM in 1966, Christoph had worked with me (HR) on pulsed high-magnetic fields, on phase diagrams, and on critical phenomena. By the end of 1978, we were quite excited about our first experimental results on the random-field problem, but when asked to participate in the new venture, Christoph did not hesitate an instant. He always liked things which were out of the ordinary, and, incidentally, was the second believer. This left me and the random-field problem without his diligent technical support. About a year later, Edi Weibel was the next one to join in, which left another project without technical support. Finally, I completed the team, leaving the random-field problem to others."

During the first few months of our work on the STM, we concentrated on the main instrumental problems and their solutions [8]. How to avoid mechanical vibrations that move tip and sample against each other? Protection against vibrations and acoustical noise by soft suspension of the microscope within a vacuum chamber. How strong are the forces between tip and sample? This seemed to be no problem in most cases. How to move a tip on such a fine scale? With piezoelectric material, the link between electronics and mechanics, avoiding friction. The continuous deformation of piezomaterial in the angstrom and subangstrom range was established only later by the tunneling experiments themselves. How to move the sample on a fine scale over long distances from the position of surface treatment to within reach of the tip? The 'louse'. How to avoid strong thermally excited length fluctuations of the sample and especially the tip? Avoid whiskers with small spring constants. This led to a more general question, and the most important one: What should be the shape of the tip and how to achieve it? At the very beginning, we viewed the tip as a kind of continuous matter with some radius of curvature. However, we very soon realized that a tip is never smooth because of the finite size of atoms, and because tips are quite rough unless treated in a special way. This roughness implies the existence of minitips as we called them, and the extreme sensitivity of the tunnel current on tip-sample separation then selects the minitip reaching closest to the sample.

Immediately after having obtained the first stable STM images showing remarkably sharp monoatomic steps, we focused our attention onto atomic resolution. Our hopes of achieving this goal were raised by the fact that vacuum tunneling itself provides a new tool for fabricating extremely sharp tips: The very local, high fields obtainable with vacuum tunneling at a few volts only can be used to shape the tip by field migration or by field evaporation. Gently touching the surface is another possibility. All this is not such a controlled procedure as tip sharpening in field-ion microscopy, but it appeared to us to be too complicated to combine STM with field-ion microscopy at this stage. We hardly knew what field-ion microscopy was, to say nothing of working with it. We had no means of controlling exactly the detailed shape of the tip. We repeated our trial-and-error procedures until the structures we observed became sharper and sharper. Sometimes it worked, other times it did not.

But first we had to demonstrate vacuum tunneling. In this endeavour, apart from the occurrence of whiskers, the most severe problem was building vibrations. To protect the STM unit also against acoustical noise, we installed the vibration-isolation system within the vacuum chamber. Our first set-up was designed to work at low temperatures and in ultra-high vacuum (UHV). Low temperatures guaranteed low thermal drifts and low thermal length fluctuations, but we had opted for them mainly because our thoughts were fixed on spectroscopy. And tunneling spectroscopy was a low-temperature domain for both of us with a Ph.D. education in superconductivity. The UHV would allow preparation and retention of well-defined surfaces. The instrument was beautifully designed with sample and tip accessible for surface treatments and superconducting levitation of the tunneling unit for vibration isolation. Construction and first low-temperature and UHV tests took a year, but the instrument was so complicated, we never used it. We had been too ambitious, and it was only seven years later that the principal problems of a low-temperature and UHV instrument were solved [9]. Instead, we used an exsicator as vacuum chamber, lots of Scotch tape, and a primitive version of superconducting levitation wasting about 20 l of liquid helium per hour. Emil Haupt, our expert glassblower, helped with lots of glassware and, in his enthusiasm, even made the lead bowl for the levitation. Measuring at night and hardly daring to breathe from excitement, but mainly to avoid vibrations, we obtained our first clear-cut exponential dependence of the tunnel current I on tip-sample separation s characteristic for tunneling. It was the portentous night of March 16, 1981.

So, 27 months after its conception the Scanning Tunneling Microscope was born. During this development period, we created and were granted the necessary elbow-room to dream, to explore, and to make and correct mistakes. We did not require extra manpower or funding, and our side activities produced acceptable and publishable results. The first document on STM was the March/April 1981 in-house Activity Report.

A logarithmic dependence of the tunnel current I on tip-sample separation s alone was not yet proof of vacuum tunneling. The slope of ln I versus s should correspond to a tunnel-barrier height of $P \approx 5$ eV, characteristic of the average workfunctions of tip and sample. We hardly arrived at 1 eV, indicating tunneling through some insulating material rather than through vacuum. Fortunately, the calibration of the piezosensitivity for small and fast voltage changes gave values only half of those quoted by the manufacturers. This yielded a tunnel-barrier height of more than 4 eV and thus established vacuum tunneling. This reduced piezosensitivity was later confirmed by careful calibration with H.R. Ott from the ETH, Zürich, and of S.Vieira of the Universidad Autónoma, Madrid [10].

U. Poppe had reported vacuum tunneling some months earlier [11], but his interest was tunneling spectroscopy on exotic superconductors. He was quite successful at that but did not measure I(s). Eighteen months later, we were informed that E.C. Teague, in his Thesis, had already observed similar I(s) curves which at that time were not commonly available in the open-literature [12].

Our excitement after that March night was quite considerable. Hirsh Cohen, then Deputy Director of our Laboratory, spontaneously asked us "What do you need?", a simple and obvious question people only rarely dare to ask. "Gerd immediately wanted to submit a post-deadline contribution [13] to the LT16 Conference to be held in Los Angeles in September. He was going there anyway with his superconducting strontium titanate, and I was sure we would have some topographic STM images by then. And indeed we had. I arranged an extended colloquium tour through the USA for Gerd, but about three weeks before his departure, a friend warned him, that once the news became public, hundreds of scientists would immediately jump onto the STM bandwagon. They did - a couple of years later. After two extended discussions on a weekend hike, he nevertheless became convinced that it was time for the STM to make its public appearance." Our first attempt to publish a letter failed. "That's a good sign", Nico Garcia, a Visiting Professor from the Universidad Autónoma de Madrid, Spain consoled us.

After this first important step with a complete STM set-up, it took us only three months, partly spent waiting for the high-voltage power supplies for the piezos, to obtain the first images of monosteps [14] on a $CaIrSn_4$ single crystal grown by R. Gambino. Here, the main problem was getting rid of the whiskers we continually created by bumping the tip into the surface. Now we were ready to turn to surface science, first to resolve surface reconstructions. We built a UHV-compatible STM (no longer with Scotch tape!) and as a quick trial, operated it in vacuum suspended from a rubber band. The results indicated that superconducting levitation might be unnecessary.

That was the state of the art for the publicity tour through the USA in September '81. Most reactions were benevolent, some enthusiastic, and two even anticipated the Nobel prize, but the STM was apparently still too exotic for any active outside engagement.

Next, we protected the STM from vibrations by a double-stage spring system with eddy-current damping [8], and incorporated it in a UHV chamber not in use at that moment. We added sputtering and annealing for sample treatment but no other surface tool to characterize and monitor the state of the sample or tip could yet be combined with that STM. Although the superconducting levitation served for three months only, it was cited for years. It would appear that something complicated is much easier to remember!

A most intriguing and challenging surface-science problem existed, namely, the 7 X 7 reconstruction of the Si(111) surface. A class of fashionable models contained rather rough features which should be resolvable by the STM. So we started to chase after the 7 X 7 structure, and succumbed to its magic. At first, with no success. The STM would function well, sometimes with resolutions clearly around 5 Å, but not our surface preparation. We occasionally found quite nice patterns with monolayer step lines [8] but usually the surface always looked rough and disordered on an atomic scale. One image even foreshadowed the 7 x 7 by a regular pattern of depressions, the precursors of the characteristic corner holes. However, a single event is too risky to make a case for a new structure obtained with a new method. But it boosted our confidence.

By spring '82, STM was already a subject talked about. Supposedly, an image of a vicinal surface expertly prepared with a regular step sequence would have eased the somewhat reserved attitude of the surface-science community. We, however, thought that the mono-, double-, and triple-steps of the $CaIrSn_4$ with atomically flat terraces [14] and the step lines of Si(111) [8] were convincing and promising enough. And instead of wasting further time on uninteresting step lines, we preferred to attack surface reconstructions with known periodicities and with a reasonable chance of learning and contributing something new.

For easier sample preparation and because the demand on resolution was only 8 Å, we changed to a gold single crystal, namely, the (110) surface known to produce a 1 X 2 reconstruction. This seemed to be well within reach of the STM resolution from what we had learned from the silicon step lines. Although some time earlier, we had returned to Karl-Heinz Rieder, the Laboratory's surface-science expert, his Si single crystal in a kind of droplet form, it did not deter him from proposing this gold experiment which meant lending us his Au crystal, and some weeks later we added another droplet to his collection! But in between, with his advice on surface preparation, we succeeded in resolving the 1 X 2 structure [15]. Contrary to expectations, we also had to struggle with resolution, because Au transferred from the surface even if we only touched it gently with our tip. The mobility of Au at room temperature is so high that rough surfaces smooth out after a while, i. e., really sharp Au-coated tips cease to exist. We should like to mention here that later, for measurements on Au(100), we formed sharp Au tips by field evaporation of Au atoms from sample to tip, and could stabilize them by a relatively high field resulting from a 0.8 V tunnel voltage.

In the case of the Au(110) surface, the atomic resolution was rather a matter of good luck and perseverance. It jumped from high to low in an unpredictable manner, which was probably caused by migrating adatoms on the tip finding a stable position at the apex for a while. We also observed an appreciable disorder leading to long but narrow ribbons of the 1 x 2 reconstruction mixed with ribbons of 1 x 3 and 1 X 4 reconstructions and step lines. Nevertheless, these experiments were the first STM images showing atomic rows with atomic resolution perpendicular to the rows. The disorder, intrinsic on this surface, but in its extent criticized from the surface-science point of view, demonstrated very nicely the power of STM as a local method, and about a year later played an important role in testing the first microscopic theories of scanning tunneling microscopy.

With gold, we also performed the first spectroscopy experiment with an STM. We wanted to test a prediction regarding the rectifying I-V characteristic of a sample-tip tunnel junction induced by the geometric asymmetry [16]. Unfortunately, the sample surface became unstable at around 5 V, sample positive, and the small asymmetry observed in this voltage range could also have been due to other reasons. But with reversed polarity, the voltage could be swept up to 20 V producing a whole series of marked resonant surface states. We consider the gold exercise during spring and early summer of '82 a most important step in the development of the method, and the STM had already exceeded our initial expectations. We had also won our first believers outside the Laboratory, Cal Quate from Stanford University [17] and Paul Hansma from the University of California at Santa Barbara [18]. We gave numerous talks on the Au work, and it attracted some attention but all in all, there was little action. We did not even take the time to write a paper - the 7 X 7 was waiting!

Meanwhile, we had also made the first attempts at chemical imaging: Small Au islands on silicon. The islands were visible as smooth, flat hills on a rough surface in the topography, but they were also clearly recognizable as regions with enhanced tunnel-barrier height [8]. Thus, the Au islands were imaged thanks to their different surface electronic properties. It would certainly have been interesting to pursue this line, but we knew that, in principle, it worked, - and the 7 X 7 was still waiting!

We started the second 7 x 7 attempt in autumn 1982 taking into consideration the advice of Franz Himpsel not to sputter the surface. This immediately worked and we observed the 7 x 7 wherever the surface was flat. We were absolutely enchanted by the beauty of the pattern.

"I could not stop looking at the images. It was like entering a new world. This appeared to me as the unsurpassable highlight of my scientific career and therefore in a way its end. Heini realized my mood and whisked me away for some days to St. Antönien, a charming village high up in the Swiss mountains, where we wrote the paper on the 7 x 7."

We returned convinced that this would attract the attention of our colleagues, even of those not involved with surface science. We helped by presenting both an unprocessed relief model assembled from the original recorder traces with scissors, Plexiglass and nails, and a processed top view; the former for credibility, the latter for analysis and discussion [19]. It certainly did help, with the result that we practically stopped doing research for a while. We were inundated with requests for talks, and innumerable visitors to our Laboratory were curious to know how to build an STM. However, the number of groups that seriously got started remained small. It seemed there was still a conflict between the very appealing, conceptual easiness of displaying individual atoms in three dimensional real space direct by recorder traces, and the intuitive reservation that, after all, it just could not be that simple.

Our result excluded all the numerous models that existed, and strangely enough also some that followed. Only one came very close: The adatom model by W. Harrison [20] with just the number of adatoms not quite right. Nowadays, a variation of the adatom model where deeper layers are also reconstructed besides the characteristic 7 x 7 adatom pattern [21], is generally accepted and compatible with most results obtained by various experimental methods like ion channeling [22], transmission electron diffraction [23], and more detailed STM results from other groups [24].

The 7 X 7 experiments also accelerated the first theoretical efforts of STM on a microscopic level. Tersoff and Hamman, and Baratoff [25] applied Bardeen's transfer Hamiltonian formalism to the small geometries of tip and an atomically corrugated surface. Garcia, Ocal, and Flores, and Stoll, Baratoff, Selloni, and Carnevali worked out a scattering approach [26]. The two approaches converged; they consoled us by roughly confirming our intuitive view on tunneling in small geometries by simply scaling down planar tunneling, and they certainly improved the acceptance of STM in physics circles. The theoretical treatments concentrated on the nonplanar aspect of tunneling of free electrons, and the STM results on Au(110), still unpublished, served as a testing ground. They remained unpublished for quite some time, since the flashy images of the 7 x 7 silicon surface somehow overshadowed the earlier Au(110) experiments. One reaction to the first attempt to publish them was: "...The paper is virtually devoid of conceptual discussion let alone conceptual novelty... I am interested in the behavior of the surface structure of gold and the other metals in the paper. Why should I be excited about the results in this paper?..." It was certainly bad publication management on our part, but we were not sufficiently familiar with a type of refereeing which searches for weak points, innocently ignoring the essence.

The gold and silicon experiments showed that STM in surface science would benefit greatly from additional, in-situ surface characterization, in particular low-energy electron diffraction (LEED). We had already learned that surfaces, even elaborately prepared, were frequently not as uniform and flat as generally assumed. The in-situ combination of LEED with STM proved extremely helpful, avoiding searching when there was nothing to be searched, and it gave us the opportunity to learn about and work with LEED and Auger electron spectroscopy (AES). The combination of STM with other established surface-science techniques also settled a concern frequently mentioned: How much did our STM images really have in common with surfaces characterized otherwise? We did not share this concern to such a degree, as we had also learned that reconstructions extended unchanged to the immediate vicinity of defect areas, and because we could detect most contaminants or defects individually. Thus, for us, the combined instrumentation was more a practical than a scientific issue.

After a short but interesting excursion with the new STM/LEED/AES combination into resolving and understanding the (100) surface of Au [27], we proceeded into the realms of chemistry. Together with A. Baró, a Visiting Professor from Universidad Autónoma de Madrid, Spain, who also wanted to familiarize himself with the technique, we observed the oxygen-induced 2 x 1 reconstruction of Ni(110) [28], interpreting the pronounced and regularly arranged protrusions we saw as individual oxygen atoms. We had seen atomic-scaled features before, which could be interpreted as adsorbates or adsorbate clusters but they were more a nuisance than a matter of interest. The oxygen on Ni experiments demonstrated that the oxygen overlayer was not irreversibly changed by the imaging tunnel tip. This was a most significant result in regard to observing, studying and performing surface chemistry with an STM tip. About a year later, when studying the oxygen-induced 2 X 2 reconstructed Ni(100) surface, we observed characteristic current spikes which we could attribute to oxygen diffusing along the surface underneath the tip [29]. We noted that the same type of spikes had already been present in our earlier images of oxygen-covered Ni(110), but had been discarded at that time. Not only could diffusing atoms be observed individually, but their migration could be correlated to specific surface features like step lines or bound oxygen atoms, imaged simultaneously. Towards the end of 1983, we also started to probe the possibilities of STM in biology together with H. Gross from the ETH, Zürich. We could follow DNA chains lying on a carbon film deposited on a Ag-coated Si wafer [30].

That year ended with a most pleasant surprise: On Friday December 9, we received a telegram from the secretary of the King Faisal Foundation, followed on Monday by a phone call from the secretary of the European Physical Society announcing the King Faisal Prize of Science and the Hewlett Packard Europhysics Prize, respectively. "The day the telegram arrived, Gerd was in Berlin delivering the Otto Klung Prize lecture. It was also my twentieth anniversary with IBM." This was an encouraging sign that Scanning Tunneling Microscopy was going to make it. It also brought a new flood of requests.

In the summer of 1984, we were finally ready to assume what we had set out to do in autumn 1978, before the notion of microscopy had ever evolved, namely, performing local spectroscopy. Together with H. Fuchs and F. Salvan, we investigated the clean 7 x 7 [1,31] and the $\sqrt{3} \times \sqrt{3}$ Au reconstructions on Si(111) [31], and - right back to the heart of the matter - a thin oxide film on Ni [1,32]. We could see that surfaces are electronically structured as known, for example, from photoemission experiments, and that we could resolve these electronic structures in space on an atomic scale. We called this (and still do) the color of the atoms. Indeed, the oxide layers were inhomogeneous and most clearly visible in scanning tunneling spectroscopy (STS) images. On the 7 X 7, we could see by STS down to the second layer, and observe individual dangling bonds between the adatoms

[1]. At that time, C. Quate and his group already had an STM running, and they had performed local spectroscopy; not yet with atomic resolution but a low temperature [33]. They had measured the energy gap of a superconductor, and later even plotted its spatial dependence. Spectroscopic imaging was not really surprising, yet it was an important development. We now had the tools to fully characterize a surface in terms of topographic and electronic structure. Although it is usually quite an involved problem to separate the property of interest from a set of STM and STS measurements, our vision of the scanning tunneling microscope had become true. But nevertheless, we heard that this view was not generally shared. Rumors reached us that scientists would bet cases of champagne that our results were mere computer simulations! The bets were probably based on the fact the STM was already three years old, and atomic resolution was still our exclusive property. This was also our concern, but in another way. In late summer '83, Herb Budd, promoter of the IBM Europe Institute and an enthusiastic STM supporter, had asked us to run an STM Seminar in summer 1984 within the framework of the Institute. This meant one week with 23 lectures in front of a selected audience of the European academia. At that time, there was no way whatsoever of filling 23 hours, let alone of committing 23 speakers. A year later, we agreed, full of optimism for summer '85. In December '84, on Cal Quate's initiative, nine representatives of the most advanced STM groups came together for a miniworkshop in a hotel room in Cancun. It was a most refreshing exchange of ideas, but there was still no other atomic resolution, and thus not a sufficient number of lectures in sight for the Seminar.

In the following few months, the situation changed drastically. R. Feenstra and coworkers came up first with cleaved GaAs [34], C.F. Quate's group with the 1 X 1 structure on Pt(100) [35], and J. Behm, W. Hoesler, and E. Ritter with the hexagonal phase on Pt(100) [36]. At the American Physical Society March Meeting in 1985, P. Hansma presented STM images of graphite structures of atomic dimensions [37], and when J. Golovchenko unveiled the beautiful results on the various reconstructions of Ge films deposited on Si(111) [38], one could have heard a pin drop in the audience. The atomic resolution was official and scanning tunneling microscopy accepted. The IBM Europe Institute Seminar in July turned into an exclusive workshop for STM'ers, and comprised some 35 original contributions, not all of them on atomic resolution, but already more than in March [39]. "A watershed of ideas" as Cal Quate expressed it.

Our story so far has dealt mainly with the striving for structural and electronic imaging in a surface-science environment with atomic resolution. Individual atoms had been seen before with field-ion microscopy, and dealt with individually by the atom probe technique [40].

The beauty of these techniques is relativized by the restriction to distinct atom sites on fine tips made from a rather limited selection of materials. Similarly, electron microscopy, the main source of present-day knowledge on submicron structures in practically all areas of science, technology, and industry, has advanced to the atomic level. Imaging of individual atoms or atomic structures, however, is still reserved for specific problems, expertise, and extraordinary equipment. The appeal and the impact of STM lie not only in the observation of surfaces atom by atom, but also in its widespread applicability, its conceptual and instrumental simplicity and its affordability, all of which have resulted in a relaxed and almost casual perception of atoms and atomic structures.

But there are many other aspects, maybe less spectacular but nonetheless significant, which have made STM an accepted and viable method now pursued in many areas of science and technology. The instruments themselves have become simpler and smaller. Their greatly reduced size allows easy incorporation into other systems, for instance, into a scanning electron microscope [41]. One type of instrument retains accurate sample positioning but is sufficiently rigid for in-situ sample and tip exchange. Other instruments are so rigid they are even insensitive to vibrations when immersed in liquid nitrogen [42] and even small enough to fit through the neck of a liquid-helium storage vessel [43]. These humming-birds of STM, some concepts of which reach back to the squeezable tunnel junctions [18], can also operate at television speed on relatively flat surfaces using single-tube scanners [43,44]. Also tip preparation has advanced to a level where well-defined pyramidal tips ending with one [45] or more [46] atoms can be fabricated in a UHV environment. Such tips are particularly important for investigations of nonperiodic structures, disordered systems and rough surfaces. They are also interesting in their own right, for example, as low-energy electron and ion point sources.

Outside the physics and surface-science communities, the various imaging environments and imaging capabilities seem as appealing as atomic resolution. Images obtained at ambient-air pressure were first reported in 1984 [47], followed by imaging in cryogenic liquids [42], under distilled water [48], in saline solutions [48], and in electrolytes [49]. Scanning tunneling potentiometry appears to have become an interesting technique to study the potential distribution on an atomic scale of current-carrying microstructures [50]. More recent advances include interatomic-force imaging with the atomic-force microscope [51], with which the structure and elastic properties of conductors and insulators are obtained, and combined imaging of electronic and elastic properties of soft materials [52]. Also the use of spin polarized electron tunneling to resolve magnetic surface structures is being explored.

Finally, we revert to the point where the STM originated: The performance of a local experiment, at a preselected position and on a very small spatial scale down to atomic dimensions. Besides imaging, it opens, quite generally, new possibilities for experimenting, whether to study nondestructively or to modify locally: Local high electric fields, extreme current densities, local deformations, measurements of small forces down to those between individual atoms [53], just to name a few, ultimately to handle atoms and to modify individual molecules, in short, to use the STM as a Feynman Machine [54]. This area has not yet reached adolescence.

The STM's "Years of Apprenticeship" have come to an end, the fundamentals have been laid, and the "Years of Travel" begin. We should not like to speculate where it will finally lead, but we sincerely trust that the beauty of atomic structures might be an inducement to apply the technique to those problems where it will be of greatest service solely to the benefit of mankind. Alfred Nobel's hope, our hope, everybody's hope.

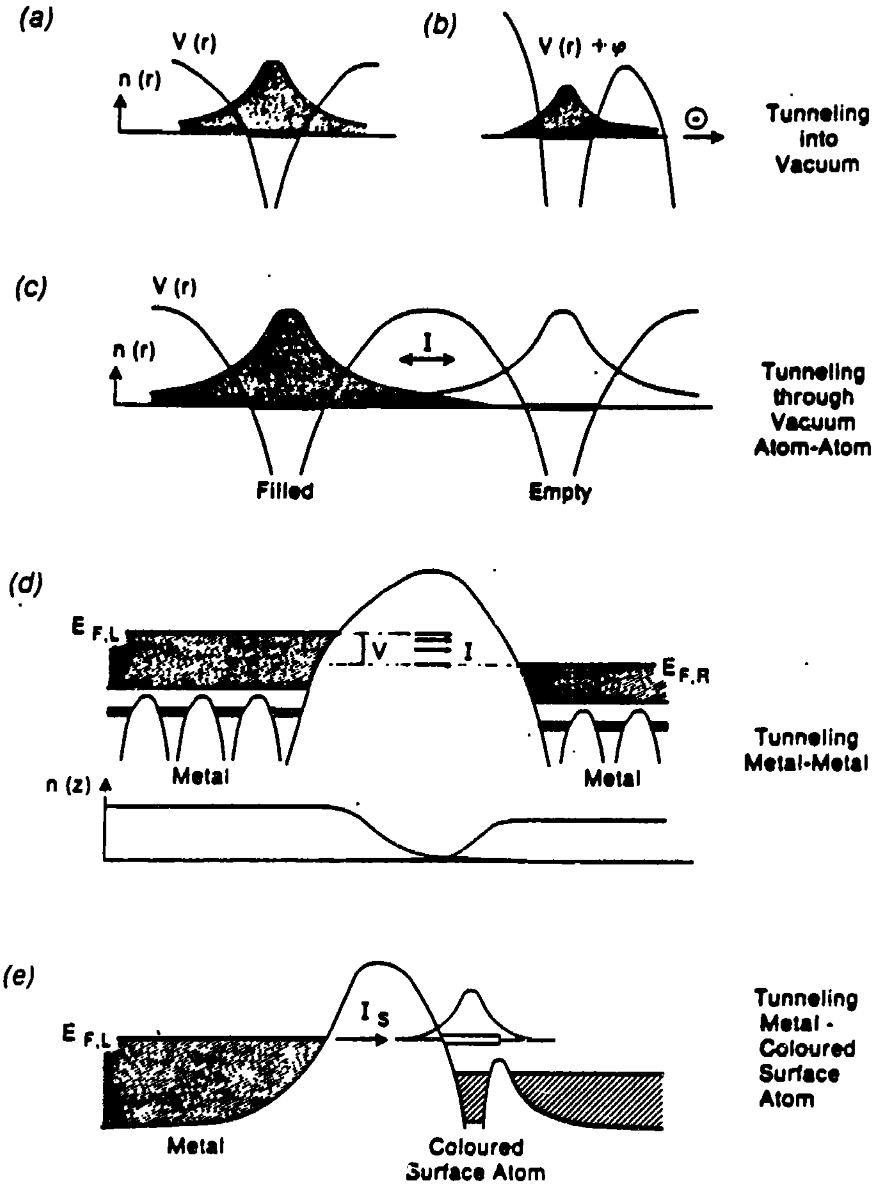

Fig. 1. Tunneling. (a) The wave function of a valence electron in the Coulomb potential well of the atom core plus other valence electrons extends into the vacuum; it "tunnels" into the vacuum. (b) Exposed to an electric field, φ, the electron can tunnel through the potential barrier and leaves the atom. (c) If two atoms come sufficiently close, then an electron can tunnel back and forth through the vacuum or potential barrier between them. (d) In a metal, the potential barriers between the atoms in the interior are quenched and electrons move freely in energy bands, the conduction bands. At the surface, however, the potential rises on the vacuum side forming the tunnel barrier through which an electron can tunnel to the surface atom of another metal close by. The voltage V applied between the two metals produces a difference between the Fermi levels $E_{F,L}$ and $E_{F,R}$, thus providing empty states on the right for the electrons tunneling from the left side. The resulting tunnel current is roughly of the form $I = f(V)\exp(-\sqrt{\varphi}s)$. The f(V) contains a weighted joint local density of states of tip an object, the exponential gives the transmittivity with φ the averaged tunnel barrier height in eV, and s the separation of the two metals in Å. Here f(V) and $\sqrt{\varphi}$ are material properties obtained by measuring dlnI/dV and dlnI/ds. (e) A simple case of local spectroscopy. A characteristic state, the "color", of a surface species is observed by the onset of the tunnel-current contribution IΣ, [see Lang, N. D. (1987). Phys. Rev. Lett. 58, 45, and references therein].

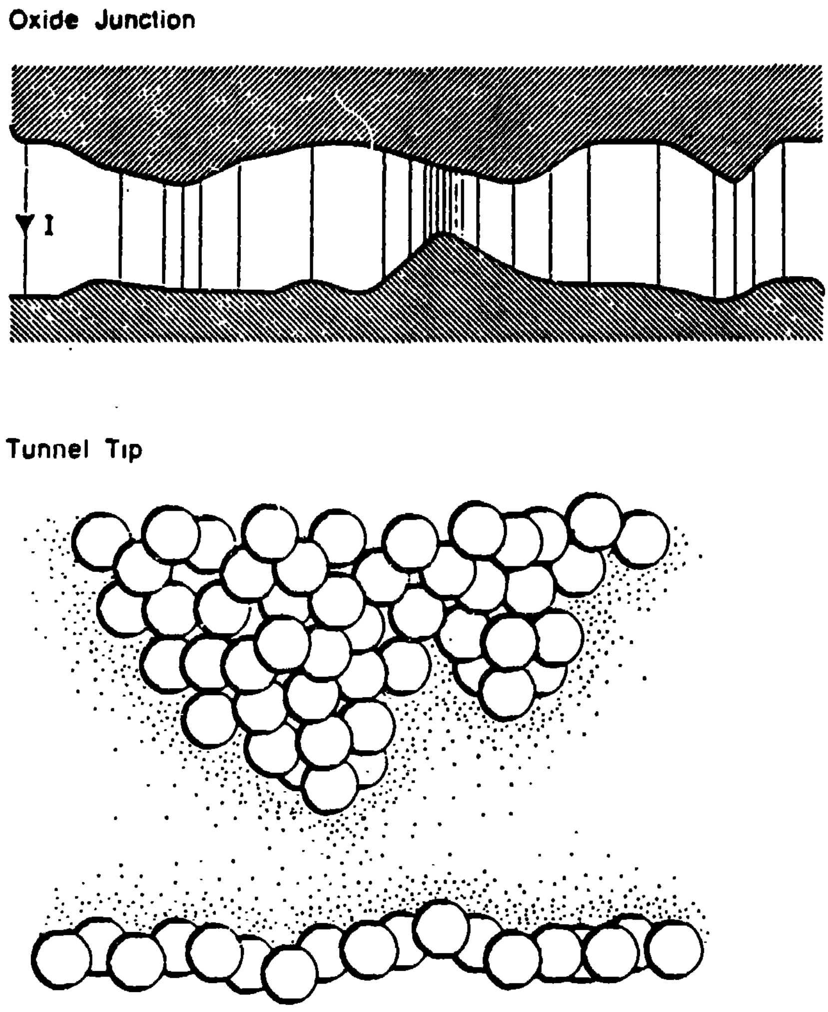

Fig. 2. The principle. The tunneling transmittivity decreases exponentially with the tunneling distance, in vacuum about a factor 10 for every Å. In an oxide tunnel junction, most of the current flows through narrow channels of small electrode separation. With one electrode shaped into a tip, the current flows practically only from the front atoms of the tip, in the best case from a specific orbital of the apex atom. This gives a tunnel-current filament width and thus a lateral resolution of atomic dimensions. The second tip shown is recessed by about two atoms and carries about a million times less current.

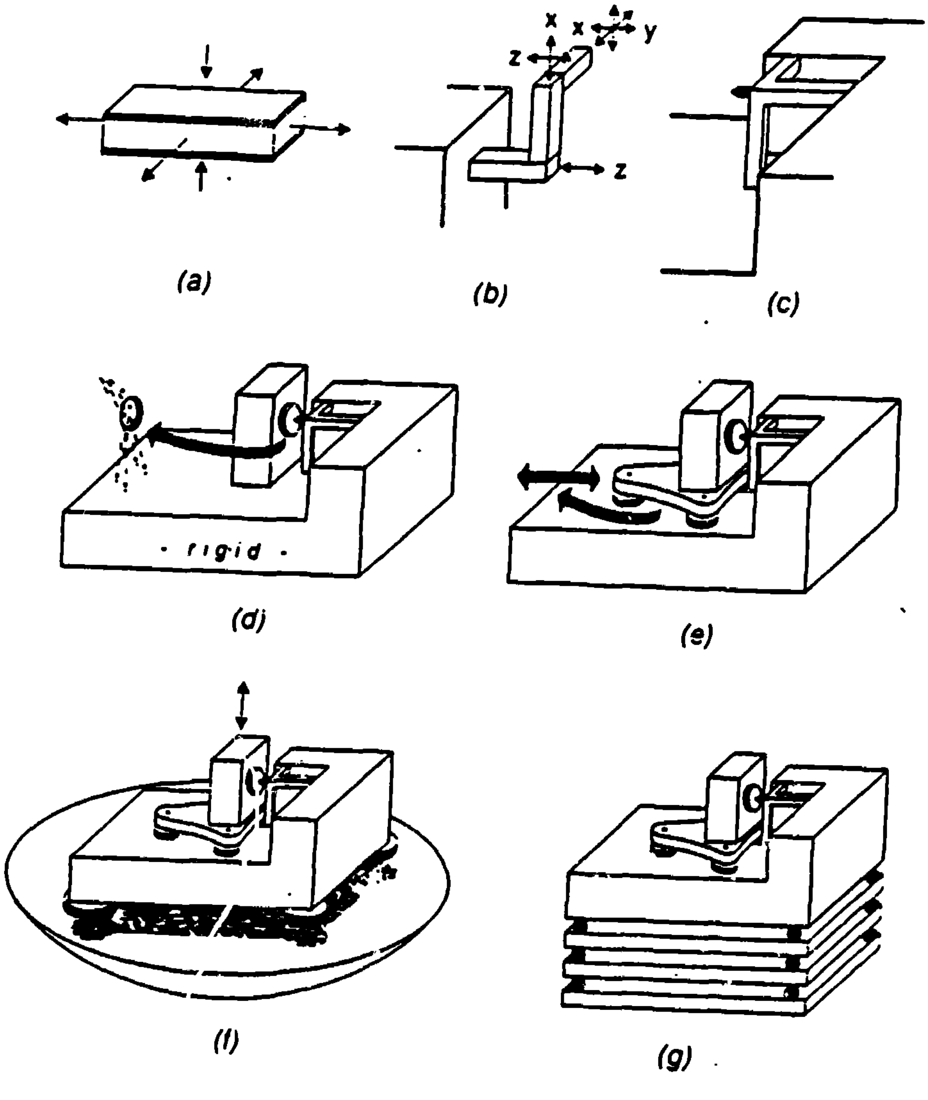

Fig. 3. The instrument. (a) A voltage applied to two electrodes contracts or expands the piezoelectric material in between. The practical total excursion of a piezo is usually in the region of micrometers. (b) A frictionless x-y-z piezodrive, which is quite vibration sensitive. (c) A rigid tripod is at present the piezodrive most used apart from the single-tube scanner. (d) Tripod and sample holder are installed on a rigid frame. The sample has to be cleared from the tip for preparation and sample transfer. (e) Positioning of the sample to within reach of the piezodrive was originally achieved with a piezoelectric 'louse' with electrostatically clampable feet. Magnetic-driven positioners and differential screws are also now in use. (f) In the first vibration-isolation system, the tunnel unit with permanent magnets levitated on a superconducting lead bowl. (g) The simple and presently widely used vibration protection with a stack of metal plates separated by viton - a UHV-compatible rubber spacer.

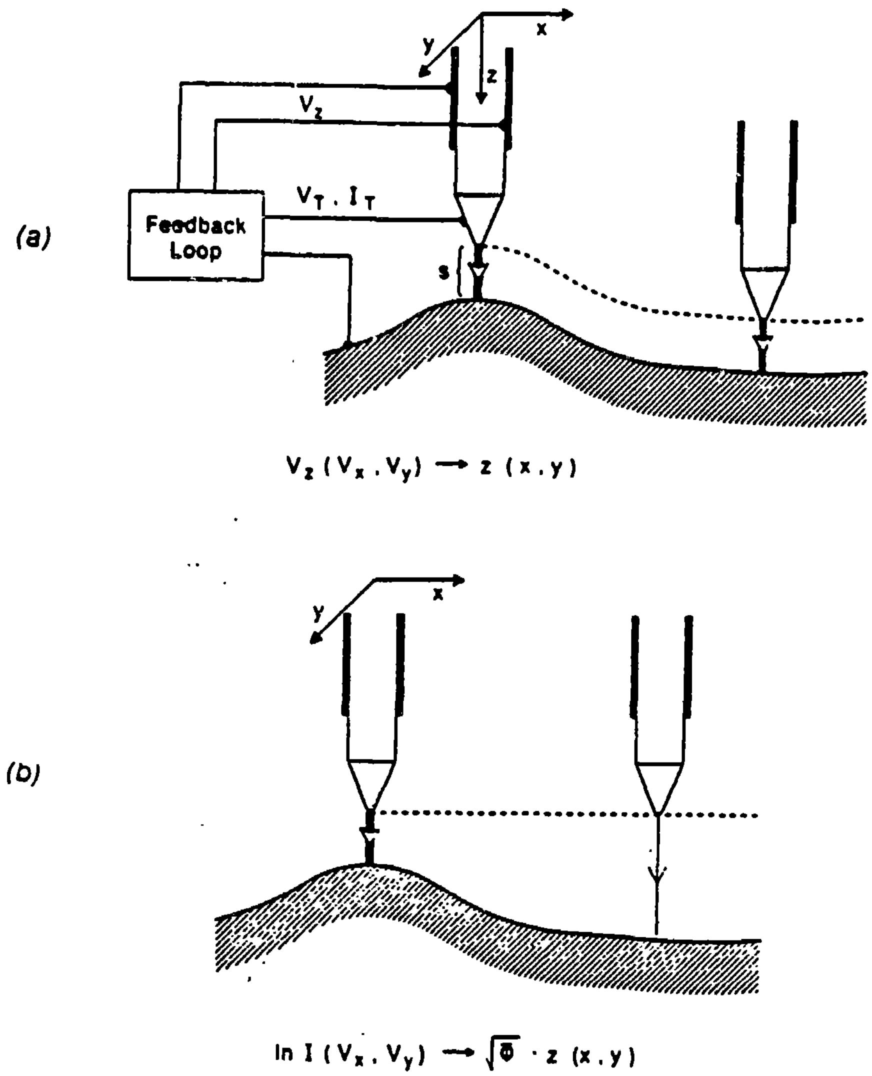

Fig. 4. Tips. (a) Long and narrow tips, or whiskers, are vibration sensitive and thermally excited. (b) A mechanically ground or etched tip shows sharp minitips, only one of which usually carries the tunnel current. Further sharpening was initially achieved with gentle contact (1), later with field evaporation (2). (c) Electrostatic and interatomic forces between tip and sample do not deform a blunt tip, or a rigid sample, but they make the tunnel gap mechanically unstable when the tip carries a whisker. The response of soft materials like graphite or organic matter to such forces, however, can be appreciable and has to be taken into account.

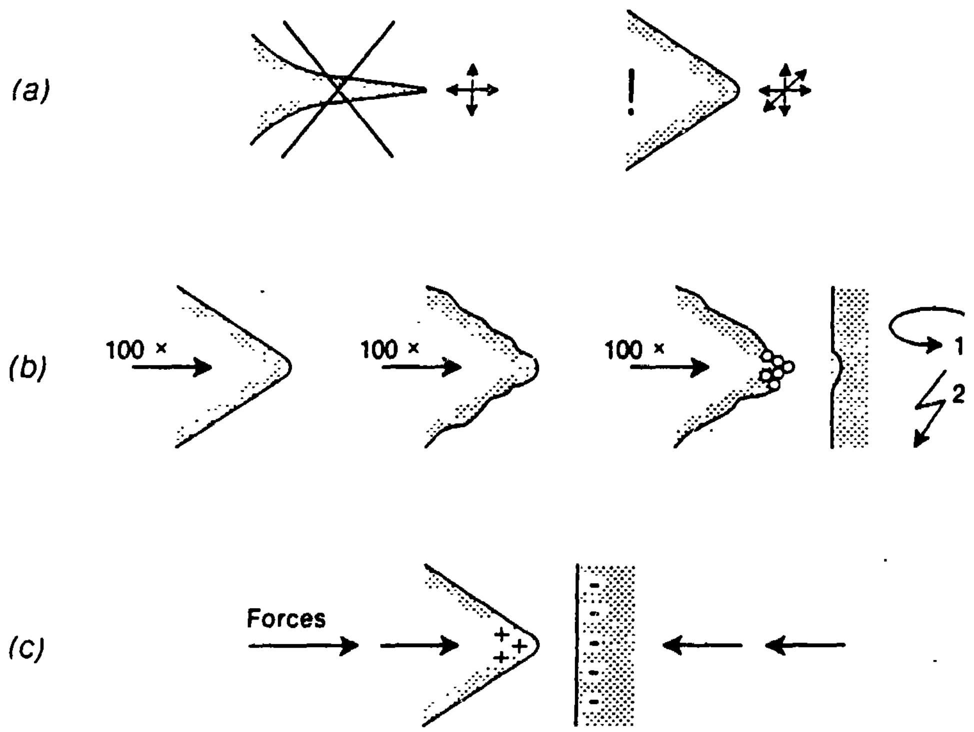

Fig. 5. Imaging. (a) In the constant current mode, the tip is scanned across the surface at constant tunnel current, maintained at a preset value by continously adjusting the vertical tip position with the feedback voltage V_Z. In the case of an electronically homogeneous surface, constant current essentially means constant s. (b) On surface portions with denivellations less than a few Å - corresponding to the dynamic range of the current measurement - the tip can be rapidly scanned at constant average z-position. Such "current images" allow much faster scanning than in (a) but require a separate determination of $\sqrt{\varphi}$ to calibrate z. In both cases, the tunnel voltage and/or the z-position can be modulated to obtain in addition, dlnI/dV and/or dlnI/ds, dlnI/ds, respectively.

REFERENCES

[1] For reviews, see Binnig, G., and Rohrer, H. (1986) IBM J. Res. Develop. 30, 355; Golovchenko, J. A. (1986) Science 232, 48; Behm, R. J., and Hoesler, W. (1986) Physics and Chemistry of Solid Surfaces, Vol VI, (Springer Verlag, Berlin), p. 361; Hansma, P. K., and Tersoff, J. (1987) J. Appl. Phys. 61, RI; Proceedings of the STM Workshops in Oberlech, Austria, July 1-5, 1985, IBM J. Res. Develop. 30, Nos. 4 and 5 (1986); Proceedings of STM'86, Santiago de Compostela, Spain, July 14-18, 1986, Surface Sci., 181, Nos. 1 and 2 (1987); An article combining technical and biographical details is presented in Dordick, Rowan L. (1986) IBM Research Magazine 24, 2.

[2] Binnig, G. K., and Hoenig, H. E. (1978) Z. Phys. B 32, 23.

[3] Rohrer, H. (1960) Helv. Phys. Acta 33, 675.

[4] Giaever, I. (1974) Rev. Mod. Phys. 46, 245.

[5] Thompson W. A., and Hanrahan, S. F. (1976) Rev. Sci. Instrum. 47, 1303.

[6] Williamson, B. P. (1967) Proc. Inst. Mech. Eng. 182, 21; Guenther, K. H., Wierer, P. G., and Bennett, J. M. (1984) Appl. Optics 23, 3820.

[7] Young, R., Ward, J., and Scire, F. (1972) Rev. Sci. Instrum. 43, 999.

[8] For technical details, see Binnig, G., and Rohrer, H. (1982) Helv. Phys. Acta 55, 726; idem (1983) Surface Sci. 126, 236; idem (1985) Sci. Amer. 253, 50.

[9] Marti, O. (1986) Ph.D. Thesis No. 8095, ETH Zurich, Switzerland; Marti, O., Binnig, G., Rohrer, H., Salemink, H., (1987) Surface Sci., 181, 230.

[10] Ott, H. R., and Rohrer, H. (1981) unpublished; Vierira, S. (1986) IBM J. Res. Develop. 30, 553.

[11] Poppe, U. (1981) Verhandl. DPG (VI) 16, 476.

[12] Teague, E. C. (1978) Dissertation, North Texas State Univ., Univ. Microfilms International, Ann Arbor, Mich., p. 141; idem (1978) Bull. Amer. Phys. Soc. 23, 290; idem (1986) J. Res. Natl. Bur. Stand. 91, 171.

[13] Binnig, G., Rohrer, H., Gerber, Ch., and Weibel, E. (1982) Physica B 109 & 110, 2075.

[14] Binnig, G., Rohrer, H., Gerber, Ch., and Weibel, E. (1982) Phys. Rev. Lett. 49, 57.

[15] Binnig, G., Rohrer, H., Gerber, Ch., and Weibel, E. (1983) Surface Sci. 131, L379.

[16] Miskowsky, N. M., Cutler, P. H., Feuchtwang, T. E., Shepherd, S.J., Lucas, A. A., and Sullivan, T. E. (1980) Appl. Phys. Lett. 37, 189.

[17] Quate, C. F. (1986) Physics Today 39, 26.

[18] Moreland, J., Alexander, S., Cox, M., Sonnenfeld, R., and Hansma, P. K. (1983) Appl. Phys. Lett. 43, 387. Actually, Paul Hansma was indisposed and could not attend the first seminar given on STM in the USA. However, his students attended, and with them Paul built the squeezable tunnel-junction.

[19] Binnig, G., Rohrer, H., Gerber, Ch., and Weibel, E. (1983) Phys. Rev. Lett. 50, 120.

[20] Harrison, W. A. (1976) Surface Sci. 55, 1.

[21] Takayanagi, K., Tanishiro, Y., Takahashi, M., and Takahashi, S. (1985) J. Vac. Sci. Tech. A 3, 1502.

[22] Tromp, R. M., and van Loenen, E. J. (1985) Surface Sci. 155, 441.

[23] Tromp, R. M. (1985) Surface Sci. 155, 432, and references therein.

[24] Becker, R. S., Golovchenko, J. A., McRae, E. G., and Swartzentruber, B. S. (1985) Phys. Rev. Lett. 55, 2028; Hamers, R. J., Tromp, R. M., and Demuth, J. E. (1986) Phys. Rev. Lett. 56, 1972.

[25] Tersoff J., and Hamann, D. R. (1983) Phys. Rev. Lett. 50, 1998; Baratoff A. (1984) Physica 127B, 143.

[26] Garcia, N., Ocal, C., and Flores, F. (1983) Phys. Rev. Lett. 50, 2002; Stoll, E., Baratoff, A., Selloni, A., and Carnevali, P. (1984) J. Phys. C 17, 3073.

[27] Binnig, G., Rohrer, H., Gerber, Ch., and Stoll, E. (1984) Surface Sci. 144, 321.[28] Baró, A. M., Binnig, G., Rohrer, H., Gerber, Ch., Stoll, E., Baratoff A., and Salvan, F. (1984) Phys. Rev. Lett. 52, 1304.

[29] Binnig, G., Fuchs, H., and Stoll, E. (1986) Surface Sci., 169, L295.

[30] Binnig, G., and Rohrer, H. (1984) Trends in Physics, editors, J. Janta and J. Pantoflicek (European Physical Society) p. 38.

[31] Baratoff, A., Binnig, G., Fuchs, H., Salvan, F., and Stoll, E. (1986) Surface Sci., 168, 734.

[32] Binnig, G., Frank, K. H., Fuchs, H., Garcia, N., Reihl, B., Rohrer, H., Salvan, F., and Williams, A. R. (1985) Phys. Rev. Lett. 55, 991; Garcia, R., Saenz, J. J., and Garcia, N. (1986) Phys. Rev. B 33, 4439.

[33] de Lozanne, A. L., Elrod, S. A., and Quate, C. F. (1985) Phys. Rev. Lett. 54, 2433.

[34] Feenstra, R. M., and Fein, A. P. (1985) Phys. Rev. B 32, 1394.

[35] Elrod, S. A., Bryant, A., de Lozanne, A. L., Park, S., Smith, D., and Quate, C. F. (1986) IBM J. Res. Develop. 30, 387.

[36] Behm, R. J., Hoesler, W., Ritter, E., and Binnig, G. (1986) Phys. Rev. Lett. 56, 228.

[37] Hansma, P. K. (1985) Bull. Amer. Phys. Soc. 30, 251.

[38] Becker, R. S., Golovchenko, J. A., and Swartzentruber, B. S. (1985) Phys. Rev. Lett. 54, 2678.

[39] Proceedings published (1986) IBM J. Res. Develop. 30, 4/5.

[40] For a review, see Ernst, N., and Ehrlich, G. (1986) Topics in Current Physics, Vol. 40, editor U. Gonser (Springer Verlag, Berlin) p. 75.

[41] Gerber, Ch., Binnig, G., Fuchs, H., Marti, O., and Rohrer, H. (1986) Rev. Sci. Instrum. 57,221.

[42] Coleman, R. V., Drake, B., Hansma, P. K., and Slough, G. (1985) Phys. Rev. Lett. 55, 394.

[43] Smith, D. P. E., and Binnig, G. (1986) Rev. Sci. Instrum. 57, 2630.

[44] Bryant, A., Smith, D. P. E., and Quate, C. F. (1986) Appl. Phys. Lett. 48, 832.

[45] Fink, H.-W., (1986) IBM J. Res. Develop. 30, 460.

[46] Kuk, Y., and Silverman, P. J. (1986) Appl. Phys. Lett. 48, 1597.

[47] Baró, A. M., Miranda, R., Alaman, J., Garcia, N., Binnig, G., Rohrer, H., Gerber, Ch. and Carrascosa, J. L. (1985) Nature 315, 253.

[48] Sonnenfeld, R., and Hansma, P. K. (1986) Science 232, 211.

[49] Sonnenfeld, R., and Schardt, B. C. (1986) Appl. Phys. Lett. 49, 1172.

[50] Muralt, P., and Pohl, D. W. (1986) Appl. Phys. Lett. 48, 514.

[51] Binnig, G., Quate, C. F., and Gerber, Ch. (1986) Phys. Rev. Lett. 56, 930.

[52] Soler, J. M., Baró, A. M., Garcia, N., and Rohrer, H. (1986) Phys. Rev. Lett. 57, 444; Dürig, U., Gimzewski, J. K., and Pohl, D. W. (1986) Phys. Rev. Lett. 57, 2403.

[53] Becker, R. S., Golovchenko, J. A., and Swartzentruber, B. S. (1987) Nature 325, 419.

[54] Feynman, R. P. (1960) Engr. and Sci., 22, February; Hameroff. S., Schneiker, C., Scott, A., Jablonka, P., Hensen, T., Sarid, D., and Bell, S. (1987).

IX

The prize was awarded in 1986 to Gerd Binnig for his design of scanning tunneling microscope.

* * *

Creativity

by Gerd Binnig.

(ENGLISH TRANSLATION TAKEN FROM THE BOOK "*AUS DEM NICHTS*").

"Creativity" is an all-embracing subject

Those were the thoughts that went through my mind to start with. But, as I looked more closely into this subject, I began to understand what I had become involved in: For, the longer I thought about it, the more general and all-embracing the subject appeared. And, at some point, I had the feeling that there couldn't be a more general topic, since it touches on practically all the fundamental issues of this world: the creation of the universe, the evolution of life and the universe, and the meaning of life. Why all of this takes place and where it is leading. Because this all has something to do with creativity, with the creation of something new.

If one looks up the traditional definitions of creativity in the dictionaries, one will see that they refer exclusively to Man -- to thought, imagination and intellect, to brainwaves and ideas. There is no mention of nature or the universe. Creativity, thus, is always connected with Man. This approach appears too restricted to me. Then, is Man really the only being or "institu- tion" capable of being creative: It was a creative act indeed by Man to invent pincers. But they were invented a long time before by Mother Nature -e.g. the claws of a crab. Or take the hypodermic needle. This was also invented by Nature a long, long time ago in the form of the poisonous snake's tooth. The similarities of both inventions, of Man's with Nature's, are evident, and one asks oneself: "Is Man's development more creative than Nature's?" Sometimes the results are so similar that one cannot talk of coincidence. It is conceivable that Man has let himself be inspired by Nature. In the scientific field of bionics there are attempts to learn from nature and to transfer "natural technology" to "human technology". Man has certainly learnt a lot from nature so far, and he will continue doing so. In addition, he will use the basis of the rest of nature to follow his own creative path. I believe that he will not come to results in many cases similar to the rest of nature, but also that the paths leading to them will reveal more similarities than differences.

Man -- "only" part of nature?

I don't see Man the way he is often seen these days: as being apart from nature. If one wants to do nature a good turn, so to speak, then one regards it as Man's partner. In my opinion, this is wrong. We are not partners, we are part of nature. We are a subsumable concept. Because Man is not principally different from the rest of nature. The fact that we separate Man from nature and place him above it probably has one of its roots in Christianity which regards Man as something very special: Only Man has an immortal soul and only he is in God's image. I regard this idea that Man is something very different from the rest of the universe as a delusion because Man is only part of nature.

Now you may think: "It is a trivial thing to say that Man is only part of nature." But I would like to show, by way of an example, that, instinctively, we cannot agree with such an opinion. Let us look at a construction, e.g. an ants' hill. From the ant's perspective, it is a giant, tall building. We don't have any problems in saying that it is a natural construction. But when we look, for example, at the Empire State Building, then we will have problems calling this a natural construction. But it is a natural construction, a natural product, because it comes from us; and because we are part of nature, our products are therefore also natural products.

Thus, the term "artificial" cannot be seen in opposition to "natural" but, rather, as a subsumable concept. Logically thinking, one can see this straightaway, but our instincts get in the opposite direction. This obviously has something to do with our upbringing. I also think, then, that some of our environmental problems would not be so serious if we subscribed to the feeling that we are part of the whole rather than being <u>above</u> it.

On a definition of creativity

I would like to proceed to a definition of creativity; to a general definition which will allow me to describe the evolution of the universe. So far, I have ignored the uniqueness of Man in my description of the conception of the world. Man with his will and consciousness is, of course, unique. But we should also ask ourselves whether animals or even stones and atoms might also have a kind of will and consciousness which, in its way, is also unique. It would even be conceivable for the future to construct computers that had a will which was comparable with Man's. Even when we regard such ideas as a devaluation because Man is then put on the same level as the computer, i.e. dead matter. Humanity had already had problems before when science concluded that the earth was not the centre of the universe. This knowledge was also regarded a devaluation. But the root of the problem is probably that man overvalues himself and underestimates the so-called inanimate nature -- it is not as inanimate as we think.

We have to get used to the idea that dead matter isn't something inferior. All the wonders of this world are contained in a stone, for example, because all the laws of nature (and thus all the possibilities which can result from these) are reflected in it. Moreover, the stone isn't the simple construction one sometimes assumes. It has a kind of life. The stone has a need to communicate with its surroundings. It interacts with its surroundings in a similar way to us. It signals its temperature to its surroundings by emitting light. When it is hotter, it transmits a completely different colour of the spectrum (e.g. bright red) and with greater intensity than when it is cold and emitting infra-red light. Moreover, it indicates its mass to its surroundings by attracting other objects. If it is magnetic, it attracts magnetic particles. If it is electrostatic, then it attracts dust and becomes covered with it. Thus, it interacts quite a lot with its surroundings. We sometimes underestimate this.

What is more, human will is difficult to prove if one puts it on the same level as the ability to make a free decision. It was Schopenhauer who told a nice story about this: Someone tries in vain to prove to his friend that he has a mind of his own. How does it function? -- By doing something for which there is no apparent reason. -- Then the reason is that he does it precisely to prove to his friend that he has a mind of his own.

To imagine that Man has no freedom of decision-making and that he is only part of a running programme is unsettling. Let us assume that such an idea is near to the truth. Then something like creativity cannot exist at all. At least, not in the way we understand it according to our instincts. However, this instinctive understanding of creativity is linked with an over-evaluation of Man and his oneness with God. We perhaps delude ourselves into thinking that we can be creative like a Creator, that we can produce completely new things. But all the existing possibilities are already anchored in the laws of nature and we cannot change these. At least, we have not yet thought of a way to do this. Things happen that have to happen, and this statistical moment, which we cannot influence either, also plays a part. I would conclude this observation by quoting from Werner Heisenberg: "Man can do what he wants, but he cannot want what he wants".

We should rethink our point of view and attitude to a programme's creativity. Strictly speaking, one really only needs programmes or computers because they are somehow able to create something new. If they are really able to create something new, then they are, in a certain sense, also creative. As the computer is "fed" by its programmer, we could also say: "The computer is creative through its programmer." Or referring to us: "Man is creative through the laws of nature." Or if one wants to formulate it from the religious perspective: "Man is creative through a Creator" The computer is our tool (hopefully this will not be reversed), and we are the tool of God or the tool of the laws of nature.

X

In 1978 the prize was awarded to Arno A. Penzias for his discovery of cosmic microwave background radiation.

* * *

Ideas

by Arno Penzias.

"Did you ask any good questions today, Isaac?"
- Jennie Teig Rabi

ISAAC ISADOR RABI, one of the twentieth century's most distinguished physicists[1], invented a sensitive technique for probing the structure of atoms and molecules and thereby opened a fruitful new field of science in the 1930s. That achievement, and many others over half a century, earned Rabi every significant honor a physicist might hope to win. When an interviewer asked Rabi to speculate on the reasons for his success, he replied with the above quote from his childhood, his mother's habitual greeting when he returned home from school each day.

I got my postgraduate education at Columbia, where Rabi was a member of the faculty. In those years, I hardly knew him, even though he taught one of my courses. Our personal

[1] Sadly, I I. Rabi died in January 1988, a few months before his ninetieth birthday.

relationship developed some twenty years later, when we began to meet at the various social functions that bring Nobel Prize winners together in the New York area. We got on very well. He treated me like a fresh young squirt, and I enjoyed acting the part around him.

"How the other kids must have hated you in school," I ventured during one of our exchanges. "A well-scrubbed, well-prepared kid who knew all the answers and just wanted to see how well everyone else was doing." "Not at all," he countered. "There are questions which illuminate, and there are those that destroy. I was always taught to ask the first kind."

"Questions which illuminate" help nourish ideas. Ideas build knowledge. Students of any age need to nurture the kind of freely inquisitive spirit that Jennie Rabi encouraged in her son.

Have you asked any good questions today?

Asking Questions

Every small child I know appears to be born curious. Young children experiment with everything they can reach. They usually ask questions from the moment they learn to talk. Unfortunately, the flow of questions often slows down once they start school - unless they get the kind of encouragement that Isaac Rabi had. Personally, I can think of no better head start to healthy use of technology than a well-developed habit of inquiry. For adults, this habit is hard to maintain, for while truly dumb people often live in blissful unawareness of the gaps in their understanding, most intelligent people secretly suspect that their personal gaps are large - far larger than the people around them realize.

My first boss at Bell Labs - and one of the smartest people I have ever known - once confided a terrible secret: he felt overrated. As he spoke, I realized we shared exactly the same feeling. I could readily picture me saying those same things about myself. "The people around me think I'm smarter than I really am ...They don't suspect that I learn things more slowly than they do and less fundamentally... When I listen to a presentation on something new, the only thing that keeps me from appearing stupid is to not ask questions about things that others obviously already understand (I know they understand it because they're not asking any questions)... The things I know about are really easier to understand, almost simple by comparison to my colleagues' areas of expertise." Since then, I've found that such monologues are common indeed.

What happens when such a secretly insecure person explains something to another with the same secret self-image (or, worse yet, to a group of them)? Since A has labeled what he understands as "simple" by self-definition, A can't insult B's intelligence by going into excessive detail. A's original estimate is reinforced by the obvious ease with which B appears to take in the material - nodding occasionally and never asking for additional clarification. Accordingly, A speeds things up even more, with no change whatever in B's demeanor. Off B goes at the end of the meeting, resolved to learn what A was transmitting during the exchange, wondering if A suspected his lack of understanding.

In my experience, the attitudes that underlie this wasteful undercommunication between technical colleagues often carry over into interactions between technical people and the rest of society. If engineers and scientists fear a public display of their "ignorance", of course nontechnical people rarely question technical presentations enough to get the information they need.

Many technical meetings are saved by those who feel free enough to ask questions. When a question is asked it's easy to spot at least half a dozen heads going up, others who would like to know the answer but couldn't ask for themselves.

In discussing this subject during a recent college lecture, I mentioned my daughter Mindy's asking how flipping a car's rearview mirror produced a new image. I went on to recount my explanation, as well as her response. "I'm really glad I asked," she said. "I thought I was the only one in the world who didn't know how it worked."

But the story didn't end there. It took an unexpected twist when I asked one of my Bell Labs colleagues to look over an edited transcript of that lecture prior to its publication. His marked-up copy came back to me with a scribbled note: "I'm not sure your flip mirror explanation is correct."

Perplexed, I called another Bell Labs physicist, and she told me that automobile companies make rearview mirrors by simply putting the reflecting surface on a wedge-shaped piece of glass. While most of the light travels through the glass, hits the reflecting surface, and bounces back - just as in an ordinary mirror - a small fraction reflects off the front surface of the glass, thereby creating a second (weaker) mirror which is tilted with respect to the first one.

I had mistakenly imagined that the manufacturer ruled lines on the back surface to create a diffraction grating - creating the same effect that sometimes produces a second (weaker) rainbow inside a larger one.

Sometimes we don't know what we don't know until someone asks the question. My daughter's ignorance of the technology in this case was no deeper than mine.

Tools

Technology is clearly improvable. It's a safe bet that tomorrow's technology will be far more powerful than today's. But people can improve, too, and people have also much room for growth. As we review the needs and opportunities for advanced technology, we should also look for growth in human capabilities. Some of that growth can come from better use of the tools technology provides.

Will the next generation be better equipped to handle technology than the present one? Like most communities, the town I live in has added "computing" to its high school curriculum. As a result, many of our high school students have learned to write elementary computer programs. Dealing with a keyboard teaches a useful basic skill, just as getting behind the wheel of a car teaches safe driving. Both skills help users "handle" technology in similar ways. They teach people to use existing tools.

Some young people learn to make imaginative use of computing tools long before they enter high school. I first met Dahlia Schwartz when she was still in second grade. Her grandmother - Lillian Schwartz, a pioneer in the computer art field, and a longtime Bell Labs consultant in computer graphics - brought her to my office for a visit. At the time, Dahlia was already quite comfortable with computers. She routinely used the wordprocessing program on her father's home computer for her letters, various games, and school assignments. Moreover, on trips to her grandparents' home, she frequently got an opportunity to explore computer graphics on some of her grandmother's more powerful machines.

During Dahlia's visit to my office, much of our conversation centered around the special features of my personal computer terminal - such as an automatic dialer which could extract phone numbers from my electronic mail messages and dial them at the touch of a button. It was a delightful visit and, as she left, I told her that I was looking forward to seeing her again soon.

Dahlia's next visit to her grandparents' house began on a Friday afternoon a few months later. While a couple of activities had been scheduled for the weekend, Dahlia suggested adding a Saturday trip to Bell Labs. Lillian tried to discourage the notion by pointing out that a visit outside normal working hours would take special permission. "Oh, I don't think that will be a problem, Grandma," the technically minded nine-year-old replied. She just sat herself down at Grandma's modem-equipped home terminal and began pressing the appropriate keys. "Arno Penzias must be able to give permission," she said. "I'll just send him mail and see."[2] If Dahlia's grandparents hadn't talked her out of it, Dahlia's imaginative use of her computer skills would have yielded the permission she sought in under five minutes.

Learning to program a computer is not an end in itself. Some of the most productive computer scientists I know never write programs. Rather, they work on underlying principles - such as the theory of algorithms. Algorithmic design is anything but rote; it calls for creative ideas rather than familiarity with computer machinery. For example, suppose I were to ask you to create an algorithm, namely: devise a method for finding all the anagrams in a piece of text (like "deal" and "lead" in "Can cane sugar lead to a good deal of acne?"). While a brute-force approach could get the job done on one page of text, only a sophisticated algorithm, working billions of times faster than brute force, would reveal the anagrams in something as large as a book. How would you do it?

If your plan consisted of matching the first word, letter by letter, against the letters in each of the words that followed it - and then did the same with the second word and so on - you can at least give yourself credit for understanding the problem. Comparing each word only to others of equal length is a further step in the right direction. But it would be better first to sort the words by length, so that words only needed to be compared within smaller groups. That's worth a part score. A simple way to keep from testing the same word more than once is to alphabetize all the words before starting the comparisons.

Jon Bentley, a friend of mine who teaches programming and writes about it, includes such examples in his courses to emphasize the need to back away from the machine and think creatively. Jon's answer to the above problem is *first* to alphabetize the letters in each word, and *then* alphabetize the resulting words. In that case the sentence "Can cane sugar lead to a good deal of acne?" would read "acn(1) acen(2) agrsu(3) adel(4) ot(5) a(6) dgoo(7) adel(8) fo(9) acen(10)" after the first alphabetization and "a(6) acen(2) acen(10) acn(1) adel(4)

2) Like most regular computer users, Dahlia uses the unmodified word "mail" to mean the electronic kind sent over computer networks.

adel(8) agrsu(3) dgoo(7) fo(9) ot(5)" after the second. Clearly words (2) and (10), as well as (4) and (8), form anagram pairs.

Alphabetizing the letters in each word makes the letters they contain easy to compare. That's the "trick".

In my view, the ability to find such creative solutions comes closer to the kind of "computer literacy" that people need in a high-tech world than does sitting at a keyboard memorizing commands.

Building Ideas

A magazine inteviewer once described me as "endlessly talkative." Most of my friends would agree with that assessment. I use words - in large numbers - to sift ideas and try them out on others.

On the other hand, a single well-chosen word can sometimes create an inspired image that can transform a familiar object into a new concept. When Akiro Morita led Sony into the production of transistorized radios, he wanted something more exciting than "the world's smallest portable radio." After some thought, Morita challenged his engineers to create a "pocketable" radio.

In practice, the early Sony models needed help from a tailor who provided Sony salesmen with oversized shirt pockets when the standard size proved a bit too snug for these radios. Nevertheless, the name "pocketable" not only gave designers a clear picture of Morita's vision, it also captured the imagination of the public. While Morita never said *which* pocket his radios would fit, his beautifully simple expression succeeded in getting his idea across to others.

There are more ways of combining words into grammatically correct English sentences than there are atoms in the universe. With so many possibilities to choose from, I see no disgrace in not finding the "right" one the first time. Moreover, a "bad" idea might be the first step to a better one.

The late Rudi Kompfner had more "bad" ideas than anyone else I've ever met. When I joined his radio research laboratory in 1961, Rudi was already one of the world's most celebrated electrical engineers. With a number of important inventions to his credit, he personally created several of the key elements that made satellite communications possible. But most of his ideas seemed hopelessly naive.

Rudi popped into my office one day with his latest "invention." At the time, our laboratory had been exploring a new communications network based on compact radio transmitters and receivers mounted on aluminum poles - somewhat larger versions of the ones used for highway lighting. But the system had a flaw. Whenever the poles swayed in the wind, the carefully aligned antennas would miss their intended targets. Rudi had "solved" the problem with a complicated rig for stabilizing the antennas atop their swaying poles. While I had to admit that this contraption could accomplish the job, I couldn't imagine that anyone would put up with all the extra struts and guy wires it called for. They took up too much real estate. There had to be a better way.

After we had gone back and forth exploring alternatives for a while, I hit upon the idea of putting the stabilizing wires inside the hollow pole itself. I was pretty sure I had something, but I knew it would take some work. Rudi said that was fine with him and promised to chat again in a few days.

By the time we met again, I had completed my design and even produced a small model of the tower - with a flashlight standing in for the antenna's radio beam - to prove my point. When Rudi gave my "tower" an energetic push, the motion caused my wires to rotate a pulley connected to the flashlight just enough to keep it level. Rudi was clearly delighted to see the flashlight's beam cast a steady spot on the far wall of my lab. "Good," he said, "now I can go think about something else." Rudi's "bad" idea had produced my first patent.

Before the German "Anschluss"*) forced Rudi to leave his native Vienna in the thirties, he had worked as an architect. His first day on the job, his new boss showed Rudi to a desk, gave him a map of a suburban building lot, and told him to design a two-bedroom house with a certain total floor space. That was it. Rudi was left on his own to carry out the task.

Stunned by the enormity of the task and afraid to put pencil to paper, Rudi sat immobilized. After a while, his boss returned, saw he was having trouble getting started, and offered to help. "Just put a square house right by the road," he said. "That way, they'll have a big backyard and won't have to shovel snow off their walk in winter." Rudi saw that his boss's plan wouldn't work. The people would want the privacy and appearance benefits a front yard offered. Furthermore, an L-shaped house would suit the triangular geometry of the lot. Rudi sketched what he had in mind.

*) the annexation of Austria by Hitler-Germany in 1938 (edt).

"Fine," his boss said. "Let's have the front door open right into the kitchen. That will make bringing in the groceries so much easier." Again Rudi objected, and pointed out a good place for a side entrance that could lead to a more conveniently located kitchen. By the time the exchange had moved to a third "suggestion," Rudi caught on to what his boss had been doing - verbalizing an idea, *any* idea, to get the creative process started.

When Rudi retired from Bell Labs, he moved on to Stanford University, where he taught a course in experimental design. One of his classes designed a wind-powered generator, which required a wind tunnel for operation. No wind tunnel being available, Rudi offered his ancient convertible in the service of research. His students mounted their models on the trunk lid, and with Rudi at the wheel, the group conducted their wind tests by driving around the Palo Alto campus as fast as the campus police would permit. While Rudi's class never succeeded in building a commercially viable windmill, his students certainly learned how to build ideas.

Estimates

Ideas flow better with needed facts at our fingertips - fewer interruptions for look-ups, and fewer detours down blind alleys. Fortunately, everyday experience has given each of us a large store of relevant information - much larger, in fact, than most of us suspect.

Do you know how fast the Mississippi River flows? No? Suppose I told you it flowed at eighty miles an hour, would you believe me? Just picture an old paddle wheeler floating gently along with the current. Surely it isn't outdistancing a speeding automobile. Even though you didn't know the speed to three decimal places, your prior knowledge let you make an *estimate* that showed my wild assertion to be invalid.

A very high barrier stands between us and the habit of making rough estimates - the fear of getting the "wrong" answer. Contrary to what most of us learned in school, however, an inexact answer is almost always good enough. All through elementary school, high school, and the first two years of college, I was taught that only the exact answer would do - -"July 14, 1789"; "5,280 feet"; "r-h-o-d-o-d-e-n-d-r-o-n." It was like a high-wire act. The slightest imbalance would send everything tumbling downward toward an inadequate safety net called "partial credit". I well remember the moment when the spell of that attitude was broken.

Henry Semat, my college atomic physics instructor, had completed a calculation on the blackboard, and several of us thought we had caught him in a "mistake." His answer was almost twice as big as the one given in the book - and he had written the book! Nevertheless, he didn't bat an eye when we pointed out the discrepancy to him. "Same order of magnitude," he shrugged with an air of total unconcern. "As long as we know that the effect is the right size, we can always fill in better numbers if we need to."

It was a revelation to find that a real-life practitioner of my intended profession didn't feel obligated to fill in every decimal place. An important lesson. Keep your eye on the left-hand digit and put zeros in all the other places. If you don't know the first digit, make a rough estimate and pick one.

I recently needed to know something about the long-term effect of radioactivity on transoceanic optical fibers. Cosmic rays flash regularly into our atmosphere from high-energy nuclear reactions, the product of massive stellar explosions within our galaxy. These rays constitute the preponderant source of the naturally created radioactivity that constantly bathes our planet. Some of these rays are so penetrating that they literally plow right through the entire earth and emerge out the other side to continue their travel through space. In addition to this natural effect, a transatlantic cable might also find itself near a drum of discarded nuclear waste, or an outcropping of radioactive uranium-bearing rock.

To learn what effects, if any, such radiation might have on the useful lifetime of glass fibers, I called a friend of mine, an expert on the subject. One input to the rough calculation we wanted to make was the amount of glass contained in a given length of optical fiber. She remembered "twenty-seven grams per meter" as the weight per unit length, but that didn't make sense to me. We'd both handled fibers very often and I couldn't imagine that a one-meter length of fiber would contain enough material to fill a shot glass.

A shot glass contains one liquid ounce, or about one-sixteenth of a pound of liquid. Two pounds (or pints) of water make up a quart, or 32 ounces, which is roughly equal to a liter. Since a liter weighs a kilogram - or 1,000 grams - dividing 1,000 (the grams) by 32 (the ounces) told me that the contents of a filled shot glass would weigh something like 30 grams. Since glass fiber is denser than water I figured it might take about half as much glass fiber to make the same weight of liquid - but still far more than one meter's worth. Once I made the connection, my friend quickly corrected her recollection to "twenty-seven grams of glass in a *kilometer* of fiber." The difference is 1,000 times.

My convoluted set of connections sounds really complicated, but that's more or less how I figured out that the original number had to be wrong. Since I'm so used to dealing with chains of numerical relationships, I completely overlooked a much easier way - one that only occurred to me as I was writing up this story. I needed simply to start (mentally) coiling up some fiber in a tight enough loop to fit into an envelope, and stop when I felt I had enough to need a second stamp. Since I had often handled optical fiber, it would have been easy for me to "see" that the coil would contain many dozens of turns - each about a foot long - so a single meter couldn't possibly weigh an ounce.

While the brute-force approach I took the first time around lacked the simple elegance of a comparison with first-class postage, it led me to the answer I needed. Both methods worked. The trick was to scan through familiar things with similar properties in order to build a path to the object in question. In order to get these, however, I had to make a number of rough estimates. Estimation gives problem solving the benefit of the imprecise knowledge our minds gather through everyday experience.

In most of my rough calculations, I treat the first digit as a 1 or a 3 and all the other digits as zeros. I like using three because it simplifies multiplication: 3 times 3 equals 10 in my system. Multiplying any given number by 10, 100, 1,000, or 1,000,000,000 can be accomplished by shifting the decimal point one, two, three, or nine places to the right, respectively. Similarly, division moves the decimal place to the left. In other words, it's simply a matter of adding or subtracting zeros.

When my son, David, was looking for his first job, one interviewer asked him, "How many barbers are there in the United States?" Not every engineering graduate would welcome that kind of question, but it was a good way of finding out whether or not a prospective co-worker could deal with the kinds of things not taught explicitly in engineering school.

David remembered that there were four barbershops on the main street of our town with a total of about ten barbers. Since our town has something over 10,000 people in it, that worked out to one barber per thousand, or about 200,000 barbers in a nation of 200 million people. The interviewer used a different method: one haircut per month for each of the 100 million people who get haircuts, and 400 haircuts per month per barber, which works out to 250,000.

When David presented me with the problem, I used money instead. I figured that each of the 100 million men who patronize barbershops spends $100 a year on haircuts, or $10 billion all together. I guessed that each barber must take in about $30,000 a year, in order to make a living and pay for a share of the shop, which gave me about 300,000 barbers.[3)] The actual number is just under 100,000, according to the U.S. Department of Labor.

While some chains of connections miss the actual answer by a wider margin than others, there is only one *wrong* method for dealing with such problems: trying to memorize answers to all the questions one might be asked. My son's interviewer was looking for someone who could find ways of probing an unfamiliar idea, rather than someone who relied on memorized facts alone.

Science

Progress depends on new ideas from inquiring minds. But not every idea is worth pursuing. Ideas need testing and refining in order to establish their validity. The scientific method offers us a particularly useful tool kit for testing ideas. While most people aren't employed as full-time professional scientists, we all need the ability to demystify the world through the creation and testing of hypotheses. That ability lies at the heart of all scientific progress.

My daughter's dog fails to greet me as I enter the house. "Where's Harvest? She isn't out in the yard because I would have heard her barking. I'll bet Laurie took her for a walk ... No, her lead is here on the hook. But the short leash is gone. Laurie must have taken Harvest in the car someplace. I think she mentioned needing to go to the veterinarian. Maybe she left a note." Test. Modify hypothesis. Test again.

The horn-shaped antenna Bob Wilson and I used for our detection of cosmic background radiation is located atop Crawford Hill in the suburban town of Holmdel, New Jersey. Geologically, the "hill" is just a pile of sand left behind by a glacier that visited our area about one million years ago, but it's high enough to offer a clear view of the surrounding countryside - including New York City's skyline some thirty miles to the northeast. While the horn had proven to be a far more precise instrument than its original design specifications had called for, it wasn't suited to the satellite communications and radio astronomy studies we wanted to carry out at much shorter wavelengths. As a result, we built a new antenna in the early 1970s to coincide with the launching of AT&T's first Comstar satellite.

Shortly after the new antenna was completed on the same hill, a Holmdel neighbor began having trouble with his television set. For no apparent reason, the picture on one of his channels would break up in the middle of a program - even though all the other channels

3) Culturally, I seem too far removed from this business to make accurate estimates. Men in New York, I'm told, often pay $25 per haircut. Apparently some barbers make quite a bit more than I imagined.

were fine. It didn't happen very often, so he paid scant attention at first. But the effect persisted. Maybe it had something to do with the new antenna? He called Bell Labs, but our community relations representative assured him that none of the antennas on Crawford Hill had transmitters, they only monitored signals from space - signals that were far too weak to be picked up by a TV set.

Since none of his neighbors had similar trouble, he got his set checked, but without result. What was left? The next time it happened, he ran outside to see whether anything unusual was going on. Apart from normal traffic, and some activity around the antenna, nothing caught his eye. But he persisted. A few more trips outside renewed his suspicions about the antenna. There seemed to be something going on over there every time his set misbehaved.

His next call to Bell Labs was more insistent - he wanted to speak to the person "in charge of the antenna." That's how I found myself on the phone with him. At first, I tried to explain that our antenna couldn't be the source of an interfering radio signal, since we weren't sending out *any* signals. "But I've looked into every other possibility," he said. "And how come it never happens when all the lights at the antenna are off and there's no one around?" That's when I decided we'd better have a look ourselves.

When Mike Gans, one of our engineers, arrived at the house, he found the neighbor's TV set working perfectly. "Get the antenna to do something," our neighbor urged. Feeling a little foolish, Mike called up his friend in the control cab and asked to have the antenna moved. "Moving up in elevation," his friend reported. "Ten degrees, twenty, thirty. . ." At "sixty," the picture began to flutter. At "seventy," it washed out completely, and then recovered as the antenna passed "eighty." Dumbfounded, Mike had the antenna moved in the opposite direction. Sure enough, as the antenna moved through the earlier position, the picture broke up again.

Confronted with the evidence, we had little trouble locating the "interfering transmitter." In those days, northern New Jersey got its TV broadcasts from New York City's Empire State Building. (These transmitters have since moved to the higher World Trade Center.) While receivers with a clear line of sight to the transmitter get the best reception, most viewers can get acceptable reception from reflected signals that bounce around obstacles. Our neighbor got an extra bounced signal from the Empire State Building - via our antenna - which exactly canceled the one his set depended on.

To get this interference, the two signals had to be exactly the same size and their paths had to differ by exactly one-half of a wavelength - so that the "crest" of one wave would arrive at the same time as the "trough" from the other. The odds against such a coincidence were so high that we never even considered the possibility.

The remedy was quite simple. We moved our neighbor's TV antenna to the other end of his roof. That move sufficiently changed the relative path lengths of the two signals to negate the exact cancellation. Had the original installer picked that end of the roof in the first place, the problem would never have come up. All's well that ends well. Our neighbor got his reception back, and we got an unexpected example of the successful application of the scientific method - from a non-scientist. In addition to solving puzzles, science also builds understanding by revealing the properties of the world and the relationships between them. Here again, the methods that scientists employ find widespread use in everyday life. From infancy onward, each person measures and classifies the properties of unfamiliar objects in order to integrate them into a larger worldview - from a ten-month-old learning to stack blocks, to Charles Darwin cataloging specimens aboard the *Beagle*.

My favorite examples of the scientific pursuit of knowledge come from antiquity. Almost 1,800 years before Ferdinand Magellan's ships sailed around the globe, an astronomer named Eratosthenes of Cyrene conceived and carried out the first accurate measurements of the earth's circumference some twenty-two centuries ago - but without leaving the African continent.

Simple observations will confirm that the world isn't flat. If you spot a ship at sea, and are sufficiently sharp-eyed, you will observe the ship "sinking" below the horizon as it moves away. At the same time, however, one can't use this effect to measure the curvature of the earth's surface reliably. Atmospheric effects blur and distort such distant images. Eratosthenes and his fellow astronomers could watch the steady motion of the sun, moon, and stars across the heavens break up into erratic fluctuations as these objects touched the horizon. The "sinking" ships might be some form of optical illusion, like a mirage in the desert.

If the earth were indeed round, proving it required a different kind of test. Eratosthenes needed some other way to establish a connecting bridge between the size of the "globe" beneath his feet and some measurable phenomena available within his own realm of personal experience. From experiments with shadows cast by sunlit objects at long distances from one another, he knew that the sun's rays moved in parallel lines. Thus, if the sun were overhead at one spot on the globe, sunlight would be tilted from the vertical at other locations.

Eratosthenes also knew from past experience that the noon sun was exactly overhead in the Egyptian city of Syene at each year's summer solstice.[4] How did he know? Because it shone to the bottom of a very deep well. Eratosthenes had also learned that, at the same moment at which the sun shone into the well at Syene, the shadow cast by an obelisk in Alexandria (some 500 miles to the north along the Nile) showed the sun to be some 7.5 degrees from the zenith.

Eratosthenes set about measuring the exact distance between the two cities. He gathered a group of people and trained them to take steps of a fixed size by having them march back and forth in a courtyard. He organized teams, gave each team a stretch of the 500 miles to pace off, and collected the total. In this way, he obtained a fairly accurate reading of the distance. (The word mile comes from the Latin word for thousand. A pace - or pair of steps - is about 5 feet long, and 1,000 paces gives us a mile.)

From geometry, Eratosthenes knew that moving 1/48th of the circumference of a circle tilted the local vertical by an angle of 7.5 degrees, the angle equal to the one cast by the obelisk. As a result, Eratosthenes concluded that the earth's circumference was 48 times the measured distance of 500 miles, or 24,000 miles (more precisely 24,662 miles). The modern value is 24,857 miles.

A few years earlier, Eratosthenes's colleague, Aristarchus of Samos, had used similar geometrical methods to measure the relation between the size of the earth and the distances to both the moon and the sun,[5] describing his work in a celebrated treatise entitled "On the Sizes and Distances of the Sun and Moon." We are fortunate that this literary treasure survived the Moslem conquest of the Roman Empire. Aristarchus's treatise was reintroduced into Europe at the end of the Dark Ages and helped pave the way for modern astronomy. (Copernicus referred to Aristarchus in his own work, for instance, but deleted the reference before publication.)

Geometrical measurements of the solar system improved as better instruments were introduced, but the principles remained the same for almost twenty centuries until

---

[4] The day each year on which the sun reaches its northernmost point, usually June 20 or 21.

[5] Aristarchus measured the size of the earth's shadow on the moon during an eclipse (with the earth between the sun and moon). He estimated the diameter of the moon to be one-third that of the earth, an error of only 8 percent. He found the sun to be twenty times more distant than the moon. While the actual value is almost four hundred, his value provided humankind's first inkling of the solar system's enormous (by terrestrial standards) size.

supplanted by modern technology in the summer of 1962 when the first radar signals bounced off the sun's surface and gave us a more accurate value.

When I described Eratosthenes's and Aristarchus's measurements of the earth, sun, and moon to a colleague at lunch one day, I ended up covering the back of a menu with lines and circles. With my explanation finished, she said, "You know, geometry sits right behind the eyeballs!" I very much agree. Geometric ideas have a dominant visual component that still appeals to the human intellect - knowledge for the sheer joy of knowing. Unlike machines, the human mind can both create ideas and enjoy them as well.

Although primitive by modern standards, the technology of the time played an essential role in Eratosthenes's and Aristarchus's exploration of ideas. Throughout the ages, technology has helped shape the facts we humans think about. As our knowledge has increased, so have our tools and the ways we employ them. Today, technology is so complex and pervasive that it dominates much of the environment in which human beings live and work. For this reason, I feel we need a better understanding of how technology affects the ways in which we now create and explore ideas.

XI

The prize was awarded in 1965 Richard P. Feynman for his fundamental work in quantum electrodynamics, with deep-ploughing consequences for the physics of elementary particles.

* * *

What is Science?

by Richard P. Feynman.

...What is science? Of course you all must know, if you teach it. That's common sense. What can I say? If you don't know, every teacher's edition of every textbook gives a complete discussion of the subject. There is some kind of distorted distillation and watered-down and mixed-up words of Francis Bacon from some centuries ago, words which then were supposed to be the deep philosophy of science. But one of the greatest experimental scientists of the time who was really doing something, William Harvey, said that what Bacon said science was, was the science that a lord-chancellor would do. He spoke of making observations, but omitted the vital factor of judgement about what to observe and what to pay attention to...

I am just going to tell you how I learned what science is. That's a little bit childish. I learned it as a child. I have had it in my blood from the beginning. And I would like to tell you how it got in. This sounds as though I am trying to tell you how to teach, but that is not my intention. I'm going to tell you what science is like by how I learned what science is like.

My father did it to me. When my mother was carrying me, it is reported - I am not directly aware of the conversation - my father said that "If it's a boy, he'll be a scientist." How did he do it? He never told me I should be a scientist. He was not a scientist: he was a businessman, a sales manager of a uniform company, but he read about science and loved it.

When I was very young - the earliest story I know - when I still ate in a high chair, my father would play a game with me after dinner. He had brought a whole lot of old rectangular bathroom floor tiles from some place in Long Island City. We set them up on end, one next to the other, and I was allowed to push the end one and watch the whole thing go down. So far, so good.

Next, the game improved. The tiles were different colors. I must put one white, two blues, one white, two blues, and another white and then two blues - I may want to put another blue, but it must be a white. You recognize already the usual insidious cleverness: first delight him in play, and then slowly inject material of educational value!

Well, my mother, who is a much more feeling woman, began to realize the insidiousness of his efforts and said, "Mel, please let the poor child put a blue tile if he wants to." My father said, "No, I want him to pay attention to patterns. It is the only thing I can do that is mathematics at this earliest level." If I were giving a talk on "what is mathematics," I would already have answered you. Mathematics is looking for patterns. (The fact is that this education had some effect. We had a direct experimental test at the time I got to kindergarten. We had weaving in those days. They've taken it out; it's too difficult for children. We used to weave colored paper through vertical strips and make patterns. The kindergarten teacher was so amazed that she sent a special letter home to report that this child was very unusual, because he seemed to be able to figure out ahead of time what pattern he was going to get, and made amazingly intricate patterns. So the tile game did do something to me.)

I would like to report other evidence that mathematics is only patterns. When I was at Cornell, I was rather fascinated by the student body, which seems to me was a dilute mixture of some sensible people in a big mass of dumb people studying home economics, etc. including lots of girls. I used to sit in the cafeteria with the students and eat and try to overhear their conversations and see if there was one intelligent word coming out. You can imagine my surprise when I discovered a tremendous thing, it seemed to me.

I listened to a conversation between two girls, and one was explaining that if you want to make a straight line, you see, you go over a certain number to the right for each row you go up, that is, if you go over each time the same amount when you go up a row, you make a straight line. A deep principle of analytic geometry! It went on. I was rather amazed. I didn't realize the female mind was capable of understanding analytic geometry.

She went on and said, "Suppose you have another line coming in from the other side, and you want to figure out where they are going to intersect. Suppose on one line you go over two to the right for every one you go up, and the other line goes over three to the right for every one that it goes up, and they start, twenty steps apart", etc. - I was flabbergasted. She figured out where the intersection was! It turned out that one girl was explaining to the other how to knit argyle socks.

I, therefore, did learn a lesson: The female mind is capable of understanding analytic geometry. Those people who have for years been insisting (in the face of all obvious evidence to the contrary) that the male and female are equally capable of rational thought may have something. The difficulty may just be that we have never yet discovered a way to communicate with the female mind. If it is done in the right way, you may be able to get something out of it.

Now I will go on with my own experience as a youngster in mathematics. Another thing that my father told me - and I can't quite explain it, because it was more an emotion than a telling - was that the ratio of the circumference to the diameter of all circles was always the same, no matter what the size. That didn't seem to me too unobvious, but the ratio had some marvelous property. That was a wonderful number, a deep number, pi. There was a mystery about this number that I didn't quite understand as a youth, but this was a great thing, and the result was that I looked for pi everywhere.

When I was learning later in school how to make the decimals for fractions, and how to make 3 1/8, I wrote 3.125, and thinking I recognized a friend wrote that it equals pi, the ratio of circumference to diameter of a circle. The teacher corrected it to 3.1416.

I illustrate these things to show an influence. The idea that there is a mystery, that there is a wonder about the number was important to me, not what the number was. Very much later when I was doing experiments in the laboratory - I mean my own home laboratory - fiddling around - no, excuse me, I didn't do experiments, I never did; I just fiddled around. I made radios and gadgets. I fiddled around. Gradually through books and manuals I began to discover there were formulas applicable to electricity in relating the current and resistance, and so on. One day looking at the formulas in some book or other, I discovered a formula for the frequency of a resonant circuit which was $f = 1/2\text{Đ}(LC)^{1/2}$ where L is the inductance and C the capacitance of the circuit. And there was pi, and where was the circle? You laugh, but I was very serious then. Pi was a thing with circles, and here is pi coming out of an electric circuit. Where was the circle? Do those of you who laughed know how that Đ comes about?

I have to love the thing. I have to look for it. I have to think about it. And then I realized, of course, that the coils are made in circles. About a half year later, I found another book which gave the inductance of round coils and square coils, and there were other pi's in those formulas. I began to think about it again, and I realized that the pi did not come from the circular coils. I understand it better now, but in my heart I still don't quite know where that circle is, where that pi comes from.

When I was still pretty young - I don't know how old exactly - I had a ball in a wagon I was pulling, and I noticed something, so I ran up to my father to say that "When I pull the wagon, the ball runs to the back, and when I am running with the wagon and stop, the ball runs to the front. Why?"

How would you answer?

He said, "That nobody knows!" He said "It's very general, though, it happens all the time to anything; anything that is moving tends to keep moving; anything standing still tries to maintain that condition. If you look close you will see the ball does not run to the back of the wagon where you start from standing still. It moves forward a bit too, but not as fast as the wagon. The back of the wagon catches up with the ball which has trouble getting started moving. It's called inertia, that principle." I did run back to check, and sure enough the ball didn't go backwards. He put the difference between what we know and what we call it very distinctly.

Regarding this business about names and words, I would tell you another story. We used to go up to the Catskill Mountains for vacations. In New York, you go to the Catskill Mountains for vacations. The poor husbands had to go to work during the week, but they would come rushing out for weekends and stay with the families. On the weekends, my father would take me for walks in the woods. He often took me for walks, and we learned all about nature, and so on, in the process. But the other children, friends of mine also wanted to go, and tried to get my father to take them. He didn't want to, because he said I was more advanced. I'm not trying to tell you how to teach, because what my father was doing was with a class of just one student; if he had a class of more than one, he was incapable of doing it.

So we went alone for our walk in the woods. But mothers were very powerful in those days as they are now, and they convinced the other fathers that they had to take their own sons out for walks in the woods. So all fathers took all sons out for walks in the woods one Sunday afternoon. The next day, Monday, we were playing in the fields and this boy said to me, "See that bird standing on the wheat there? What's the name of it?" I said, "I haven't got the slightest idea." He said, "It's a brown-throated thrush. Your father doesn't teach you much about science."

I smiled to myself, because my father had already taught me that doesn't tell me anything about the bird. He taught me "See that bird? It's a brown-throated thrush, but in Germany it's called a Halzenflügel, and in Chinese they call it a chung ling and even if you know all those names for it, you still know nothing about the bird. You only know something about people; what they call that bird."

"Now that thrush sings, and teaches its young to fly, and flies so many miles away during the summer across the country, and nobody knows how it finds its way", and so forth. There is a difference between the name of the thing and what goes on.

The result of this is that I cannot remember anybody's name, and when people discuss physics with me they often are exasperated when they say "the Fitz-Cronin effect" - and I ask "What is the effect?" and I can't remember the name.

I would like to say a word or two - may I interrupt my little tale - about words and definitions, because it is necessary to learn the words. It is not science. That doesn't mean just because it is not science that we don't have to teach the words. We are not talking about what to teach; we are talking about what science is. It is not science to know how to change Centigrade to Fahrenheit. It's necessary, but it is not exactly science. In the same sense, if you were discussing what art is, you wouldn't say art is the knowledge of the fact that a 3-B pencil is softer than a 2-H pencil. It's a distinct difference. That doesn't mean an art teacher shouldn't teach that, or that an artist gets along very well if he doesn't know that. (Actually you can find out in a minute by trying it; but that's a scientific way that art teachers may not think of explaining.)

In order to talk to each other, we have to have words, and that's all right. It's a good idea to try to see the difference, and it's a good idea to know when we are teaching the tools of science, such as words, and when we are teaching science itself.

To make my point still clearer, I shall pick out a certain science book to criticize unfavorably, which is unfair, because I am sure that with little ingenuity, I can find equally unfavorable things to say about others.

There is a first-grade science book which, in the first lesson of the first grade, begins in an unfortunate manner to teach science, because it starts off on the wrong idea of what science is. There is a picture of a dog, a windable toy dog, and a hand comes to the winder, and then the dog is able to move. Under the last picture, it says "What makes it move?" Later on, there is a picture of a real dog and the question "What makes it move?" Then there is a picture of a motor bike and the question "What makes it move?" and so on.

I thought at first they were getting ready to tell what science was going to be about: physics, biology, chemistry. But that wasn't it. The answer was in the teachers edition of the book; the answer I was trying to learn is that "energy makes it move."

Now energy is a very subtle concept. It is very, very difficult to get right. What I mean by that it is not easy to understand energy well enough to use it right, so that you can deduce something correctly, using the energy idea. It is beyond the first grade. It would be equally well to say that "God makes it move", or "spirit makes it move", or "movability makes it move." (In fact one could equally well say "energy makes it stop.")

Look at it this way: That's only the definition of energy. It should be reversed. We might say when something can move that it has energy in it, but not "what makes it move is energy." This is a very subtle difference. It's the same with this inertia proposition. Perhaps I can make the difference a little clearer this way.

If you ask a child what makes the toy dog move, you should think about what an ordinary human being would answer. The answer is that you wound up the spring; it tries to unwind and pushes the gear around. What a good way to begin a science course. Take apart the toy; see how it works. See the cleverness of the gears; see the ratchets. Learn something about the toy, the way the toy is put together, the ingenuity of people devising the ratchets and other things. That's good. The question is fine. The answer is a little unfortunate, because what they were trying to do is teach a definition of what is energy. But nothing whatever is learned.

Suppose a student would say, "I don't think energy makes it move." Where does the discussion go from there?

I finally figured out a way to test whether you have taught an idea or you have only taught a definition. Test it this way: You say, "Without using the new word which you have just learned, try to rephrase what you have just learned in your own language. Without using the word "energy", tell me what you know now about the dog's motion." You cannot. So you learned nothing except the definition. You learned nothing about science. That may be all right. You may not want to learn something about science right away. You have to learn definitions. But for the very first lesson is that not possibly destructive?

I think, for lesson number one, to learn a mystic formula for answering questions is very bad. The book has some others - "gravity makes it fall"; "the soles of your shoes wear out because of friction." Shoe leather wears out because it rubs against the sidewalk and the little notches and bumps on the sidewalk grab pieces and pull them off. To simply say it is because of friction, is sad, because it's not science.

My father dealt a little bit with energy and used the term after I got a little bit of the idea about it. What he would have done I know, because he did in fact essentially the same thing - though not the same example of the toy dog. He would say, "It moves because the sun is shining" if he wanted to give the same lesson. I would say "No. What has that to do with the sun shining? It moved because I wound up the springs."

"And why, my friend, are you able to move to wind up the spring?"

"I eat."

"What, my friend, do you eat?"

"I eat plants."

"And how do they grow?"

"They grow because the sun is shining."

And it is the same with the dog. What about gasoline? Accumulated energy of the sun which is captured by plants and preserved in the ground. Other examples all end with the sun. And so the same idea about the world that our textbook is driving at is phrased in a very exciting way. All the things that we see that are moving, are moving because the sun is shining. It does explain the relationship of one source of energy to another, and it can be denied by the child. He could say, "I don't think it is on account of the sun shining", and you can start a discussion. So there is a difference. (Later, I could challenge him with the tides, and what makes the earth turn, and have my hand on mystery again.)

That is just an example of the difference between definitions (which are necessary) and science. The only objection in this particular case was that it was the first lesson. It must certainly come in later, telling you what energy is, but not such a simple question as "What makes a dog move?" A child should be given a child's answer. "Open it up; let's look at it."

During those walks in the woods, I learned a great deal. In the case of birds, for example, I already mentioned migration, but I will give you another example of birds in the woods. Instead of naming them, my father would say, "Look, notice that the bird is always pecking in its feathers. It pecks a lot in its feathers. Why do you think it pecks the feathers?"

I guessed it's because the feathers are ruffled, and he's trying to straighten them out. He said, "Okay, when would the feathers get ruffled, or how would they get ruffled?"

"When he flies. When he walks around, it's okay; but when he flies it ruffles the feathers."

Then he would say, "You would guess then when the bird just landed he would have to peck more at his feathers than after he has straightened them out and has just been walking around the ground for a while. Okay, let's look."

So we would look, and we would watch, and it turned out, as far as I could make out, that the bird pecked about as much and as often no matter how long he was walking on the ground and not just directly after flight.

So my guess was wrong, and I couldn't guess the right reason. My father revealed the reason.

It is that the birds have lice. There is a little flake that comes off the feather, my father taught me, stuff that can be eaten, and the louse eats it. And then on the louse, there is a little bit of wax in the joints between the sections of the leg that oozes out, and there is a mite that lives in there that can eat that wax. Now the mite has such a good source of food that it doesn't digest it too well, so from the rear end there comes a liquid that has too much sugar, and in that sugar lives a tiny creature, etc.

The facts are not correct. The spirit is correct. First I learned about parasitism, one on the other, on the other, on the other.

Second, he went on to say that in the world whenever there is any source of something that could be eaten to make life go, some form of life finds a way to make use of that source; and that each little bit of left over stuff is eaten by something.

Now the point of this is that the result of observation, even if I were unable to come to the ultimate conclusion, was a wonderful piece of gold, with a marvelous result. It was something marvelous.

Suppose I were told to observe, to make a list, to write down, to do this, to look, and when I wrote my list down, it was filed with 130 other lists in the back of a notebook. I would learn that the result of observation is relatively dull, that nothing much comes of it.

I think it is very important - at least it was to me - that if you are going to teach people to make observations, you should show that something wonderful can come from them. I learned then what science was about. It was patience. If you looked, and you watched, and you paid attention, you got a great reward from it (although possibly not every time). As a result, when I became a more mature man, I would painstakingly, hour after hour, for years, work on problems - sometimes many years, sometimes shorter times - many of them failing, lots of stuff going into the wastebasket; but every once in a while there was the gold of a new understanding that I had learned to expect when I was a kid, the result of observation. For I did not learn that observation was not worthwhile.

Incidentally, in the forest we learned other things. We would go for walks and see all the regular things, and talk about many things; about the growing plants, the struggle of the trees for light, how they try to get as high as they can, and to solve the problem of getting water higher than 35 or 40 ft, the little plants on the ground that look for the little bits of light that come through, all that growth, and so forth.

One day after we had seen all this, my father took me to the forest again and said, "In all this time we have been looking at the forest we have only seen half of what is going on, exactly half."

I said, "What do you mean?"

He said, "We have been looking at how all these things grow; but for each bit of growth, there must be the same amount of decay, otherwise the materials would be consumed forever. Dead trees would lie there having used up all the stuff from the air, and the ground, and it wouldn't get back into the ground or the air, and nothing else could grow, because there is no material available. There must be for each bit of growth exactly the same amount of decay."

There then followed many walks in the woods during which we broke up old stumps, saw funny bugs and funguses growing - he couldn't show me bacteria, but we saw the softening effects, and so on. I saw the forest as a process of the constant turning of materials.

There were many such things, descriptions of things, in odd ways. He often started to talk about a thing like this: "Suppose a man from Mars were to come down and look at the world." It's a very good way to look at the world. For example when I was playing with my electric trains, he told me that there is a great wheel being turned by water which is connected by filaments of copper, which spread out and spread out and spread out in all directions; and then there are little wheels, and all those little wheels turn when the big wheel turns. The relation between them is only that there is copper and iron, nothing else, no moving parts. You turn one wheel here, and all the little wheels all over the place turn, and your train is one of them. It was a wonderful world my father told me about.

You might wonder what he got out of it all. I went to MIT. I went to Princeton. I came home, and he said, "Now you've got a science education. I have always wanted to know something that I have never understood; and so, my son, I want you to explain it to me." I said yes.

He said, "I understand that they say that light is emitted from an atom when it goes from one state to another, from an excited state to a state of lower energy."

I said, "That's right."

"And light is a kind of particle, a photon, I think they call it."

"Yes."

"So if the photon comes out of the atom when it goes from the excited to the lower state, the photon must have been in the atom in the excited state."

I said. "Well, no."

He said, "Well, how do you look at it so you can think of a particle photon coming out without it having been in there in the excited state?"

I thought a few minutes, and I said, "I'm sorry; I don't know. I can't explain it to you."

He was very disappointed after all these years and years of trying to teach me something, that it came out with such poor results.

What science is, I think, may be something like this: There was on this planet an evolution of life to a stage that there were evolved animals, which are intelligent. I don't mean just human beings, but animals which play and which can learn something from experience (like cats). But at this stage each animal would have to learn from its own experience. They gradually develop, until some animal could learn from experience more rapidly and could even learn from another's experience by watching, or one could show the other, or he saw what the other one did. So there came a possibility that all might learn it, but the transmission was inefficient and they would die, and maybe the one who learned it died too, before he could pass it on to others.

The question is: Is it possible to learn more rapidly what somebody learned from some accident than the rate at which the thing is being forgotten, either because of bad memory or because of the death of the learner or inventors?

So there came a time, perhaps, when for some species the rate at which learning was increased, reached such a pitch that suddenly a completely new thing happened; things could be learned by one individual animal, passed on to another, and another fast enough that it was not lost to the race. Thus became possible an accumulation of knowledge of the race.

This has been called time-binding. I don't know who first called it this. At any rate, we have here some samples of those animals, sitting here trying to bind one experience to another, each one trying to learn from the other.

This phenomenon of having a memory for the race, of having an accumulated knowledge passable from one generation to another, was new in the world. But it had a disease in it. It was possible to pass on ideas which were not profitable for the race. The race has ideas, but they are not necessarily profitable.

So there came a time in which the ideas, although accumulated very slowly, were all accumulations not only of practical and useful things, but great accumulations of all types of prejudices, and strange and odd beliefs.

Then a way of avoiding the disease was discovered. This is to doubt that what is being passed from the past is in fact true, and to try to find out *ab initio*, again from experience, what the situation is, rather than trusting the experience of the past in the form in which it is passed down. And that is what science is; the result of the discovery that it is worthwhile rechecking by new direct experience, and not necessarily trusting the race experience from the past. I see it that way. That is my best definition.

I would like to remind you all of things that you know very well in order to give you a little enthusiasm. In religion, the moral lessons are taught, but they are not just taught once, you are inspired again and again, and I think it is necessary to inspire again and again, and to remember the value of science for children, for grown-ups, and everybody else, in several ways; not only that we will become better citizens, more able to control nature and so on. There are other things.

There is the value of the world view created by science. There is the beauty and the wonder of the world that is discovered through the results of these new experiences. That is to say, the wonders of the content which I just reminded you of; that things move because the sun is shining. (Yet, not everything moves because the sun is shining. The earth rotates independent of the sun shining, and the nuclear reaction recently produced energy on the earth, a new source. Probably volcanoes are generally moved from a source different from the shining sun.)

The world looks so different after learning science. For example, trees are made of air, primarily. When they are burned, they go back to air, and in the flaming heat is released the flaming heat of the sun which was bound in to convert the air into tree, and in the ash is the small remnant of the part which did not come from air, that came from the solid earth, instead.

These are beautiful things, and the content of science is wonderfully full of them. They are very inspiring, and they can be used to inspire others.

Another of the qualities of science is that it teaches the value of rational thought, as well as the importance of freedom of thought; the positive results that come from doubting that the lessons are all true. You must here distinguish - especially in teaching - the science from the forms or procedures that are sometimes used in developing science. It is easy to say, "We write, experiment, and observe, and do this or that." You can copy that form exactly. But great religions are dissipated by following form without remembering the direct-content of the teaching of the great leaders. In the same way, it is possible to follow form and call it science, but that is pseudoscience. In this way, we all suffer from the kind of tyranny we have today in the many institutions that have come under the influence of pseudoscientific advisers.

We have many studies in teaching, for example, in which people make observations, make lists, do statistics, and so on, but these do not thereby become established science, established knowledge. They are merely an imitative form of science - analogous to the South Sea islands airfields, radio towers, etc., made out of wood. The islanders expect a great airplane to arrive. They even build wooden airplanes of the same shape as they see in the foreigners' airfields around them, but strangely enough, their wood planes do not fly. The result of this pseudoscientific imitation is to produce experts, which many of you are. You teachers who are really teaching children at the bottom of the heap can maybe doubt the experts once in a while. Learn from science that you *must* doubt the experts. As a matter of fact, I can also define science another way: Science is the belief in the ignorance of experts.

When someone says, "Science teaches such and such", he is using the word incorrectly. Science doesn't teach anything; experience teaches it. If they say to you, "Science has shown such and such", you might ask, "How does science show it? How did the scientists find out? How? What? Where?" It should not be "science has shown," but "this experiment, this effect has shown". And you have as much right as anyone else, upon hearing about the experiments (but be patient and listen to *all* the evidence) to judge whether a sensible conclusion has been arrived at.

In a field which is so complicated that true science is not yet able to get anywhere, we have to rely on a kind of old-fashioned wisdom, a kind of definite straightforwardness. I am trying to inspire the teacher at the bottom to have some hope, and some self-confidence in common sense and natural intelligence. The experts who are leading you may be wrong.

I have probably ruined the system, and the students that are coming into Caltech no longer will be any good. I think we live in an unscientific age in which almost all the buffeting of communications and television words, books, and so on are unscientific. As a result, there is a considerable amount of intellectual tyranny in the name of science.

Finally, with regard to this time-binding, a man cannot live beyond the grave. Each generation that discovers something from its experience must pass that on, but it must pass that on with a delicate balance of respect and disrespect, so that the rate (now that it is aware of the disease to which it is liable) does not inflict its errors too rigidly on its youth, but it does pass on the accumulated wisdom, plus the wisdom that it may not be wisdom.

It is necessary to teach both to accept and to reject the past with a kind of balance that takes considerable skill. Science alone of all the subjects contains within itself the lesson of the danger of belief in the infallibility of the greatest teachers of the preceding generation.

So carry on. Thank you.

Permissions for reprinting

Albert Einstein: "Autobiographical Notes"

"Reprinted from Albert Einstein; Autobiographical Notes by Albert Einstein, translated by Paul A. Schilpp, pp 7-17, by permission of the publisher, Open Court Publishing Company, La Salle, Illinois. The book, Autobiographical Notes, is itself reprinted from The Library of Living Philosophers, Volume VII, Albert Einstein: Philosopher-Scientist, edited by Paul A. Schilpp, La Salle, Illinois: Open Court, 1970." USA, 1990.

Nevill Mott: "Can we really use solar energy?"

(c) by Prof. Dr. Nevill Mott: 63 Mount Pleasant, Aspley Guise, GB- Milton Keynes, MK 17 8JX. GB 1989

Arthur L. Schawlow: "Discovering Science"

(c) by Arthur L. Schawlow, Stanford University, Stanford, California 94305-4060, USA 1989.

Leon M. Lederman: "Low Pay and Long Hours"

(c) by Physics Today, 335 E. 45th Street, New York, NY 10017, USA 1990.

S. Chandrasekhar: "The pursuit of science: Its motivations"

(c) by The University of Chicago Press, 5801 South Ellis Avenue, Chicago, Illinois 60637, USA 1990.

N. Bloembergen: "Physics in our daily lives and physics as an intellectual adventure"

(c) by Prof. Dr. N. Bloembergen, Division of Applied Sciences, Harvard University, Pierce Hall, Cambridge, Mass. 02138, USA 1989.

J. Bardeen: "Semiconductor Research Leading to the Point Contact Transistor."

(c) The Nobel Foundation 1989.

Gerd Binnig and Heinrich Rohrer:
"Scanning tunneling microscopy - from birth to adolescence"

(c) The Nobel Foundation 1989.

Gerd Binnig: "Creativity"

English Translation taken from the book "Aus dem Nichts", pp. 15-17 and 22-24, by Gerd Binnig, (c) by R. Piper Gmbh & Co. KG, München 1989, ISBN 3-492-03353-9, FRG 1990

Arno Penzias: "Ideas"

Richard P. Feynman: "What is Science?"