

17

3136

94

PROFESSOR MAX BORN
Physics and Metaphysics

PROFESSOR ALBERT EINSTEIN
My Attitude to Quantum Theory

The Early History of Man - Finding Water - Semiconductors -
The Shower Cloud - The Instrumentation of the Living Body -
Hydraulic Models - Research Report and some Starter Notes



SCIENCE NEWS

17



3136-

34

verified
in

**SCIENCE
NEWS**

EDITED BY J. L. CRAMMER

17

PENGUIN BOOKS

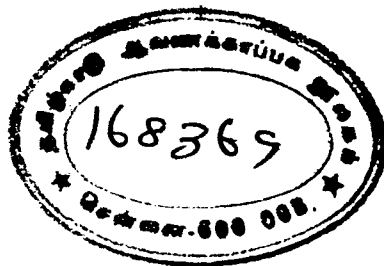
HARMONDSWORTH · MIDDLESEX

7L
Am1, N46

First published September 1950

N50-17 *

*Subscriptions to
SCIENCE NEWS
and
NEW BIOLOGY
are accepted at
7/- each
for four issues
including postage*



*Made and printed in Great Britain
for Penguin Books Ltd
by C. Nicholls & Company Ltd*

CONTENTS

EDITORIAL	7
PHYSICS AND METAPHYSICS Professor Max Born, F.R.S.	9
MY ATTITUDE TO QUANTUM THEORY Professor Albert Einstein, For.Mem.R.S.	28
THE EARLY HISTORY OF MAN L. S. B. Leakey	36
FINDING WATER Sir Cyril S. Fox	51
THE SHOWER CLOUD James Paton	67
SEMICONDUCTORS James T. Kendall	78
THE INSTRUMENTATION OF THE LIVING BODY J. W. S. Pringle	100
RESEARCH REPORT A. W. Haslett	115
YOUR GALVANIZED IRON ROOF D. J. Swaine	132
HYDRAULIC MODELS B. Standing	137
THE CARBON ARC Hugh Soden	149
ARTIFICIAL GEOLOGY Leslie Bor	153
ABOUT OUR CONTRIBUTORS	155
BOOKS RECEIVED	157
GLOSSARY	158
INDEX	159

24 illustrations between pages 64 and 65, including
NEAR THE ABSOLUTE ZERO

EDITORIAL

WHEN in 1926 Heisenberg contributed his principle of Indeterminacy to the development of the theory of quantum physics, a new epoch opened in popular philosophy. What was intended to be a specialized theoretical contribution to a restricted field of physics was quickly twisted by certain popularisers of science and by delighted clergymen into meaning the general overthrow of the laws of strict causality. The nineteenth-century view of the Universe as a great machine was suddenly seen to be hazy in fine detail; there was still room for human freewill, for miracles and the intervention of the hand of God; for once, modern science had been compelled to grant living-space to religion – thus ran the new popular philosophy. It was a philosophy based on a misunderstanding at its scientific root. Heisenberg had been concerned to point out that we cannot by the nature of our methods measure simultaneously both the velocity and the position of an atomic particle; thereby stimulating closer scientific scrutiny of methodology in physics. But he said nothing about the behaviour of particles being causeless, or any need to abandon belief in causality: this was added by the popularisers. Planck pointed all this out nearly twenty years ago in an interesting little book *Where is Science Going?* (1933) which seems to have been missed by the popular writers and the religious philosophers. There is an appendix recording conversations between Planck, Einstein, and the English translator Murphy, in which Einstein gives a clear warning against scientific writers in England who are illogical and romantic in their popular books, though quite the reverse in their scientific work. 'You must distinguish between the physicist and the *littérateur*,' he says, 'when both professions are combined into one.'

The question of causality was taken up again by Planck in an essay written just before his death in 1947 and recently published in English translation by Frank Gaynor in *A Scientific Autobiography and other Papers** by Max Planck. He draws a clear

* Williams and Norgate, 1950, 8s. 6d.

distinction (as does Professor Born on p. 11 of this issue) between determinism and causation: night always follows day, but we do not therefore believe that one causes the other, though the sequence is clearly determined. 'An occurrence is causally determined if it can be predicted with certainty,' he writes, but '... it is never possible to predict a physical occurrence with unlimited precision.' For him, the world picture of physics is composed of symbols, so that a physical occurrence is not an actual measured event in the objective real world, though the two are closely linked.

In this issue Professor Born and Professor Einstein discuss further the whole concept of causality and its meaning in modern physics; a re-introduction of philosophy (which has been excluded from our pages since *Science News* 3) as another experiment in our promised widening of scope. Perhaps it may do something to stimulate not only physicists and philosophers, but also churchmen and laymen, towards a careful mental spring-cleaning on a matter which is central to our understanding of the Universe.

PHYSICS AND METAPHYSICS

PROFESSOR M. BORN, F.R.S.

THE matter I wish to treat of is on the border-line of two fields of research, and this seems to imply that I am familiar with both of them. However, though I feel on fairly stable ground when speaking about physics, I cannot claim in any way to be expert in what is customarily treated in philosophical books and lectures under the title metaphysics. What I know of it is more or less the recollection from my student days, refreshed by some sporadic reading. Long years of neglect have not deleted the deep impression made in my youth by the age-old attempts to answer the most urgent questions of the human mind: the questions about the ultimate meaning of existence, about the Universe at large and our part in it, about life and death, truth and error, goodness and vice, God and eternity. But just as deep as this impression of the importance of the problems is the memory of the futility of the endeavour. There seemed to be no steady progress as we find in the special sciences, and like so many others, I turned my back on philosophy and found satisfaction in a restricted field where problems can actually be solved. Yet getting old, I feel, again like many others whose productive powers are declining, the desire to summarise the results of the scientific search in which, during several decades, I have taken a small part, and that leads unavoidably back to those eternal questions which go under the title of metaphysics.

Let me quote two definitions of metaphysics by modern philosophers. William James says: 'Metaphysics is an unusually stubborn effort to think clearly.' Bertrand Russell says: 'Metaphysics, or the attempt to conceive the world as a whole by means of thought.' These formulations stress two important aspects; one the method: stubborn clear thinking; the other the object: the world as a whole. But is every case of stubborn clear

thinking metaphysics? Every scientist, every historian, philologist, even theologian would claim to think clearly. On the other hand, the world as a whole is a subject not only vast, but definitely not closed, open to new discovery at any moment, therefore not exhausted and probably inexhaustible; in short the world known to us is never a whole. I shall return to this point at the end.

I propose to use the word metaphysics in a more modest way, in regard to both method and subject, namely as an investigation of the general features of the structure of the world and of our methods of dealing with this structure. I wish to discuss in particular the question whether the progress of physics has contributed anything essential to this problem. This progress of physics has been, as we are all aware, somewhat sensational during the last few years, and the aspect of the physical world has thoroughly changed in the half century of my own scientific life. Yet the methods of the physicist have always remained essentially the same: experimenting, observing regularities, formulating mathematical laws, predicting new phenomena with the help of these laws, combining the different empirical laws in coherent theories which satisfy our sense of harmony and logical beauty, and testing these theories again by prediction. These successful predictions are the highlights of theoretical physics, as we have witnessed in our day in the case of de Broglie's waves, of Dirac's positron, Yukawa's meson and many more such cases.

The power of prediction is the main claim of physics. It is based on the acceptance of the principle of causality, which, in its most general form, means the assumption of invariable laws of Nature. Yet you will all have heard that modern physics has been led to doubt this principle. Here is the first metaphysical conception on which I wish to make some comments.

Closely connected with it is the conception of reality. The sceptical attitude in regard to causality has arisen in atomic physics where the objects are not immediately accessible to our senses but only indirectly, with the help of more or less complicated apparatus. These ultimate objects of physics are

particles, forces, fields, etc.; what kind of reality can one ascribe to them? This leads to the more general question of the relation between subject and object, of the existence of an objective physical world independent of the observing subject, and thus back to Russell's problem, whether a conception of the world as a whole is actually possible.

The cause-effect relation is used in ordinary life in two rather different ways, which may be illustrated by the following two statements: 'The capitalistic system is the cause of economic crises,' and 'The economic crisis of 1930 was caused by a panic on the New York exchange.' One states a general rule or law, independent of time; the other declares one definite event to be the necessary sequence of another definite event. Both cases have in common the idea of necessity, a conception of a somewhat mysterious character which I feel completely unable to analyze further, and which I am willing to accept as metaphysical. Classical physics has officially adopted the second form of causality, as a necessary sequence in time. This came about through the discovery of the fundamental laws of mechanics by Galileo and Newton, laws which allow the prediction of future events from previous ones – or vice versa. In other words, these laws are deterministic: a world governed solely by them would be a gigantic machine, the complete knowledge of the situation at a given time would determine the situation at any other time. This kind of determinism was regarded by the physicists of the last century as the only rational interpretation of causality, and by using it they boasted that they had eliminated from physics the last remnants of metaphysical thinking.

Now it seems to me that this identification of determinism and causality is quite arbitrary and confusing. There are deterministic relations which are not causal; for instance, any timetable or programmatic statement. To take an absurdly obvious case, you could predict from the programme of a pantomime the sequence of the scenes, but would hardly say that the acrobats of scene No. 5 had caused the love scene No. 6. To return to science, the Ptolemaic system of the Cosmos is a deterministic but not a causal interpretation, and the same can be said about

Copernicus' cycles and Kepler's ellipses. They are all, in the usual scientific terminology, kinematic descriptions, but not causal explanations. For no cause of the phenomena is given except the ultimate cause of the creator's will. Then came the dynamical theories of Galileo and Newton. If one sticks to the programme that the only aim of a theory is deterministic prediction, the progress made by the introduction of dynamics into astronomy could merely be seen in a considerable condensation and simplification of the laws. When I was a student in Germany, fifty years ago, this standpoint, skilfully formulated by Kirchhoff, was dominant, and it is still widely shared.

I think that the discovery of mechanics was a much more fundamental affair. Galileo showed that a certain quantity connected with the motion of a body, namely its acceleration, is independent of the body and of its motion and dependent only on its position relative to the earth; and Newton showed the same for the planets, where the acceleration depends only on the distance from the sun. This appears to me something more than a short and efficient description of facts. It means the introducing of a quantitative expression of the cause-effect relation in its most general form, through the concept of force. It introduces the idea, foreign to the older kinematic theories, that one set of data (here positions) 'causes' another set of data (here accelerations). The word 'causes' means just 'determines quantitatively', and the law of force expresses in detail how the effect depends on the cause.

This interpretation of the laws of mechanics brings them in line with the ordinary practice of the scientist. An experiment is planned, i.e., certain conditions of observation are produced; then the effect is observed, sometimes at a future date, but more often all the time while the conditions hold. It is the timeless relation between observation and conditions of observation (apparatus) which is the real object of science. I suggest that this is the actual meaning of the principle of causality, as distinct from determinism, which is a special and almost accidental property of the mechanical laws (due to the fact, that one class of the quantities involved is accelerations, i.e., time derivatives).

If one looks on the history of physics during the last centuries from this point of view (as I have tried to do in my *Waynflete Lectures*, which have recently been published under the title *Natural Philosophy of Cause and Chance*), one gets the following impression.

Physics has used just this timeless cause-effect relation in its everyday practice, but another notion in the theoretical interpretation. There causality was taken as synonymous with determinism, and as the deterministic form of the mechanical laws is an empirical fact, this interpretation was hailed as a great achievement in eliminating dark metaphysical concepts. However, these concepts have a strange way of asserting themselves. Causality has in everyday life two attributes, which for shortness I shall call the principles of contiguity and of antecedence. The first states that things can act only on neighbouring things, or through a chain of things in contact, and the second that if cause and effect refer to situations at different times the cause should be prior to the effect.

Both principles are violated by Newtonian mechanics, as the gravitational force acts over any distance of empty space, and as the laws of motion connect two configurations at different times in a perfectly symmetric and reversible way. One can regard the whole development of classical physics as a struggle to re-establish these two essential features of the concepts of cause and effect. The methods of preserving contiguity were mathematically developed, by Cauchy and others, by extending mechanics to continuous media; the idea of contiguity played a leading part in Faraday's researches in electricity and magnetism, and led to Maxwell's concept of a field of force propagating itself with finite velocity, which was soon confirmed by Hertz's discovery of electro-magnetic waves. Finally Newton's case was brought into line with contiguity through Einstein's relativistic theory of the gravitational field. No modern theory of interaction is thinkable which violates this principle.

Antecedence has a much more tortuous history and not a happy end. It took much effort to discover that in physics the distinction between past and future was linked with the

irreversibility of heat phenomena – here we remember Joule as one of the central figures – and to reconcile this result with the reversibility of mechanics through the development of atomistics and statistical methods. I think that this work, initiated by Maxwell, Boltzmann, Gibbs and Einstein, is one of the greatest achievements in science. The deterministic interpretation of causality could be maintained for the atomic world and yet the apparent validity of antecedence understood as an effect of the statistical law of large numbers. However, this interpretation carried the germ of self-destruction of one of its pillars: it opened the way to the study of the atomic world, and the result was that the presupposed validity of Newtonian mechanics in this microscopic world was wrong. The new quantum mechanics does not allow of a deterministic interpretation, and since classical physics has identified causality with determinism, the doom of the causal explanation of nature seems to have come.

I am much opposed to this view. It does not matter much in discussions between scientists who know exactly what they are talking about; but it is harmful if used in describing the latest results of science to the non-scientific world. Extremes are always harmful. The deterministic mechanistic view produced a philosophy which shut its eyes to the most obvious facts of experience; but a philosophy which rejects not only determinism, but causation altogether, seems to me just as absurd. I think that there exists a reasonable definition of the cause-effect relation which I have already mentioned: that a certain situation depends on another one (irrespective of time) in a way describable by quantitative laws.

I shall indicate how this is still true in quantum mechanics in spite of its indeterministic character, and how the apparent loss is compensated by another fundamental principle, called complementarity, which will be of great philosophical and practical importance.

This new conception is due to Niels Bohr, the great Danish physicist, who was one of the leaders in the development of quantum mechanics, in regard not only to physics itself, but also to the philosophical implications. I was fortunate enough

to listen to his Gifford Lectures, given in Edinburgh last autumn, and which I hope will be published in the not too far distant future. I cannot give you an account of his ideas in the short space left to me, but try only to outline the main points and to bring them into line with my slightly different formulations.

As you will know, Planck's fundamental law of quantum theory connects an energy E with a frequency ν , by the simple formula $E = h\nu$, where h is a constant. This was later extended by Einstein and de Broglie, from the number of vibrations ν per unit time to the number of waves k per unit length, which is connected with a mechanical momentum p by the corresponding formula $p = hk$, with the same constant h .

That this is so has been confirmed by innumerable direct experiments and more or less indirect inferences from observations. Whenever a process can be resolved into periodic components with definite periods in time and space, i.e., with definite ν and k , the effect of it on the motion of particles consists in transferring energy and momentum according to this law. This empirical fact must be accepted as undeniable before its implications can be discussed.

Now this fact is so extremely strange that it took many years before physicists began to consider it seriously, and Niels Bohr himself has used the word 'irrational' to describe the new feature of the physical world discovered by Planck. Why irrational? Because the energy and momentum of a particle are, by their definition, related to an extremely small region of space, practically to a point, while frequency and wave number, also by their definition, are related to a very large, theoretically infinite, extension of space and time. This latter point will perhaps not appear so obvious as the first; you may say, I hear a tone of a piano string well defined even if it is played extremely staccato. This is practically true, because our ear is not a very sensitive instrument to discover tiny distortions. But the telecommunication engineer is familiar with the fact that there is a distortion. A tone lasting only a short time, comparable with the period, is no longer pure but is accompanied by other tones, with frequencies spread out over a little interval $\Delta\nu$ around the original

one; and if the duration becomes shorter and shorter this interval becomes larger and larger, until no tone is heard, but a noise, a crack. As modern telecommunication is based on the principle of modulation, i.e., of interrupting a high frequency current in the rhythm of signals or modifying its strength according to the relatively slow vibrations of speech or music, it is obvious that there is a limit to the perfection of transmission: if Δt is the duration of a tone of frequency ν , there is a relative limit of recognisability given in order of magnitude by $\Delta t \Delta \nu \sim 1$. An excellent account of these problems has been given by Dr Gabor in this country, and in America a book with the title *Cybernetics*, i.e. the science of governing – namely by sending out signals and orders – has been published by Norbert Wiener, which, though full of rather abstruse mathematics, has made quite a sensational stir. The mathematical analysis of these relations had its roots, almost one and a half centuries ago, in an investigation by Fourier on the conduction of heat. The main point is that the ideal, or pure, or harmonic vibration to which alone a sharp frequency can be ascribed, appears in a time-amplitude diagram as an endless train of sinusoidal waves. Every other curve, for instance a wave restricted to a finite interval of time, is a superposition of harmonic waves and has a whole 'spectrum' of ν -values. The same holds for real waves expanding in space where, apart from the periodicity in time, one has a periodicity in space measured by the wave number k . Between the length Δl of a train of waves and the width Δk of the k -spectrum one has the relation

$$\Delta l \Delta k \sim 1.$$

There is no logical way of dealing with periodic processes or waves other than this Fourier analysis, and practical applications have amply confirmed the theory.

Let us return to quantum physics. The 'irrationality' can now be formulated more precisely. In order to define ν and k sharply, one has to have $\Delta \nu$ and Δk very small, hence a very long duration $\Delta t \sim 1/\Delta \nu$ and spatial extension $\Delta l \sim 1/\Delta k$. So far nothing is different from the case of telecommunication, and

nothing paradoxical. But if one uses the relations $E = h\nu$, $P = \hbar k$, and re-writes the limiting relations in the form

$$\Delta t \Delta E \sim \hbar, \quad \Delta l \Delta p \sim \hbar,$$

they indicate a paradoxical situation: that with a tiny particle of sharp energy and momentum (i.e. small ΔE and Δp) there are associated long intervals of time and space Δt and Δl . What can the meaning of Δt and Δl be?

The only possible answer is, that they mean the limits for determining the position of the particle in time and space. They are indeed nothing other than Heisenberg's much discussed uncertainty relations.

Thus it is seen that the very first quantum laws lead necessarily to a mutual restriction in the accuracy attainable in space-time location on the one hand, and energy-momentum determination on the other. As Bohr has stressed again and again, we are confronted here with a logical alternative: either to deny the validity of an enormous amount of experience confirming the quantum laws $E = h\nu$, $p = \hbar k$, or to accept the existence of those limits for the determination of such pairs of quantities as time-energy, co-ordinate-momentum, which in the mechanical terminology are called conjugate. The most remarkable thing is that in spite of the completely new and revolutionary basic situation it was possible to develop a quantum mechanics which is a straightforward generalization of classical mechanics, extremely similar in mathematical form and considerably more perfect in its structure. It is true that the simple way of describing variable quantities as functions of time has to be given up and a more abstract method introduced where physical quantities are represented by non-commuting symbols (i.e. symbols with which one can form sums and products; the value of the latter, however, depends on the order of the factors). I shall never forget the thrill which I experienced when I succeeded in condensing Heisenberg's ideas on quantum conditions in the mysterious equation $pq - qp = \hbar/2\pi i$, which is the centre of the new mechanics and was later found to imply the uncertainty relations.

The transition from the symbols to actual quantities which can be measured is made by the introduction of a quantity called wave function, which describes the state in which a system is found so far as it can be described: its square is the probability density for finding the given data (e.g. co-ordinates of particles) in a given small region, analogous to the distribution function of ordinary statistics. There is, however, a fundamental difference. Suppose two beams of particles coming from the same source, counted separately, give the results ψ_1^2 and ψ_2^2 ; if by a suitable arrangement they can be made to overlap and be counted together, the result is $(\psi_1 + \psi_2)^2$, which differs from the sum $\psi_1^2 + \psi_2^2$ (by $2\psi_1\psi_2$). One has 'interference' of probabilities, as is well known from the case of light quanta or photons, the particles whose abundance is measured by the square of the intensity of an electro-magnetic wave. But I cannot enter into a technical description of wave mechanics, which has been developed from the foundations laid by de Broglie, through the ingenuity of Schrödinger, Dirac, and others. It suffices to say that a wave function ψ can be regarded as a packet of harmonic waves of different ν and k , and that the physical quantities like co-ordinates, momenta, energies q, p, E , are operators distorting the ψ -function, and thus determining the strengths of the harmonic components of the packet from which by squaring the probability of the appearance of particles with given $E = h\nu$, $p = \hbar k$ is obtained.

Thus the new mechanics is essentially statistical and, in regard to the distribution of particles, completely indeterministic. Yet it preserves, strangely enough, some similarity to classical mechanics, as the law of propagation of the function ψ , the so-called Schrödinger equation, is of the same type as the wave equations of elasticity or electro-magnetism. We have therefore the somewhat paradoxical situation, that there is no determinism for physical objects, like small particles, but for the probability of their appearance. Yet this determination of the ψ -function needs very many more data than those we are accustomed to in classical mechanics (initial positions and velocities of particles). In fact it needs a knowledge, or at least a hypothetical knowledge,

of ψ everywhere at a given time, and at the boundaries at all times, for the region and period in question; or in other words, predictions even of probabilities alone can be made only with reference to the whole situation, to the apparatus used. One must decide beforehand which feature one wishes to investigate, and one must construct the instrument accordingly. Then the effect can be predicted in terms of particles as a probability of their appearance under the conditions of the experiment (e.g. with given momentum), either at a certain finite region independent of time, or at a later time. That is in complete harmony with the meaning of causality which I have suggested. The use of this terminology is not a mere decoration; for it is essential to be clear that here the metaphysical, irreducible concept of necessity in the relation of two sets of things is postulated, which is the characteristic feature of the scientific attitude to the world.

Summarizing, we may say that while classical physics assumes natural phenomena to be going on independent of the incident of observation and describable without reference to observation, quantum physics claims only to describe and predict a phenomenon in relation to a well-defined mode of observation or instrumental arrangement. But one can, of course, use different instruments for observing the same class of phenomena; the propagation of light, for instance, can be investigated by prisms or gratings, with the help of photographic plates or Geiger counters. If every arrangement, from the standpoint of quantum mechanics, has to be considered separately, what is the common feature of all of them? For instance, if by one arrangement we can determine the spatial distribution of electrons, by others their distribution in energy, how can we know if and when we have exhausted all possibilities?

This question has been discussed in detail by Niels Bohr under the title Complementarity. It is true that he presents his ideas in a slightly different way: he is keenly intent to show by simple examples how one can intuitively understand the wholeness of an experimental situation and the mutual exclusiveness and complementarity of two such situations, by using nothing but the

uncertainty principle in its simplest form. I think that his motive in spending much ingenuity and effort on this task is the tragic situation that the philosophical attitude accepted by him and presented here by myself, accepted also by the whole international community of atomic physics, has not found favour in the eyes of just those men who have contributed most to the development of quantum theory, Planck and Einstein. Planck always preserved a cautious attitude to the revolutionary consequences of his own discovery, but Einstein went further and made repeated efforts to show by simple examples that the renunciation of determinism and the uncertainty relation are wrong. Just these examples have been studied by Bohr, in collaboration with Rosenfeld; in every case Einstein's objections could be refuted by a refined study of the experimental situation. The main point is that an instrument, by its very definition, is a physical system whose structure can be described in ordinary language and its functioning in terms of classical mechanics. Indeed, this is the only way in which we can communicate about it with one another. For instance, any spatial location needs a rigid frame, any measurement of time a mechanical clockwork, while on the other hand a determination of momentum and energy needs a break of rigidity and mechanical connection, a freely moveable part of the instrument to which the laws of conservation can be applied. Now Bohr shows that these two types of arrangement are mutually exclusive and complementary, in exact agreement with the results of the theory. If you use a diaphragm with a slit for fixing a co-ordinate of a particle passing through it, the diaphragm must be fixed to the frame of the instrument; if you wish to know whether a particle has really passed the slit, the diaphragm must be moveable so as to be able to recoil. You cannot have it both ways. By taking this complementarity into account, one can describe experiments without contradictions. Sometimes this is not very easy. I cannot refrain from indicating one example, which Einstein brought forward at the Solvay Conference in 1930, with the purpose of showing that it was possible to determine the exact time of an atomic event and the change of energy simultaneously, namely by making use of the relation

$E = mc^2$, derived from the theory of relativity. One has only to determine the mass m by weighing, to find the energy E . Assume radiation enclosed in a box with a shutter which is worked by clockwork inside the box and allows the escape of a given amount of energy, one or several photons, at a moment fixed with any desired accuracy. Moreover, you could weigh the whole box before and after this event and thus measure the energy released with any accuracy wanted, in contradiction to the reciprocal indeterminacy of time and energy assumed by quantum mechanics (Fig. 1). This seems to be a serious challenge. Bohr's answer is that the emission of energy is equivalent to a change of weight and therefore a displacement of the balance which must be compensated. But this displacement in the gravitational field of the earth is coupled with a change of rate of the clock. All these effects can be determined with some latitude of accuracy which depend on one another and produce the result: that Einstein's method does not work.

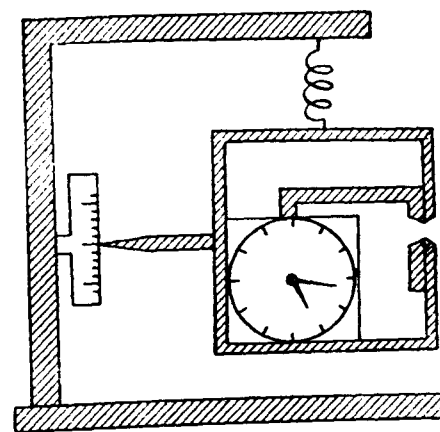


Fig. 1. The box with clockwork-operated shutter hangs from a form of spring balance.

I shall now describe this in more detail. As the uncertainty ΔT of a measurement of time is proportional to the time

measured, we must avoid delay and hang our box directly on the balance. If the shutter is opened the balance will move and can then be readjusted with an accuracy Δq . As this happens in the field of the earth g , there is a change of gravitational potential $\phi = gq$, at the place of the clock, which is fixed with a latitude $\Delta\phi = g \Delta q$. The reading of the clock in the time interval T necessary for this, according to the general theory of relativity, will have a relative uncertainty

$$\frac{\Delta T}{T} = -\frac{\Delta\phi}{c^2} = -\frac{g}{c^2} \Delta q.$$

If in this time T the weight of the box is determined with an accuracy Δm , we have from Newton's law of motion, for the latitude in measuring the momentum of the box, $\Delta p = g \Delta m T$. By substituting the values of Δq and Δp from these relations in

$$\Delta p \cdot \Delta q \sim h, \text{ one finds } h \sim \Delta p \cdot \Delta q = g \Delta m T \cdot \frac{c^2}{g} \frac{\Delta T}{T} =$$

$c^2 \Delta m \cdot \Delta t = \Delta E \cdot \Delta T$, according to the relativistic connection between mass and energy. Hence it is impossible to determine energy and time of release as well, with arbitrary accuracy.

You will find this and other examples in Bohr's Gifford Lectures. While I wrote this, a new book came into my hands, *Albert Einstein, Philosopher and Scientist* (The Library of Living Philosophers, ed. by Paul Arthur Schilpp, 1949), which contains articles by many philosophers and theoretical physicists on different aspects of Einstein's work, amongst them one by Niels Bohr and one by myself. The most interesting part of the work is a scientific autobiography by Einstein, and a summarizing article in which he answers the criticism in the previous essays.* This is most fascinating reading, but with all respect to the great physicist, I cannot accept his arguments against the philosophy of the quantum physicists. All essential points are treated in Bohr's article where he gives a delightful account of a number of discussions he had with Einstein. But the latter persists in his opposition, and declares himself firmly convinced

* See also page 28.

that the present theory, though logically consistent, is an incomplete description of physical systems. His main arguments are derived not so much from considerations of causality, but from the new attitude to the meaning of physical reality which it implies. Let me quote his words (p. 72): 'For me ... the expectation that the adequate formulation of the universal laws involves the use of all conceptual elements which are necessary for a complete description, is more natural', namely than the ideas of the quantum physicists, and he insists that the emission of, say, an α -particle by a radioactive atom with definite energy must happen at a definite time predictable from theory - otherwise he calls the description conceptionally incomplete. Yet he himself has taught us in the case of relativity that this argument is wrong. There you have an infinite number of equivalent inertial systems, each of which can be assumed to be at rest with the same right. But there is no way of deciding experimentally which is truly or absolutely at rest. Einstein's opponents pointed out that they regarded a description of the world as conceptionally incomplete which denied the existence of a system absolutely at rest, even if there is no experimental way of finding it. This anti-relativistic argument is just as strong as Einstein's anti-quantistic, as everybody has experienced who was asked to conceive a light wave without a material ether as a carrier of the vibrations.

The generation to which Einstein, Bohr and I belong was taught that there exists an objective physical world, which unfolds itself according to immutable laws independent of us; we are watching this process as the audience watches a play in a theatre. Einstein still believes that this should be the relation between the scientific observer and his subject. Quantum mechanics, however, interprets the experience gained in atomic physics in a different way. We may compare the observer of a physical phenomenon not with the audience at a theatrical performance, but with that at a football game, where the act of watching, accompanied by applauding or hissing, has a marked influence on the speed and concentration of the players, and thus on what is watched. In fact, a better simile is life itself, where

audience and actors are the same persons. It is the action of the experimentalist who designs the apparatus, which determines essential features of the observations. Hence there is no objectively existing situation, as was supposed to exist in classical physics. Not only Einstein, but also others who are not opposed to our interpretation of quantum mechanics, have said that under these circumstances there is no objectively existing external world, no sharp distinction between subject and object. There is of course some truth in it, but I do not consider this formulation very fortunate. For what do we mean by speaking of an objectively existing world? This is certainly a pre-scientific notion, never questioned by ordinary man. If he sees a dog, he sees a dog whether it sits beside him, jumps about, or runs away and disappears in the distance as a tiny spot. All these innumerable and vastly different sense impressions are united by an unconscious process in this mind to the one conception *dog*, which remains the same dog under all these aspects. I propose to express this by saying that the mind constructs, by an unconscious process, invariants of perception, and that these are what the ordinary man calls real things. And I think that science does exactly the same, only on a different level of perception, namely using all the magnifying devices which are the essence of observing and measuring.

The innumerable possible observations are linked again by some permanent features, invariants, which differ from those of ordinary perception, but are nevertheless in the same way indicators of things, objects, particles. For in describing what we observe even with the most refined instrument we have no other language than the ordinary one. Thus atomistic objects have, it is true, not all the properties of ordinary objects, but they have enough definite properties for us to ascribe to them physical reality of the same kind as to a dog. I think the fact that various observations of electrons give always the same charge, rest-mass and spin, perfectly justifies speaking of them as real particles.

Here is another point where I disagree with Einstein's philosophy. He accepts the doctrine of conventionalism which in

my youth was powerfully advocated by the great French mathematician Henri Poincaré. According to this view all human concepts are free inventions of the mind and conventions between different minds; they are justifiable only by their usefulness in ordinary experiences. This may be true in a restricted sense, namely for the abstract parts of theories, but not for the connection of the theories with observations, with real things. It neglects the psychological fact that the building of language is not a conscious process. And even in the abstract part of science the use of concepts is often decided by facts, not by conventions.

An instructive example is Schrödinger's attempt to interpret his electronic waves as a diffuse cloud of electricity, sacrificing the particle concept. It was soon abandoned, since electrons could be counted. The corpuscular character of the electron is certainly not a convention.

If we thus have to attribute a definite reality to the particles, what about the waves? Are they also real, and in what sense? It has been said that electrons appear sometimes as waves, sometimes as particles, perhaps changing over every Sunday and Wednesday, as a great experimentalist mockingly remarked, obviously in a fit of anger about the somersaults of the theorists. I cannot agree to this view. In order to describe a physical situation, one has to use both waves, describing a 'state', i.e., the whole experimental situation, and particles, the proper objects of atomic research. The square of a wave function representing probability has the character of reality. That probability has some kind of reality cannot be denied. How otherwise could a prediction based on probability calculus have any application to the real world? I am not deeply interested in the numerous attempts to make this more understandable. It seems to me, just like the necessity of the causal relations of classical physics, something beyond physics, a metaphysical idea. The same holds for the wave functions of quantum mechanics. One could call for the wave functions and particles in physics a duality in the description, which should be strictly distinguished from complementarity.

Let us now finally ask whether these new developments in

physics have any bearing on other subjects, and principally on the great problems of metaphysics. There is first the eternal dispute between idealism and realism in philosophy. I do not think that the new ways in physics can produce any weighty argument for one side or the other. Whoever believes that the only important reality is the realm of ideas, of the spirit, should not occupy himself with science. The scientist must be a realist, he must accept his sense impressions as more than a hallucination, as messages from a real outer world. In disentangling these messages he uses ideas of a very abstract kind, group theory in spaces of many or even infinitely many dimensions and things like that, but finally he has his observational invariants, representing real things with which he learns to operate like any craftsman with his wood or metal. Modern theory has made the part of the ideas more extended and refined, but has not changed the whole situation.

But a real enrichment of our thinking is the idea of complementarity. The fact that in an exact science like physics there are mutually exclusive and complementary situations which cannot be described by the same concepts, but need two kinds of expressions, must have an influence, and I think a welcome influence, in other fields of human activity and thought. Here again Niels Bohr has shown the way. In biology the concept of life itself leads to a complementary alternative: the physico-chemical analysis of a living organism is incompatible with its free functioning and leads in its extreme application to death. In philosophy there is a similar alternative in the central problem of free will. Any decision can be considered on the one side as a process in the conscious mind, on the other as a product of motives, implanted in the past or present from the outside world. If one sees in this an example of complementarity, the eternal conflict between freedom and necessity appears to be based on an epistemological error. But I cannot enter into the discussion of these questions, which are only just beginning to be seen in this way. Let me conclude by a remark on Russell's definition of metaphysics from which I started: that it is an attempt to conceive the world as a whole by means of thought. Has the lesson

in epistemology which we learned from physics any bearing on this problem? I think it has, in showing that even in restricted fields a description of the whole of a system in one picture is impossible; there are complementary images which do not apply simultaneously but are nevertheless not contradictory and exhaust the whole only together. This is, I think, a very healthy doctrine, which properly applied may remove many violent disputes not only in philosophy but in all ways of life. For instance, in politics. The president of the Russian Academy of Sciences, Professor Vavilov, has published (in *Vox*) an interesting article in which he explained the ideas of dialectical materialism and used as example the development of optics. The thesis 'light consists of particles' and the antithesis 'light consists of waves' fought with one another until they were united in the synthesis of quantum mechanics, and the same holds for electrons and other constituents of matter. That is all very well and indisputable. Only why not apply it to the thesis Liberalism (or Capitalism), the antithesis Communism, and expect a synthesis, instead of a complete and permanent victory for the antithesis? There seems to be some inconsistency. But the idea of complementarity goes deeper. In fact this thesis and antithesis represent two psychological motives and the corresponding economic forces, both justified in themselves but, in their extremes, mutually exclusive. Complete freedom of the individual in economic behaviour is incompatible with the existence of an orderly state, and the totalitarian state is incompatible with the development of the individual. There must exist a relation between the latitudes of freedom Δf and of regulation Δr , of the type $\Delta f \cdot \Delta r \sim p$, which allows a reasonable compromise. But what is the 'political constant' p ? I must leave this to a future quantum theory of human affairs. The world which is so ready to learn means of mass-destruction from physics would do better to accept the message of reconciliation contained in the philosophy of complementarity.

(This was the Joule Memorial Lecture, read to the Manchester Literary and Philosophical Society in February 1950.)

MY ATTITUDE TO QUANTUM THEORY

PROFESSOR ALBERT EINSTEIN, FOR.MEM.R.S.

I NOW come to what is probably the most interesting subject which absolutely must be discussed in connection with the detailed arguments of my highly esteemed colleagues Born, Pauli, Heitler, Bohr, and Margenau. They are all firmly convinced that the riddle of the double nature of all corpuscles (corpuscular and undulatory character) has in essence found its final solution in the statistical quantum theory. On the strength of the successes of this theory they consider it proved that a theoretically complete description of a system can, in essence, involve only statistical assertions concerning the measurable quantities of this system. They are apparently all of the opinion that Heisenberg's indeterminacy-relation (the correctness of which is, from my own point of view, rightfully regarded as finally demonstrated) is essentially prejudicial in favour of the character of all thinkable, reasonable physical theories in the sense mentioned. In what follows I wish to adduce reasons which keep me from falling into line with the opinion of almost all contemporary theoretical physicists. I am, in fact, firmly convinced that the essentially statistical character of contemporary quantum theory is to be ascribed solely to the fact that this [theory] operates with an incomplete description of physical systems.

Above all, however, the reader should be convinced that I fully recognize the very important progress which the statistical quantum theory has brought to theoretical physics. In the field of *mechanical* problems – i.e. wherever it is possible to consider the interaction of structures and of their parts with sufficient accuracy by postulating a potential energy between material points – [this theory] even now presents a system which, in its closed

character, correctly describes the empirical relations between statable phenomena as they were theoretically to be expected. This theory is until now the only one which unites the corpuscular and undulatory dual character of matter in a logically satisfactory fashion; and the (testable) relations, which are contained in it, are, within the natural limits fixed by the indeterminacy-relation, *complete*. The formal relations which are given in this theory – i.e. its entire mathematical formalism – will probably have to be contained, in the form of logical inferences, in every useful future theory.

What does not satisfy me in that theory, from the standpoint of principle, is its attitude towards what appears to me to be the programmatic aim of all physics: the complete description of any (individual) real situation (as it supposedly exists irrespective of any act of observation or substantiation). Whenever the positivistically inclined modern physicist hears such a formulation his reaction is that of a pitying smile. He says to himself: 'There we have the naked formulation of a metaphysical prejudice, empty of content, a prejudice, moreover, the conquest of which constitutes the major epistemological achievement of physicists within the last quarter-century. Has any man ever perceived a "real physical situation"? How is it possible that a reasonable person could to-day still believe that he can refute our essential knowledge and understanding by drawing up such a bloodless ghost?' Patience! The above laconic characterization was not meant to convince anyone; it was merely to indicate the point of view around which the following elementary considerations freely group themselves. In doing this I shall proceed as follows: I shall first of all show in simple special cases what seems essential to me, and then I shall make a few remarks about some more general ideas which are involved.

We consider as a physical system, in the first instance, a radioactive atom of definite average decay time, which is practically exactly localized at a point of the co-ordinate system. The radioactive process consists in the emission of a (comparatively light) particle. For the sake of simplicity we neglect the motion of the residual atom after the disintegration process. Then it is possible

for us, following Gamow, to replace the rest of the atom by a space of atomic order of magnitude, surrounded by a closed potential energy barrier which, at a time $t = 0$, encloses the particle to be emitted. The radio-active process thus schematized is then, as is well known, to be described – in the sense of elementary quantum mechanics – by a ψ -function in three dimensions, which at the time $t = 0$ is different from zero only inside of the barrier, but which, for positive times, expands into the outer space. This ψ -function yields the probability that the particle, at some chosen instant, is actually in a chosen part of space (i.e., is actually found there by a measurement of position). On the other hand, the ψ -function does not imply any assertion concerning the time instant of the disintegration of the radioactive atom.

Now we raise the question: Can this theoretical description be taken as the complete description of the disintegration of a single individual atom? The immediately plausible answer is: No. For one is, first of all, inclined to assume that the individual atom decays at a definite time; however, such a definite time-value is not implied in the description by the ψ -function. If, therefore, the individual atom has a definite disintegration-time, then as regards the individual atom its description by means of the ψ -function must be interpreted as an incomplete description. In this case the ψ -function is to be taken as the description, not of a singular system, but of an ideal ensemble of systems. In this case one is driven to the conviction that a complete description of a single system should, after all, be possible; but for such complete description there is no room in the conceptual world of statistical quantum theory.

To this the quantum theorist will reply: This consideration stands and falls with the assertion that there actually is such a thing as a definite time of disintegration of the individual atom (an instant of time existing independently of any observation). But this assertion is, from my point of view, not merely arbitrary but actually meaningless. The assertion of the existence of a definite time-instant for the disintegration makes sense only if I can in principle determine this time-instant empirically. Such

an assertion, however, (which, finally, leads to the attempt to prove the existence of the particle outside of the force barrier), involves a definite disturbance of the system in which we are interested; so that the result of the determination does not permit a conclusion concerning the status of the undisturbed system. The supposition, therefore, that a radioactive atom has a definite disintegration-time is not justified by anything whatsoever; it is, therefore, not demonstrated either that the ψ -function cannot be conceived as a complete description of the individual system. The entire alleged difficulty proceeds from the fact that one postulates something not observable as 'real'. (This is the answer of the quantum theorist.)

What I dislike in this kind of argumentation is the basic positivistic attitude, which from my point of view is untenable, and which seems to me to come to the same thing as Berkeley's principle, *esse est percipi*. 'Being' is always something which is mentally constructed by us, that is, something which we freely posit (in the logical sense). The justification of such constructs does not lie in their derivation from what is given by the senses. Such a type of derivation (in the sense of logical deducibility) is nowhere to be had, not even in the domain of pre-scientific thinking. The justification of the constructs, which represent 'reality' for us, lies alone in their quality of making intelligible what is sensorily given (the vague character of this expression is here forced upon me by my striving for brevity). Applied to the specifically chosen example this consideration tells us the following:

One may not merely ask: 'Does a definite time instant for the transformation of a single atom exist?' but rather: 'Is it, within the framework of our theoretical total construction, reasonable to posit the existence of a definite point of time for the transformation of a single atom?' One may not even ask what this assertion means. One can only ask whether such a proposition, within the framework of the chosen conceptual system – with a view to its ability to grasp theoretically what is empirically given – is reasonable or not.

Insofar, then, as a quantum-theoretician takes the position

that the description by means of a ψ -function refers only to an ideal systematic totality but in no wise to the individual system, he may calmly assume a definite point of time for the transformation. But, if he represents the assumption that his description by way of the ψ -function is to be taken as the *complete* description of the individual system, then he must reject the postulation of a specific decay-time. He can justifiably point to the fact that a determination of the instant of disintegration is not possible on an isolated system, but would require disturbances of such a character that they must not be neglected in the critical examination of the situation. It would, for example, not be possible to conclude from the empirical statement that the transformation has already taken place, that this would have been the case if the disturbances of the system had not taken place.

As far as I know, it was E. Schrödinger who first called attention to a modification of this consideration, which shows an interpretation of this type to be impracticable. Rather than considering a system which comprises only a radioactive atom (and its process of transformation), one considers a system which includes also the means for ascertaining the radioactive transformation – for example, a Geiger-counter with automatic registration-mechanism. Let this latter include a registration-strip, moved by clockwork, upon which a mark is made by tripping the counter. True, from the point of view of quantum mechanics this total system is very complex and its configuration space is of very high dimension. But there is in principle no objection to treating this entire system from the standpoint of quantum mechanics. Here too the theory determines the probability of each configuration of all its co-ordinates for every time instant. If one considers all configurations of the co-ordinates, for a time large compared with the average decay-time of the radioactive atom, there will be (at most) *one* such registration-mark on the paper strip. To each co-ordinate-configuration corresponds a definite position of the mark on the paper strip. But, inasmuch as the theory yields only the relative probability of the thinkable co-ordinate-configurations, it also

offers only relative probabilities for the positions of the mark on the paper strip, but no definite location for this mark.

In this consideration the location of the mark on the strip plays the rôle played in the original consideration by the time of the disintegration. The reason for the introduction of the system supplemented by the registration-mechanism lies in the following. The location of the mark on the registration-strip is a fact which belongs entirely within the sphere of macroscopic concepts, in contradistinction to the instant of disintegration of a single atom. If we attempt [to work with] the interpretation that the quantum-theoretical description is to be understood as a complete description of the individual system, we are forced to the interpretation that the location of the mark on the strip is nothing which belongs to the system *per se*, but that the existence of that location is essentially dependent upon the carrying out of an observation made on the registration-strip. Such an interpretation is certainly by no means absurd from a purely logical standpoint; yet there is hardly likely to be anyone who would be inclined to consider it seriously. For, in the macroscopic sphere it simply is considered certain that one must adhere to the programme of a realistic description in space and time; whereas in the sphere of microscopic situations one is more readily inclined to give up, or at least to modify, this programme.

This discussion was only to bring out the following. One arrives at very implausible theoretical conceptions, if one attempts to maintain the thesis that the statistical quantum theory is in principle capable of producing a complete description of an individual physical system. On the other hand, those difficulties of theoretical interpretation disappear, if one views the quantum-mechanical description as the description of ensembles of systems.

I reached this conclusion as the result of quite different types of considerations. I am convinced that everyone who will take the trouble to carry through such reflections conscientiously will find himself finally driven to this interpretation of quantum-theoretical description (the ψ -function is to be understood as the description not of a single system but of an ensemble of systems).

Roughly stated the conclusion is this: Within the framework of statistical quantum theory there is no such thing as a complete description of the individual system. More cautiously it might be put as follows: The attempt to conceive the quantum-theoretical description as the complete description of the individual systems leads to unnatural theoretical interpretations, which become immediately unnecessary if one accepts the interpretation that the description refers to ensembles of systems and not to individual systems. In that case the whole 'egg-walking' performed in order to avoid the 'physically real' becomes superfluous. There exists, however, a simple psychological reason for the fact that this most nearly obvious interpretation is being shunned. For if the statistical quantum theory does not pretend to describe the individual system (and its development in time) completely, it appears unavoidable to look elsewhere for a complete description of the individual system; in doing so it would be clear from the very beginning that the elements of such a description are not contained within the conceptual scheme of the statistical quantum theory. With this one would admit that, in principle, this scheme could not serve as the basis of theoretical physics. Assuming the success of efforts to accomplish a complete physical description, the statistical quantum theory would, within the framework of future physics, take an approximately analogous position to the statistical mechanics within the framework of classical mechanics. I am rather firmly convinced that the development of theoretical physics will be of this type; but the path will be lengthy and difficult.

I now imagine a quantum theoretician who may even admit that the quantum-theoretical description refers to ensembles of systems and not to individual systems, but who, nevertheless, clings to the idea that the type of description of the statistical quantum theory will, in its essential features, be retained in the future. He may argue as follows: True, I admit that the quantum-theoretical description is an incomplete description of the individual system. I even admit that a complete theoretical description is, in principle, thinkable. But I consider it proven that

the search for such a complete description would be aimless. For the lawfulness of nature is so constituted that the laws can be completely and suitably formulated within the framework of our incomplete description.

To this I can only reply as follows: Your point of view – taken as theoretical possibility – is incontestable. For me, however, the expectation that the adequate formulation of the universal laws involves the use of *all* conceptual elements which are necessary for a complete description, is more natural. It is furthermore not at all surprising that, by using an incomplete description (in the main), only statistical statements can be obtained out of such description. If it should be possible to move forward to a complete description, it is likely that the laws would represent relations among all the conceptual elements of this description which, *per se*, have nothing to do with statistics.

(The above excerpt reprinted from *Physics Today* is taken from 'Albert Einstein: Philosopher-Scientist', Volume VII of the Library of Living Philosophers, Inc., Northwestern University, Evanston, Illinois, who hold the copyright, with the kind permission of Paul Arthur Schilpp, editor of the series. Agents in Great Britain: Cambridge University Press.

THE EARLY HISTORY OF MAN

L. S. B. LEAKEY

UNTIL relatively recently most scientists who were interested in the origin of man looked to Central Asia as the place most likely to provide evidence of the 'cradle of the human race'. This was not surprising, as Asia had already yielded a good deal of evidence in support of such a theory.

In northern India sites in the Siwalik Hills had provided a number of fragmentary remains of fossil apes of several genera, including *Sivapithecus*, *Bramapithecus*, *Ramapithecus*, *Dryopithecus* and *Surgivapithecus*. These fossil ape remains came from deposits variably dated as either Pliocene or Upper Miocene.*

Then, too, it was in Asia – in Java and at Choukoutien near Peking – that remains of what were commonly described as the 'most primitive' types of man, *Pithecanthropus* and *Sinanthropus*,* had been found, and there were many who were prepared to agree with the late Professor Weidenreich in seeing in these types the forms from which true man later evolved.

There were always some scientists, however, who did not subscribe to the theory that Asia was the cradle of humanity, pointing out that the earliest known remains of the true ape stock came from the African continent – from the Oligocene deposits of the Fayum in Egypt – and also emphasising that of the still-surviving great apes the two with most points of resemblance to man were the two African genera, the Gorilla and the Chimpanzee. Professor Dart's discovery of *Australopithecus* had, moreover, strengthened the case of those who looked to Africa for evidence of man's origin.

To-day there are probably few left who would seriously give Central Asia pride of place as the region most likely to be the

* See Glossary for a table of these geological periods.

'cradle of the human race', for recent discoveries in South and Central Africa have consistently altered the whole picture.

The very important new finds of fossil apes made by Broom and Dart are not really our concern in this article, which deals primarily with East Africa, but they must be referred to briefly.

A very large number of skulls and jaws, as well as some limb bones of several genera of fossil apes, have now been found in the Transvaal by these two scientists. All the evidence from the fossils points to the conclusion that these creatures walked almost erect and in many ways come nearer to the conception of a 'missing link' between man and ape than anything found hitherto. In certain respects, however, these South African apes were already rather specialized and they are also of a relatively recent geological date (some of them Pleistocene and others possibly Upper Pliocene), so that they cannot be regarded as directly ancestral to true man. Their bearing upon the question of whether Africa was the place where man was evolved is, however, very great, for if they are not true ancestors they stand very close indeed to the stage from which man probably did evolve.

The other series of discoveries that have turned opinion towards Africa as the place where man and ape were both evolved are those of the many forms of fossil apes found in Kenya. Very rich fossiliferous deposits are now known in that country, at Rusinga Island, Songhor, Koru, Maboko, Ombo and a number of other sites, and the quantity of fossil ape remains that have come from these deposits in the last few years is almost unbelievable.

More than three hundred specimens representing fossil apes have now been found in Kenya in the Lower Miocene deposits, and a detailed report of the discoveries is now in the press.

These specimens show that at this early date in East Africa there was not only a very large ape population, but that it was one which included an astonishing variety of genera and species living side by side and ranging in size from creatures as large as a gorilla to little apes about the size of a gibbon.

The mere fact that such a large variety of fossil apes had been

found in deposits which are geologically considerably older than those of the Siwalik Hills would in itself tend to point to Africa as the potential place where the ape and human stock was evolved, but there is other evidence in the fossils themselves which is also of great significance.

Hitherto there has been a marked tendency to regard many of the characters which are found in the modern great apes as 'primitive' – characters that had been retained unchanged from an earlier ancestral stage – while man was supposed to have gradually lost these characters during the process of his gradual evolution away from the ape.

Following this line of thought, the presence of the so-called 'primitive' ape characters in certain human fossil types was regarded as an indication that the fossil men in which these characters were found were 'primitive', and were possibly to be regarded as ancestral to modern man.

For example, the skulls of the modern great apes are characterized by the presence of massive bony brow-ridges over the eyes and by a ledge of bone on the inner aspect of the middle line of the jaw which is known as the simian shelf.

The presence of massive brow-ridges in Peking man, Java man and Neanderthal man, thus came to be regarded as evidence of their 'primitive' status, while the simian shelf on the mandible of Piltown man was similarly interpreted.

Among the many fossil ape remains from the Lower Miocene deposits in Kenya are numerous jaws of several different genera and species in which the middle line – or symphyseal area – is well preserved, and examination shows that contrary to expectation none of them have any sign of a simian shelf. We can therefore no longer regard the simian shelf as a primitive character retained in the modern apes and in Piltown man from a very early and primitive stage of evolution of the ape-human stock, but we must instead regard it as a highly specialised character of the jaw which developed only in certain branches, just as the chin developed in the ultimate human branch.

The symphyseal region of the Lower Miocene apes has neither a simian shelf nor a chin, but has instead a form which could by

gradual development evolve into either, and this form may be regarded as the truly primitive form of this region of the jaw.

Skulls of Miocene and Lower Pliocene apes have hitherto been unknown and were therefore the subject of much speculation, but in October 1948 a fairly complete skull (Inset, Nos. 15–16) of an adult *Proconsul africanus* was discovered by my wife while working with me on Rusinga Island. This ape skull may be regarded as truly primitive, and indeed it shows a number of characters more reminiscent of the monkeys than of the apes.

One of the most remarkable and unexpected things about this skull is that it shows absolutely no trace of the supposedly 'primitive' ape character, the heavy brow-ridge of bone over the orbits,* a character which for years has been regarded as evidence of the primitiveness of the modern apes and of such fossil human skulls as also show it. The evidence from this single specimen is not, of course, conclusive and the search continues for further ape skulls in the Kenya Miocene deposits. But it certainly now seems likely that in the early and truly primitive members of the ape stock this character was not developed and that its presence in the living great apes and in human skulls like that of Peking man must be interpreted as showing a high degree of specialization, and not primitiveness in the true sense.

Then again, in all the living great apes we find that the arms are very long in relation to the legs and that this character is linked with a mode of progression which is usually spoken of as 'brachiating'. It has often been assumed in the past that man's immediate ancestors must have passed through such a stage and that they eventually changed from a brachiating to an upright gait, at the same time undergoing a reduction in the length of the arms.

Thus, long arms indicating a brachiating stage were regarded as a 'primitive' character and Neanderthal man, in whom the proportion of the arms to the legs tends in the direction seen in the great apes, was regarded as a very primitive human type.

* Hollows in the skull to hold the eyeballs.

A preliminary report on Olorgesailie is in course of preparation, but a great deal more work remains to be done.

One of the facts revealed at Olorgesailie is that, whatever the position may have been elsewhere, here Acheulean man was without the knowledge of fire. Another interesting point is the evidence that Acheulean man probably used the *bolas* as his chief hunting weapon, while handaxes and cleavers were probably more in the nature of domestic tools for chopping and skinning (Inset, No. 19).

Round stones suggestive of the *bolas* have long been known in prehistoric sites, but at Olorgesailie the finding of groups of three such stones on the floors of the living sites leaves little doubt as to their use.

So far we still know next to nothing about the men who were responsible for the Hand-axe culture. It might have been expected that sites such as Olduvai and Olorgesailie, with their vast quantities of tools and with excellent conditions for the fossilization of bone, would have yielded fossil human remains to solve this problem. But so far we have been disappointed.

It would seem that Hand-axe man (unlike the later cave-dwellers who made the Mousterian and Aurignacian cultures) did not bury his dead, but probably left the bodies to be devoured by hyaenas and vultures and other carrion feeders, much as some native tribes of Africa did until twenty or thirty years ago. Even if Acheulean man had been a cannibal we might well have found remains to give us a clue as to his racial type, for we have the fossil bones of a great variety of the animals upon which he fed, strewn over the living floors.

We are thus still not completely sure of the human type responsible for the Hand-axe culture, although the evidence of the Swanscombe skull from the high-level gravels of the Thames, and of the Kanjera skull fragments from Kenya, suggests that he was of *Homo sapiens* type.

The story of Hand-axe man in Europe is to be studied against a background of changing climate – from the advances of the ice-sheets during the glacial maxima to the periods when warmth-loving animals roamed England and France and the

hippopotamus sported in the waters of the Thames and the Cam.

In an exactly similar way the story in East Africa is seen and studied in relation to climatic changes which vary from very wet or pluvial periods to dry and semi-desert interpluvial stages.

There can be little doubt that the changes of climate which in Europe are represented by glacial and interglacial periods and in East Africa by pluvials and interpluvials were world changes and not merely local events. But there is still no general agreement among scientists as to whether pluvials were contemporary with glacial and interpluvials with interglacial periods; on the whole, however, the balance of opinion seems to favour this interpretation. There is a likelihood that this problem may be solved in the not too distant future, for there is evidence which can be used to link the climatic changes in Europe with variations in the relative land and sea level.

Similar changes in land and sea level can also be seen and studied along the African coast, and if the climatic changes here can be linked with these changes in sea level, a solution to the problem may be found.

We already know that along the Kenya coast the sea level has on at least two occasions during stone-age times been high enough to inundate old land surfaces upon which prehistoric tools occur. The first of these rises laid down marine sands and even coral over the surface on which were lying tools of the Hand-axe culture.

Turning back to Olorgesailie, we have, in addition to the sites with hand-axes, two sites with assemblages of implements of a much more primitive character. This second culture is now in process of being studied and it seems that it may have affinities with the earlier stages of the Clacton culture of Europe.

When the Hand-axe culture was past its zenith in East Africa two quite distinct cultures, each of which has its roots in the Chelles-Acheul culture, appear upon the scene. These are the cultures known as the Fauresmith and the Sangoan. The first is confined to the east-central regions of Kenya, but probably extends southwards through Tanganyika, since it is also found in

South Africa. The second, the Sangoan, is found in western Kenya, all through Uganda, and also has a very wide distribution in the Congo and south-west to Angola.

Until recently, much of what is now classified as Sangoan was referred to as Tumbian, and a recent paper using this term described the later stages of the culture in western Kenya. As a result of the first Pan-African Congress in Prehistory, held in Kenya in 1947, the term Tumbian was abandoned in favour of Sangoan. This culture is of great interest, for it seems that it came into being somewhere in the well-watered regions of the south-west Congo and northern Angola at a time when climatic conditions over much of East and South Africa were very dry, probably too dry to support man and the game animals he needed for his food.

A great deal more work on this culture is now required in order to get a clearer picture of the relationship of the cultural stages which occur in west Kenya and Uganda to the evolution of the culture in the Congo-Angola area.

Like the Sangoan, the Fauresmith (formerly called Nanyukian in Kenya) reached its greatest development during the dry period at the end of the Middle Pleistocene, and it is mainly found in the high country bordering mountain regions in Kenya, where, in the dry periods, there was more rainfall than over the rest of the country.

Of these two cultures which followed the Chelles-Acheul, the Sangoan in the western zone continued to evolve for a very long time, passing through a number of stages during the Upper Pleistocene, even into neolithic times.

On the other hand, the Fauresmith – in East Africa at least – was short-lived, though there is evidence that in South Africa it was the stock from which some of the branches of the so-called Middle Stone Age complex were derived.

But we are not concerned here with the details of the South African story. In East Africa, in the zone where the Fauresmith had flourished for a time, this culture was replaced during the early Upper Pleistocene by two distinct but contemporary cultures; a developed stage of the true Levalloisian, and a

culture akin to the Lower Aurignacian of Europe. This is now called the Kenya Capsian (Inset, No. 20), although in all earlier works it was referred to as the Kenya Aurignacian.

The Kenya Capsian folk were essentially dwellers in rock shelters, like their European counterparts, and they made and used a great diversity of tool types in stone, each adapted for a special purpose. The commonest tools were blunt-backed knife blades, scrapers, chisels and small crescent-shaped tools which were used as barbs for arrows. They buried their dead – in an ultra-contracted position – in the floors of the rock shelters in which they lived, and we have found enough skeletons to have a good idea of what they were like. They were a well-built people of *Homo sapiens* type, not unlike the Combe Capelle man, who existed in France at the same period. They had already begun to adorn themselves with beads and pendants and to use red ochre and, of course, fire was known to them.

There are a number of rock shelters in Tanganyika where there are well-preserved prehistoric paintings, and it is likely, but as yet not proved, that the earliest phases of this art may have been the work of the Kenya Capsian people.

The makers of the contemporary Developed Levalloisian, unlike the Kenya Capsian folk, were a people who lived in the open near streams or lakes, much as the earlier Hand-axe people used to do. Their culture as found in East Africa very closely resembles the classical European Levalloisian of France. It has very few specialised tools, but the flakes from which the tools were made were obtained by a very specialised technique, known as the *tortoise core* technique. While this culture has been known in Kenya for many years, the best and most prolific site is one found by the late Archdeacon Owen and studied during the war years. It is at Muguruk, near Lake Victoria.

In Kenya no human remains have been found associated with the Levalloisian, but near Lake Eyasi, in Tanganyika, Dr Kohl-Larsen found parts of more than one skull and the report on the most complete skull, published by Dr Weinert of Kiel, shows that it was a man of brutish features, akin in many respects to Peking man, Java man and Neanderthal man and also

with some resemblances to Rhodesian man. Dr Weinert created a new genus for this human type, calling it *Africanthropus*, and he suggested that it might be of very early date. Recently, however, Dr Reeve, a geologist, and I, have published a report showing that the Eyasi skull is of Upper Pleistocene age and was contemporary with the *Homo sapiens* makers of the Kenya Capsian culture.

The Eyasi skull (Inset 21) (which we cannot refer to by Weinert's name of *Africanthropus*, since that name is invalid, having been used previously in a quite different connection), must be regarded as representing a very over-specialised branch of the human stock which survived until Upper Pleistocene times in Africa, in the same way that Neanderthal man survived in Europe, and eventually became extinct.

Towards the close of the Upper Pleistocene period the Levalloisian was replaced by a culture known as the Stillbay, which contains elements of both the preceding cultures. There is some reason to believe – but it has yet to be conclusively proved – that the makers of this culture may have been responsible for some of the finest naturalistic paintings in the prehistoric art of Tanganyika.

So far in East Africa we have no evidence of the physical appearance of the Stillbay people, who lived not only in rock-shelters, but also sometimes in the open, thus combining the living habits of the Levalloisian and Kenya Capsian peoples.

The later stages of the Kenya Capsian and of the Stillbay cultures, which were contemporary, both have derivative cultures which extend into mesolithic and neolithic times. A number of new sites of these derived cultures have been found and some of them studied during the last few years.

Both the Magosian, and the branch of the Wilton culture known as Kenya Wilton B, may be regarded as having their roots in the Stillbay, but both have been considerably influenced by the contemporary derivatives of the Upper Kenya Capsian culture.

The Magosian, which precedes the Wilton B, is known from both rock shelters and open station sites, and it is usually found

in places where tectonic springs assured a water supply and attendant animal and bird life even when the climate generally was very dry.

The Wilton B culture contains many degenerate Magosian elements, and it is found only in rock shelters, presumably because the climate was then much wetter and made the occupation of open sites undesirable.

It seems almost certain that these Wilton B people were closely akin to the bushmen, and probably the later stages of the Tanganyika stone-age art were the work of the makers of this culture.

The Mesolithic Elmenteitan culture, which was discovered and studied in 1928–29, seems to have been very localised, and very few new sites of this cultural stage have been found during the subsequent years. Its origin is still not clear, though it has certain affinities with the Kenya Capsian and may be an aberrant offshoot.

Of neolithic cultures which have been studied in recent years, the oldest is probably the Hyrax Hill variant of the Stone Bowl Culture, which was published during the war. This culture is a late derivative of the Kenya Capsian, and an obsidian industry very similar to that of the Capsian is found in association with flat, platter-like stone bowls. The makers of the culture were a people with a well-organised community life, as evidenced by their large communal cemeteries in which blocks of stone – requiring eight to ten men to lift – were used in building up the mounds over the bodies. The pottery associated with the Hyrax Hill neolithic culture is characterised by a beaker of a type not known in any other stone-age culture in East Africa.

The Hyrax Hill site, which was excavated and fully described by my wife, is not far from the Nakuru burial site belonging to the Gumban B variant of the Stone Bowl Culture, excavated in 1926. The contrast between the Hyrax Hill and Gumban B Neolithic cultures is very great, the latter having deep stone bowls and burying their dead in an unusual manner against a cliff face; a second burial of this type was found during 1949, within a mile of the site excavated in 1926. At this site, as at the

Nakuru burial site, the sloping mound of stones built up against the cliff face over the grave contained the remains of a number of bodies that had been thrown in as the mound was constructed. It seems possible that these may have been slaves.

While burials of the Gumban B are known at the Nakuru burial site and at this new site, the work carried out at Hyrax Hill revealed a village of pit-dwellings that were also of the Gumban B culture, later in date than the Hyrax Hill variant of the Stone Bowl culture, but earlier than the Iron age burial pits and stone circles that also occur at the same site.

About three miles from Hyrax Hill there is too, an immense and at present unexcavated settlement of the Gumban B culture which includes two earthwork 'forts' enclosed by large ditches as well as many pit-dwellings.

An interesting parallel between the Gumban B people and a tribe still living in Tanganyika, called the Iroqui, seems to exist, for this tribe, even at the present day, inhabits underground huts or pit-dwellings of a very similar structure.

Yet another variant of the neolithic Stone Bowl culture is that from the Njoro River cave. This is probably a very late branch of the neolithic, but in the obsidian industry it exhibits clear links with the mesolithic Elmenteitan, while some of the stone bowl types recall those from Hyrax Hill and others those of the Gumban B. The Njoro cave variant is remarkable because of the very unusual method of cremation which was practised. Bodies were first tightly bound into an ultra-contracted position and then placed in a shallow grave in the floor of the cave and covered over with a little soil. A large fire was then lit over the grave and the body, together with the grave goods, was 'baked'. In this way perishable objects like wooden vessels, basket work, calabashes, etc., were burned to charcoal by a process of closed combustion, and were thus preserved.

The Njoro river site is also of interest because of the presence of hundreds of well-made beads of such materials as agate, carnelian, chalcedony and amazon stone - beads which strongly recall specimens from pre-dynastic Egypt, but which were undoubtedly made locally by these stone-age people who were

not negro or even negroid, as far as their skull characters show. It is to be hoped that a living-site of these people will be found before long, as it might throw light on their origin.

Polished axes have been known in Kenya for some time and during the past few years many more have come to light, but there is still no real evidence of what culture they are associated with, though it may be assumed that they are very late neolithic and that their use probably extended well into the Iron Age. Some forms, in fact, suggest that they may be stone copies of metal prototypes. A brief paper illustrating the principal types of polished axes known in Kenya was published during the war.

It is not easy in Kenya to differentiate between the end of the Neolithic period and the beginning of the Early Iron Age, any more than the dividing line between the Neolithic and Bronze Age is distinct in Europe.

The study of the Iron Age in Kenya is still relatively untouched, but a brief report on the Iron Age site overlying the neolithic cemetery was included in the report on the Hyrax Hill excavations, and there has also been published a report on some curious 'dimple-based' pottery of Iron Age date.

The Iron Age people of Hyrax Hill built rough stone enclosures, probably for cattle kraals, and small stone hut circles. A number of burial pits linked with the site contained decapitated bodies.

The 'dimple-based' pottery was first found by the late Archdeacon Owen at sites near Ngia, in Central Kavirondo, and more recently a number of other sites have been discovered on islands in Lake Victoria. This pottery includes a wide variety of bowls, beakers, etc., and is associated with evidence of intensive cultivation, and instead of the usual movable saddle-quern or pestle and mortar, corn was ground on the tops of huge granite boulders - in one case more than eighty grinding hollows were found on a single large rock slab (Inset, Nos. 22-23).

The story of prehistoric man in Kenya is beginning to take shape and although much detail remains to be filled in, a general picture of his gradual evolution can now be seen. As an introduction to the story we have the evidence of a major centre of

primate evolution in Lower Miocene times, and from the dawn of the Pleistocene onwards, successive stages of culture can be traced almost without a break until the Iron Age is reached.

BIBLIOGRAPHY

- Report on the Kenya Miocene Apes.* To be published by the British Museum of Natural History.
- Report on the Olduvai Cultures.* To be published by the Cambridge University Press.
- 'A Contribution to the Study of the Tumbian Culture in E. Africa' by L. S. B. Leakey and Archdeacon Owen. *Coryndon Memorial Museum Occasional Papers*, No. 1. 1945.
- 'Report on a Visit to the Site of the Eyasi Skull found by Dr Kohl-Larsen' by L. S. B. Leakey and W. H. Reeve. *Journal E. Africa Natural History Society*, 1946.
- 'Report on the Excavations at Hyrax Hill, Kenya Colony, 1937-8' by Mary D. Leakey. *Trans. Royal Society of S. Africa*, 1945.
- Excavations at the Njoro River Cave* by M. D. and L. S. B. Leakey, to be published by Oxford University Press, 1950.
- 'Notes on the Ground and Polished Stone Axes of E. Africa' by Mary D. Leakey. *Journal E. Africa Natural History Society*, 1943.
- 'Dimple-Based Pottery from Central Kavirondo, Kenya Colony' by M. D. Leakey, the late Archdeacon Owen and L. S. B. Leakey. *Coryndon Memorial Museum Occasional Papers*, No. 2, 1948.

FINDING WATER

SIR CYRIL S. FOX

ALTHOUGH long spells of sunny days are always enjoyable our experience is that, when these summer conditions of cloudless weather are prolonged, lawns become dry, streams cease to flow, springs and wells fail, and water restrictions may be imposed. This direct connection between lack of rain and drought is so obvious that it would be unnecessary to refer to it but for the fact that it is insufficiently appreciated. When a drought becomes severe, or when water has to be found in a desert region, it is an instinctive proceeding to turn to anyone reputed to be a water-diviner or 'dowser'. It is therefore interesting to know that Cicero described the *virgula divina* or divining-stick for discovering hidden articles, before the Christian era. Also that, in 1546, Georg Agricola (Bauer) wrote a treatise on prospecting and mining, and illustrated the use of the *virgula furcata*, or forked twig (of hazel or willow), in prospecting for hidden metalliferous veins in the Erzgebirge of Bohemia. Similar methods of prospecting for minerals were known in the adjacent mining districts of Saxony and the Harz. It was from this latter area, about 1575, that German miners were recruited in Queen Elizabeth's reign, to open up and work the mines in Cornwall in England. The German miners from the Harz brought with them their *schlagruthe*, or striking-rod (forked hazel twigs), and - so it is claimed - when the mines proved unprofitable, transferred the use of their divining methods to water finding.

There appears to be little doubt that there are people who possess the gift of detecting the presence of water underground, but, after many years' experience as a mining geologist, the writer can remember only three cases of this non-scientific method of finding water worthy of mention. The first instance was in the Deccan of India, near Manmad, where water was urgently

needed for an important railway station (for watering locomotives). A well-known professional water-diviner had indicated positions for wells, but the results were disappointing and the money was wasted. Geologically, there was little hope of getting the required quantities of water at the place (a watershed) from relatively shallow wells (in basaltic lavas). Arrangements were finally made to impound water in a near-by river, which, although dry in the hot months, carried floods in the rainy season.

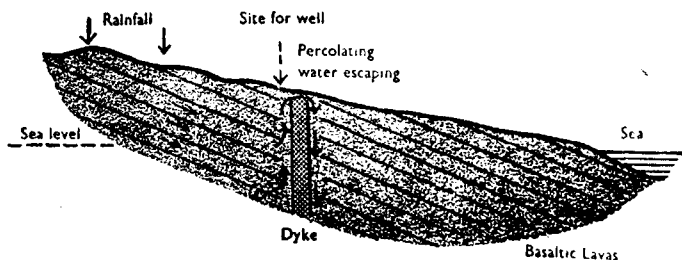


Fig. 2. Geological Section to show effect of a dyke on underground water (see also Inset, No. 24).

The next case was among the parched defiles in Waziristan, North-West Frontier of Pakistan, where the rainfall is small, but may be precipitated in a few hours and cause disastrous floods. The professional water-diviner indicated a place which was unsatisfactory as a source of water-supply and militarily unsuitable for the outpost (fort). The geological investigation led to the discovery of a water-bearing stratum into which a well was sunk and from which water could be drawn under protection from the fort.

The third example does not really involve a professional water-diviner, but is of considerable interest. In 1942, when the Japanese army was occupying Burma and their navy was raiding the Bay of Bengal, a British naval base was contemplated in the Gulf of Cutch, west of Jamnagar in Kathiawar. The writer was entrusted with the task of finding fresh water in great quantity. The region receives a moderate rainfall during June to October, but very little for the remaining eight months of the year, which

ends in three months of hot weather, March to May, when all the streams are dry and only very deep wells hold any supplies of water. The country is open, and underlaid by basaltic lavas (Deccan trap) which slope northwards to the sea. The inclination of the beds, or sheets, of lava (each roughly 50 feet thick) is slightly steeper than the slope of the land. The structure is the opposite to the tiles on a roof, and, consequently, rain falling on the exposed lava surface (see Fig. 2) tends to penetrate into the lava (underground) rather than to flow over the surface. Cutting vertically upwards through the lava sheets there was found one wall-like mass, or dyke, of rock (dolerite), which could be traced across country as a low ridge. As this dyke, filling a fissure up which the lavas had welled to the surface, would function as a 'hidden dam' and obstruct all water percolating underground towards the north, it seemed obvious to fix the site of a well in the bed of a large stream and immediately at the upstream side of where the dyke crossed (cf. Inset, No. 24).

I explained the position to the Engineer Officer with me, and we traced the dyke (ridge) to a village, and so to the big stream. The bed was dry, but, just at the point we had decided on for a large well, there were five villagers digging; they were standing in water barely 3 feet below the surface of the stream bed. On enquiry they told us they were making a well; that a very holy man (a Brahman priest who had stayed at the village on his journey to Dwarka) had fixed the position of the well two days previously. It appeared that, when leaving the village, the priest had said that he could not pay for the food and shelter given to him, but he would try to help the village people in any reasonable way. He was promptly informed that the village had suffered from shortage of water for generations. After a considerable time in contemplation and prayer, but without using a divining-rod or twig of any kind, the priest had gone down to the stream and indicated a place where they could dig a well for their water supplies for the hot months of the year. We left the village people free to enjoy the Brahman's gift. It was quite clear that the supply would not in any case be adequate for the projected naval base, and, as the dyke crossed another larger stream

to the east, it was proposed to erect a dam using the dyke as a core for the foundations. A small reservoir also existed in this larger stream, but, as a result of our success in Burma shortly afterwards, the project for a naval base was abandoned, and there was no need to incur the expense of a storage dam and gravity feed pipe-line.

RAINFALL AND PERCOLATION

Practically all the water met with on the land, in lakes, rivers and glaciers, is derived from rain and snow which has fallen on the surface. A part of this rain or snowfall is evaporated back into the air, a considerable portion flows away to replenish the rivers, lakes and icefields, and the remainder sinks into the ground to soak the strata and rocks, and maintain the flow of springs. It is difficult to ascertain the proportion of the rainfall which percolates into the ground, but it is evident that where the strata are porous, as in the case of coarse-grained sandstones, or the rocks are fissured, as in the case of fractured quartzites, very large amounts of water will go underground. In those cases where percolation occurs through the pores of the rock, there is some resistance to the flow of the water, if the beds are fine-grained; there may be little resistance if the strata are highly porous, and even less obstruction if the rocks are traversed by fissures. Granites are frequently traversed by joint planes which, so to say, cut up the mass into enormous blocks several yards along each side, if the jointing is regular. Basaltic rocks are also jointed, but the fissures are smaller and much closer together than in granites. Most rocks are cut by joints or fissures along which water can travel by percolation, if not by actual flow. In the case of Limestones the rock itself may be impervious, as is the case with granite, basalt, slate, etc.; but water dissolves limestone and so enlarges the fissures in these rocks into great underground caves and true subterranean rivers.

Unless the supply of water is maintained from rain, the movement of water underground, to saturate the rocks and to emerge as springs, will slow down and cease. In some instances the water percolates down to such great depths as to become heated

to boiling point. If we reckon a descent of 100 feet (30 metres) to increase the temperature 1°C ., then it requires a depth of less than 10,000 feet for any normal surface water to be heated to 100°C . The hot springs, which are so common in some areas and famous in most places, thus tell us of the depths to which they have been and come up again. As in the case of waters from springs, the point of emergence (the spring) is entirely different, and probably very distant from the area at which the water percolated into the ground. When Chicago was first settled on the shores of Lake Michigan and wells were sunk, the water came up under sufficient pressure to make a fountain 100 feet or so high. It was evident that this water came from the hills to the north, 150 miles away and perhaps 300 feet higher than Chicago, so that there must have been very easy percolation for a 'head' of 300 feet to force water 150 miles underground and then exert a pressure of nearly 50 lb. to the square inch at that distance. Owing to the numerous wells which have been put down since those early days the water has been drawn off and the pressure has been lost. It has been suggested that the water-holes in the oases in the Libyan Desert are supplied with water that percolates through the strata (Nubian Sandstones) from the Nile, at distances of upwards of 500 miles.

In the 'Artesian Basin' of Australia, in south-west Queensland, water has been found in borings over 2,000 feet deep, with a pressure, at the top of the well, sufficient to work a Pelton Wheel* (as at Thargomindah). The intake of the water for the artesian basin is considered to be from the rainfall on the Great Dividing Range, 500 miles away to the north-east. The waters tapped in south-west Queensland by deep borings are, as is to be expected, both warm and often brackish from the dissolved matter taken up during long percolation underground. It might be thought surprising that the water is as good as it is, if it has been 'standing' so long at such depths. The belief is that the basin goes to depths of 6,000 feet, but that the water

* A Pelton Wheel is an impulse water-turbine. It is a wheel furnished with specially constructed 'buckets' on the periphery. These are struck by a jet of water from a movable nozzle.

is in slow movement, and that the point of emergence is near Lake Eyre, South Australia, at the so-called Mound Springs. These 'mounds' are due to the matter deposited from the water, which, still hot, emerges after its great underground passage of perhaps 750 miles and to a depth of 6,000 feet. Explanations regarding great fresh-water springs in the sea – such as those of the Mediterranean at Spezia, Taranto, Syracuse, etc., in Italy; the gulf of Argos in Greece; the underground river from Postumia into the head of the Adriatic; etc. – are of the same, if not of an even simpler, type. In these examples the rainfall sinks into the limestone plateau (karstlands) through 'sink-holes' or dolinas, and gravitates through solution channels of considerable size into the limestone; and, if trapped in the limestone, the water escapes upwards at the first opening.

In contrast to the relatively easy percolation which appears to be in progress in the artesian basin of Australia, it is necessary to refer to a remarkable case of considerable resistance to the flow of water through sands. When Robert Stevenson, the famous engineer, was faced with the driving of a tunnel at Kilsby, near Rugby, through water-bearing strata of the nature of quicksand, he thought he would have to pump out all the water above the level of the projected tunnel. For this purpose he sank shafts and put in the pumps along the line of the tunnel. When the de-watering operation was in progress and observations were made he '... was soon surprised to find how slightly the depression of the water-level in one shaft influenced that of the other, notwithstanding a free communication existed between them through the sand, which was very coarse and open. It occurred to me that the resistance which the water encountered in its passage through the sand to the pumps would be accurately measured by the slope or inclination which the surface of the water assumed towards the pumps, and that it would be unnecessary to draw the whole of the water off the quicksand, but to persevere in pumping only in the precise level of the tunnel, allowing the surface of the water flowing through the sand to assume that inclination which was due to resistance. The simple result, therefore, of all the pumping was merely to establish and

maintain a channel of comparatively dry sand in the immediate line of the intended tunnel, leaving waters heaped up on each side by the resistance which the sand offered to its descent to that line on which the pumps and shaft were situated.'

GEOLOGICAL CONSIDERATIONS

The scale on which the geologist operates is large by comparison with the problems for which a 'dowsers' may be called in, but, nevertheless, even in cases of village wells or factory borings, the geologist is obliged to give an explanation as to why he prefers one site to another. In Fig. 2 an obstruction to underground percolation, by an intrusive igneous rock (a dolerite dyke (Inset, No. 24)), forced water practically to the surface at the upstream side of the dyke. On the downstream side, barely 40 yards away, it is almost certain that a well would be dry down to a depth of 40 feet. Fig. 3 shows another section of the rocks

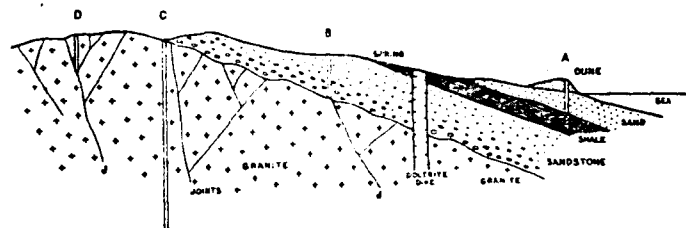


Fig. 3. Geological Section showing coastal problems of water supply. (From *The Geology of Water Supply* (Fox), by kind permission of the Technical Press Limited.)

near a coast where the problem of securing fresh-water may be considered. On the shore at A there is a sand dune deposited on the sand of the shore. Two effects are produced: (i) the dune will hold water (after rain) which may be tapped from the edge of the dune near the sea; (ii) the weight of the dune will compress the main body of sand below it and so make it hold up any percolating water in the sands and also keep back sea water. A boring through the dune should thus find small supplies of fresh

water, both in the dune and in the sand below the dune (even though these may be below sea level). The boring shown at B is of course intended to tap the water held up by the dyke in the gravels (which water seeps at the spring when these sands and gravels are overflowing). The borings at C and D, respectively, show an unsuccessful well in granite and a successful well in the same rocks. The first has unfortunately missed water-holding fissures.

A somewhat more complicated section, across a silted river valley, is shown in Fig. 4. At the point A a normal well has been

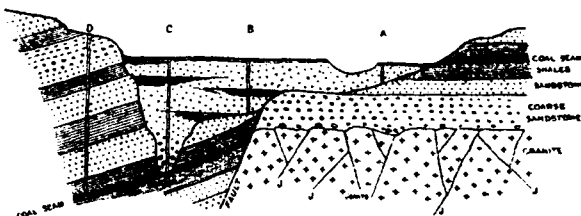


Fig. 4. Geological Section across silted-up river valley. (From *The Geology of Water Supply* (Fox), by kind permission of the Technical Press Limited.)

sunk in the river bed in the concave bend of the stream, probably successfully if the well is 10 or 15 feet deep. The well at B is shown into deeper sands which invariably carry percolating water even when the river bed is quite dry and dusty. The well at C in this section has probed into a filled-up gorge and encountered pebbles and boulders and almost certainly easy-flowing water in relative abundance. The only danger in this case will be that, when the river comes down in flood, it will be necessary to plug well C and abandon it until it is required again in the dry period. As an alternative to closing down C the illustration shows a shaft sunk at D on the bank, and from a known stratum (ascertained by the geologist's mapping) a gallery is driven up under water-bearing gravels and the water drawn into the pumping shaft D. A geologist studying these sections (Figs. 3 and 4) will no doubt recognise other alternatives in each, but he will also

know that, while there may be some similarity in different cases, it is seldom that one problem of water supply is exactly like another. As a general rule, each problem requires independent judgment, and this can be properly exercised only by a clear understanding of the movement of the underground water and the geological structure which must influence the flow of water underground.

Perhaps a few words may help to explain how a geological map is prepared and exactly how it may be used in an investigation for water. Fig. 5 is a map showing the position of three



Fig. 5. Skeleton geological map showing field data (see fig. 6 for key). (Redrawn from *Geikie's Structural and Field Geology* by courtesy of Oliver and Boyd Ltd.)

streams with their tributaries, and the observations made by a geologist who has carefully traversed each stream and marked the position of the rocks met with. These observations include the tilt of the strata, the direction of dykes and the line of dislocation of the strata. With the detailed observations (as full as can be made in the field) the geologist sketches in the rock formations and completes the map shown in Fig. 6; and, to elucidate the underground structure he draws a section across the map (see the line A to B and the section immediately under the map in Fig. 6). From such a geological map it is seen that the high ground lies to the north (top) so that run-off rainfall

and percolation will gravitate southwards. The conglomerates seen exposed to the left and at the head of most of the streams will take in large amounts of water, and this formation will be a good water-bearing stratum below ground. Indeed, a

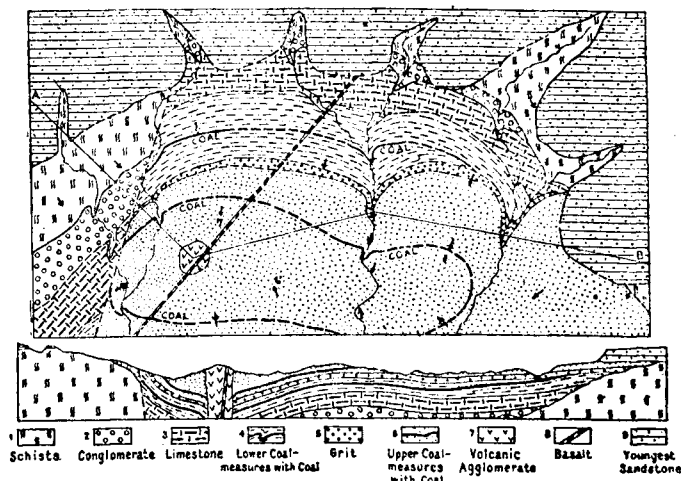


Fig. 6. Completed geological map with boundaries interpolated. (Redrawn from Geikie's Structural and Field Geology by courtesy of Oliver and Boyd Ltd.)

basin structure is present in the lower part of the map (outlined by the outcrop of a coal seam). Within this basin, and near the basalt dyke (where it cuts the volcanic agglomerate), a boring should strike artesian water. If the line of dislocation near the left of the map is an open fissure, wells put down where streams cross the dislocation (or fault) should, if they are sited in the conglomerate or the limestone, prove excellent.

An essential factor in all problems of finding water is the personal investigation. Looking at the map of rock formations (Fig. 6) and noting the succession of strata from the basement schists to the topmost sandstones, there is a danger of assuming that a very large proportion of any rainfall must sink into such a

widespread series of porous rocks. A field-visit over the area might show that the sandstones are fine-grained, that the grits include much interstitial feldspathic material, and that even the conglomeratic bed has clay filling in the spaces between the pebbles. Indeed, the personal inspection may show that the limestones absorb, through sink-holes and underground drainage, the percolating rain-water to a greater degree than any of the other formations. Also, that the fissure (on the line of faulting or dislocation) plane at the left of the map is sealed up with clay material, and therefore is not a plane for movement of underground water. A very careful examination of the map, followed by a more careful study of the behaviour of the outcrop of the coal seam and of the exposures in the top right hand area of the map, may prove that the successive layers are not absolutely parallel (not a conformable series). The outcrops as shown on the map also reveal a plane of discordance at the base of the grit bed, where it crosses the outcrop of the coal seam (in the right half of Fig. 6). There is every likelihood of such a plane being a zone of easy flow for underground water, so that a well sunk in the stream on the right, at a point nearest to the coal seam, and carried down to the base of the grit bed, should prove a good source of water. If still larger quantities are desired the well may be continued down, as a boring perhaps, into the underlying limestone.

SEASONAL INFLUENCES

Since it is recognised as a fundamental consideration in practically every example of water supply (although not always as evident for underground as for surface supplies) that rainfall is the source of the water, variations are to be expected with the season of the year. No better example of seasonal influence could be given than that described by Sir Francis de Guingand (*Operation Victory*, 1946, page 212), after the battle of El Alamein (1942). He wrote '... By the evening of November 6th, 1st Armoured Division was nearing this locality (Mersa Matruh), and it looked as if Rommel's fate was sealed. As luck would have it one of those very heavy rain storms occurred which for a short

intimate local knowledge is required for sinking such wells, which may be 150 to over 200 feet deep, and exceedingly costly. In the case of Bikaner city, where the wells are 200 to 300 feet deep in jointed sandstones, it may be soon demonstrated that the water drawn is more than the rainfall on the catchment can provide. Thus the underground water tapped by the city wells in Bikaner has percolated from a more distant region, which is believed to be 250 to 300 miles to the north-east, in the Jumna drainage area.

In the dry, stony stream beds which drain the karstland of Dhofar, in Southern Arabia (Figs. 7 and 8) there are abundant signs of great floods having been discharged northwards into the Rubi al Khali or 'Empty Quarter' of South-East Arabia. I have evidence of flood waters which must have been over 100 yards across, between 3 to 4 deep and flowing at roughly 1.5 feet per second (say 1,500 cubic feet per second). In a width of 60 miles at least half a dozen such streams with flood signs occur, so that, reckoning floods of only 12 hours duration, the total discharge carried into the desert and lost in the Sands (Ar Rimal) is of the order of 9,000 cubic feet per second (cusecs), or a total of nearly 400,000,000 cubic feet. The Dhofar karstland, known as the Jabal Qara, is a limestone watershed, to the south of which lies the 10 mile wide plain of Salala (Dhofar) bordering the Arabian Sea. Along the base of Jabal Qara at the edge of the plain there are several well known springs. One of these, the Ar' Zat in the Wadi Rizat, discharges fresh water at 18 cubic feet per second during the dry months (February to June), and larger amounts during the wet season (June to October) and immediately after (November to January). This water does not reach the sea as stream flow except during the rains. Nevertheless, a small lagoon at El Baleed, in which the water level rises and falls with the tide, and which is separated from the sea by 100 yards of beach, contains fresh water. There is similar evidence along this coastal strip near Salala. Indeed, the wells used for watering the coconut palms west of Salala draw fresh-water from wells which must be below sea level. The well at the Salala Landing Ground goes well below sea level, but is 3 miles from

THE SHOWER CLOUD (see pp. 67-77)

The development of cumulus during the day



1, 2. Illustrations show the growth of the cloud during the forenoon (Its appearance after midday is shown in illustration 3).



3. By about two hours after midday the cloud has reached the cumulonimbus stage.



4. Cumulus in lanes. 'Cloud streets' of the sailplane pilot (Glen Feshie, Inverness).



5. The anvil of the cumulonimbus.



6. Evening after a showery summer day. The cumulus clouds settle down and slowly dissipate (Castle Stalker, Appin, Argyll).

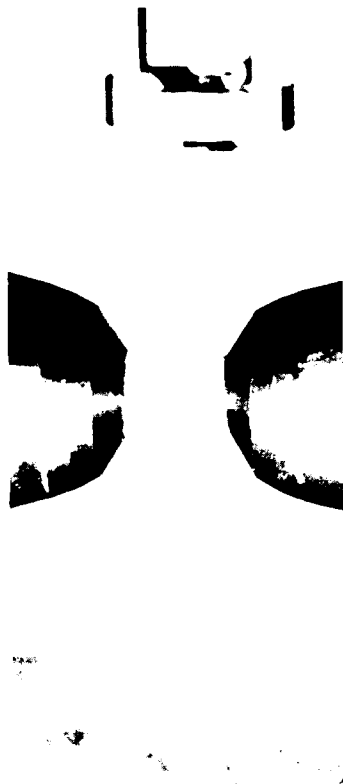
(Illustrations 1-5 are from Kodachrome transparencies by John F. Steljes.)

NEAR THE ABSOLUTE ZERO

The magnetic properties of liquid and solid oxygen, demonstrated in a powerful magnetic field

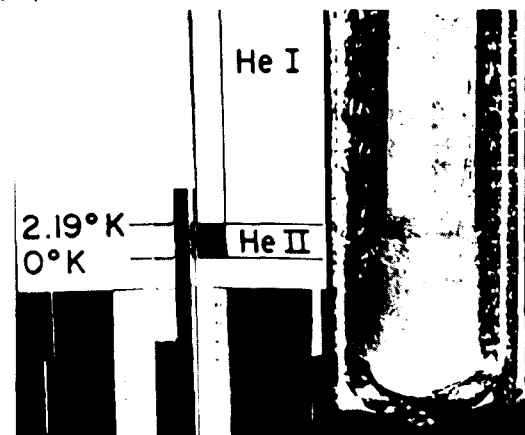


7. A vacuum (Thermos) flask of liquid oxygen is held between the poles of the magnet: the liquid surface becomes sharply U-shaped as the liquid is attracted.

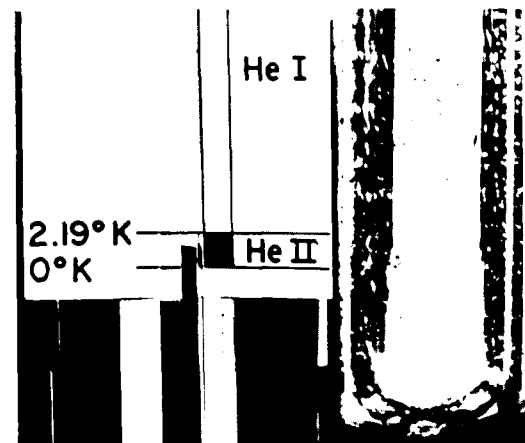


8. The liquid oxygen emptied between the magnetic poles has frozen solid – further chilled by the rapid evaporation of some of it. The solid oxygen adheres to the poles.

When cooled down to 2.19 degrees absolute (about -271°C) ordinary liquid helium (helium I) is transformed into helium II, a substance with some very peculiar properties.



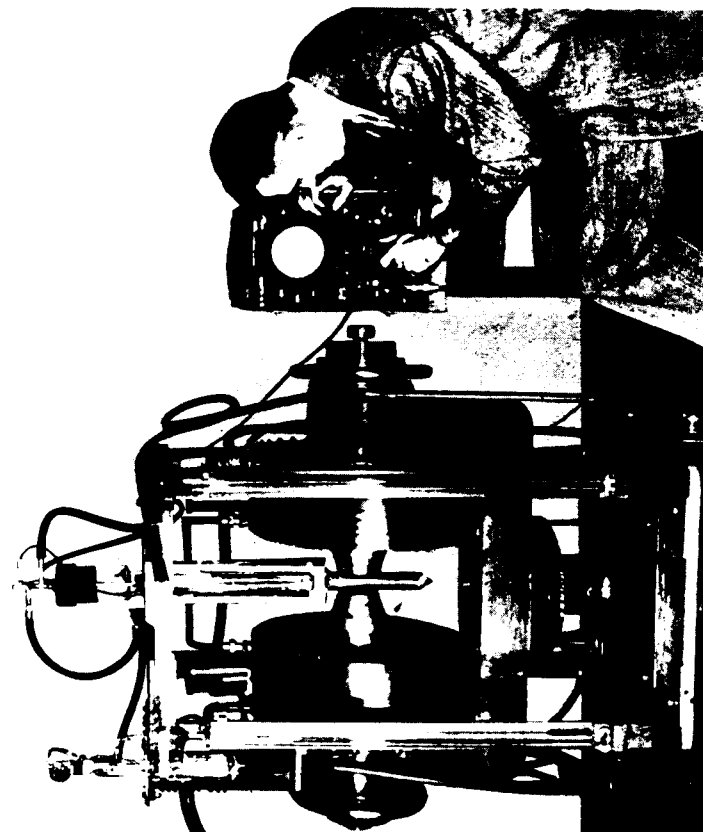
9. Helium I, lower right, supported in a vacuum flask within a second outer flash of liquid air, is boiling vigorously under reduced pressure. The manometer on the left indicates the temperature.



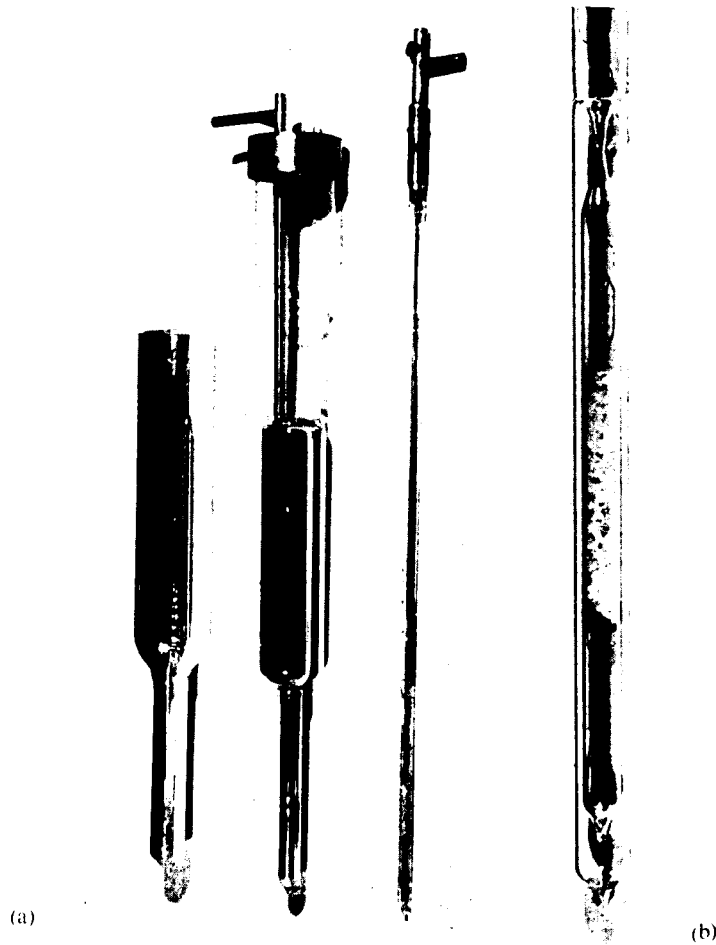
10. Further evacuation of the inner flask has cooled the helium below 2.19°K . Boiling suddenly stops, and the helium II remains still and clear, because it is a superconductor of heat, which spreads through it as a wave disturbance of measurable velocity, known as 'Second Sound' (see page 125).



11. The fountain effect in liquid helium II (see page 123). Just left of centre a J-shaped glass tube, somewhat larger in diameter at the bottom, has been suspended by means of a central rod in a quantity of liquid helium II surrounded by liquid air in an outer flask. The glass tube projects out of the helium and is approximately level with the liquid air surface, which is seen, disturbed by boiling, as a wavy horizontal white line running round in an ellipse, since it is photographed obliquely from above. The lower wider part of the glass tube contains emery powder, and when the liquid helium here is warmed by exposure to light it spurts through the emery powder and out of the top of the tube in a spectacular fountain (seen just to the left of the central rod, near the top).



12. Measuring the velocity of second sound in helium II in the low temperature laboratory of the U.S. National Bureau of Standards. Cooling below 1° K is obtained by adiabatic demagnetization: crystals of iron ammonium alum are immersed in liquid helium II within the innermost of three coaxial vacuum flasks (see illustration 13 and Fig. 27, page 124) supported between the poles of a large water-cooled electromagnet; when the magnetic field is suddenly switched off, the salt crystals are demagnetized and pick up random vibrational energy from the helium atoms. Pulses of electrically-generated heat pass within the helium II and are detected by a temperature-sensitive resistor as a small voltage pulse. This is amplified and displayed on the oscilloscope screen on the right.



13. (a) The three vacuum flasks for measuring the velocity of second sound (see also illustration 12). The outermost (*extreme left*) contains liquid air, the middle one liquid helium transferred directly from the liquefier through the pipe at the top, the innermost (*right*) the apparatus, seen enlarged in (b). The inside chamber of this is sealed off and suspended in an evacuated tube. It contains liquid helium II, the crystals, and near the bottom, two black discs about two inches apart - the transmitter and detector of the heat pulses (see also Fig. 27, page 124).

(Illustrations 7-13 supplied by the U.S. National Bureau of Standards.)



14. A skeleton of a Lower Miocene rhinoceros in the excavations at Rusinga (Kenya).



15. Full face view of the *Proconsul* skull found on Rusinga Island in 1948. Note the rectangular eye socket (orbit).



16. Profile view of the *Proconsul* skull. Note the complete absence of brow ridges, and the contour of the front part of the lower jaw.



17. General view of the Olorogesallie prehistoric site. Hundreds of hand-axes lie on the surface in the foreground, having been worked out by erosion from the living site on the original land surface (the hard horizontal band in the mound on the right).



18. Close-up of the foreground, showing hand-axes.



19. Part of the excavation of an old living floor on land surface 7 at Olorogesallie, showing the arrangements of implements on the camp site.



20, 21. The upper Kenya Capsian skull from Naivasha, a typical *Homo sapiens* type. Contrast it with the contemporary Eyasi skull from Tanganyika (*below*), found in Gambian (upper Pleistocene) deposits with a developed Levalloisian culture.



22, 23. A huge granite boulder, on Lolue island, in the centre of a living site of the Iron Age 'Dimple-base' pottery people. The tops of such boulders (*below*) were used as communal corn-grinding places – note the many circular depressions.



24. Two dolerite dykes side by side across a large river. They run down thousands of feet below ground cutting through horizontal sandstone strata acting as a dam (Photo by Sir Cyril S. Fox.)

the shore and yields slightly brackish water. I was of the opinion that it had been sited without geological advice, possibly by someone who was reputed to be a water-finder. There is no difficulty in finding water on the Dhofar plain, but a geologist would easily have found fresh water.

At Aden the difficulty of obtaining water was met with in the days of the Queen of Sheba (if the 'tanks' at Aden were really built under the advice of King Solomon). These 'tanks' to hold run-off rain-water are themselves an admission of an ancient scarcity which has prevailed at Aden ever since. Even when Aden was known as Eudaemon Arabia, a port or mart of Arabia Felix, the Romans failed to put things right. This is not surprising in a volcanic area, where the average rainfall is under 2.5 inches annually, where for years the fall may be nil, and where, at several-year intervals, there may be a storm yielding 7 inches in a couple of hours. In 1870 when a geologist appears, for the first time, to have examined the Aden problem of water supply, F. R. Mallet found a well, the Bamian, which had been sunk to a depth of 185 feet (70 feet below sea level) and had yielded 4,000 gallons a day of slightly saline water. As there was likely to be contamination from sea water on the one hand, and percolating water through the volcanic material would dissolve matter on the other, especially with so poor a rainfall, it was suggested to search inland. After an examination of the country within 25 miles or so of Aden, Mallet reported that an abundance of good water could be secured near Mahilla, 20 miles from Aden. He pointed out that the high lands to the north, draining towards Lahej to the south, provided large supplies for the wells in the plains adjacent to Aden. The distance was evidently considered too far, and for the next 60 years Aden depended on the evaporation and condensation of sea water, with a regular expenditure on coal for the heating of the plant. In 1934 when I called at the Residency at Aden I walked on the lawn and saw the fountain turned on. It was water piped (under gravity) from near Lahej.

It has long been admitted that a Geological Survey Department should be maintained as a State institution, but this

should include somewhat more than routine surveys for the purpose of preparing geological maps. There can be no doubt that the dowsers and professional water finders cannot operate with any greater reliability than an amateur equipped with some type of geophysical apparatus and a smattering of geological information. Neither could be depended upon to find water, or at least to indicate in detail the expectation of finding water for the largest quantities that might be required. This is the essence of the work of a trained and experienced geologist, but the principles are so simple that almost any observant individual, who has made himself familiar with the different kinds of rocks and their mode of occurrence, should readily understand these geological considerations. Added to this knowledge, there is the recognition of the rainfall as the source of supply, and the movement of percolating water underground. The subject is primarily dependent on field observations: making or using a geological map; collection of rainfall data; gauging of streams and rivers; behaviour of wells after pumping; noting the nature of spring and well waters; and avoidance of any kind of 'blind' work.

THE SHOWER CLOUD

JAMES PATON

*Sometimes we see a cloud that's dragonish,
A vapour sometimes like a bear or lion,
A towered citadel, a pendant rock,
A forked mountain, a blue promontory,
With trees upon't that nod unto the world
And mock our eyes with air.*

SHAKESPEARE'S 'towered citadel' is undoubtedly the fully developed cumulus cloud and the 'forked mountain' the cumulonimbus, the shower cloud, into which the cumulus may later develop. These are the most striking of all clouds, partly because of their great vertical development and partly because of their isolation and solidity, which cause remarkable variations in light and shade, the side next the sun being a dazzling white while the base is grey, sometimes almost black. As a massive turreted cumulus changes to cumulonimbus, the hollows in the cauliflower-like top disappear as they fill with ice crystals and soon the whole summit, in glaciating, becomes filmy and tenuous as the characteristic anvil takes shape. The cumulonimbus is the centre of violent convection and copious condensation that produces heavy showers, often of hail, and sometimes a thunderstorm. Much work has been done in the last decade on the development and structure of this cloud and on its electrification. This article will be concerned with the development of the cloud and its precipitation.

Convection results from instability in the atmosphere. The atmosphere is said to be unstable when, at the vertical displacement of a 'bubble' of air, the bubble continues to move with acceleration upwards or downwards from its initial position. The temperature of an ascending bubble of air falls approximately $5\frac{1}{2}^{\circ}\text{F}$ for every 1,000 feet of ascent. The fall in temperature is caused by the expansion as the bubble rises through, and immediately assumes the pressure of, its surroundings. The conductivity of air is negligibly small, and assuming that there is

neither sensible mixing with the environment nor gain or loss of heat by radiation, the expansion is adiabatic, i.e. the bubble may be supposed to be enclosed by a perfectly flexible skin through which heat cannot pass, in or out. The rate of fall of temperature may then be readily calculated to be $5\frac{1}{2}^{\circ}\text{F}$ for 1,000 feet, and this is known as the dry adiabatic lapse rate, the word 'dry' signifying, not that the air is devoid of water vapour, but merely that no condensation occurs. So, if the *existing* lapse rate of temperature is greater than $5\frac{1}{2}^{\circ}\text{F}$ per 1,000 feet, then the rising bubble will find itself always surrounded by air that is cooler and denser than itself. When a super-adiabatic lapse rate exists in the atmosphere, it requires only some trigger action to cause an initial slight upward motion to develop into an uprushing torrent of air.

Instability, then, may be caused by the heating of the land by the sun (solar radiation does not heat the atmosphere to any extent during its passage to earth), or by the movement of cold air over a warmer underlying surface, land or sea. In each case, the heating of the atmosphere from below increases the rate at which temperature falls off with height in the layer next the earth. For instance, during the forenoon of a clear sunny day, as the ground heats up, instability develops first in a shallow layer near the surface, and as the day proceeds the depth of this unstable layer increases. Here and there over the land there will be columns of ascending air – the 'thermals' of the sailplane pilot – while outside these scattered columns the air slowly subsides to become heated by contact with the earth and later to feed into the base of a thermal. In very dry weather, the position of the base of the thermal may sometimes be seen by the dust that it stirs up.

The air in the ascending column may eventually be cooled to the point where it becomes saturated by its contained water vapour, which condenses to become visible as a cloud of water droplets. During the early part of the day the ascending columns do not reach to the condensation level – there is 'dry convection'. Indeed, if the air is very dry, the condensation level may be so high that even when the unstable layer attains its maximum thickness at about two hours after mid-day, the top is still below the condensation level. But usually on a fine summer day small

scattered clouds, all at about the same height, develop early. Each marks the summit of a thermal and is a cumulus cloud.

The further development of the cumulus depends on whether the temperature of the cloud air is greater or less than that of the surrounding air into which it has penetrated. If the cloud air is at the same temperature as, or is cooler than, its environment, there is no tendency for further ascent. The clouds remain small and indicate fair weather. This is the cumulus humilis, the humble cumulus! (Inset, No. 1).

But if the cloud-bearing air is warmer and so less dense than its environment, it continues to ascend. The continuous condensation that proceeds in the ascending air is accompanied by the release of latent heat. (Heat must be supplied to evaporate water. The reverse process of condensation releases this heat.) The heat reduces the rate of fall of temperature in the ascending cloud-bearing air and so makes it more likely that the cloud remains at a higher temperature than its environment. Further, if the temperature lapse rate in the environment remains in excess of that in the ascending cloud to a great height, the cloud develops rapidly upwards, indeed often so rapidly that it is possible to observe the growth of the turrets that boil out of its summit. This is the cumulus congestus (Inset, No. 2). So long as the cloud consists entirely of water droplets, its boundaries remain well defined and solid, but when ice crystals appear, the cap becomes diffuse and tenuous. This change often takes place suddenly and marks the transition from the cumulus into the cumulonimbus (Inset, No. 3). Precipitation begins to fall out of the cloud very soon after the glaciation occurs. The anvil that often forms the summit of the cloud is caused by the spreading out of the rising air on reaching a stable environment, i.e. on attaining the level where the ascending air becomes cooler and so denser than its surroundings (Inset, No. 5). In the absence of strong winds, the cloud may extend from a base at 1,500 feet to a summit at a height of five miles. In strong winds, the accompanying wind-shear with height hinders vertical development and the cloud forms in long rolls extending down-wind (Inset, No. 4).

In the late afternoon when the cooling ground causes the atmosphere to become stable, the clouds shrink and collapse, leaving evaporating remnants at different levels (Inset, No. 6).

PRECIPITATION

It was the observation that precipitation is suddenly released very soon after glaciation is seen to occur at the cloud summit (i.e. at the change from cumulus to cumulonimbus), that led Bergeron to propose his now well-known theory of rain formation.

Even when air ascends in the cumulus above the level where its temperature is 0°C , condensation to liquid droplets continues, so that the upper portion still consists of water droplets though their temperature is below the normal freezing point. The droplets are *supercooled*. This state of affairs may persist up to levels where the temperature is as low as -35°C .

The initial process of ice formation, the first appearance of ice in the cloud, is by no means understood. Laboratory experiments indicate the necessity of a nucleus, called a *sublimation nucleus* to distinguish it from the *condensation nucleus* that is required for the initiation of the change to liquid water. The nature and properties of the sublimation nucleus are being investigated by many experimenters. Space forbids the discussion of the present state of this work, but it is sufficient to say that at temperatures above about -32°C it seems that few sublimation nuclei are present or effective, but that below this temperature, there is either a sudden increase in their efficiency or another type of nucleus springs into activity.

So, a few ice crystals, perhaps about one to every 100,000 water droplets, are likely to be present in the cloud below the level where the temperature is -32°C . Above that level the number of ice crystals will increase substantially.

Let us consider what happens when supercooled water droplets and ice crystals exist together in the cloud. The equilibrium vapour pressure over ice is less than that over supercooled water at the same temperature. The existing vapour pressure then adjusts itself to a value between the two equilibrium pressures and the atmosphere will be unsaturated for the water and super-

saturated for the ice. So the water droplets continuously evaporate while the ice crystals rapidly grow to the size when they fall through the cloud, eventually to melt and arrive at the surface as rain. This process, first proposed by Bergeron, is now generally accepted as one of the most important ways in which rain may form.

It is interesting to examine the consequences of (a) a very small and then of (b) a large concentration of ice crystals among the supercooled droplets.

(a) When only a few ice crystals are present, each will grow very rapidly by extracting water vapour from the surrounding supersaturated air, which is kept replenished by evaporation from the abundant water droplets. As each ice particle falls, it collides with and captures the water droplets that come in its way, the supercooled water beginning to freeze immediately on collision. The rate of freezing is conditioned by the speed at which the liberated latent heat of fusion may be removed. At very low temperatures, this heat is quickly conducted to the surroundings. So, if the accretion of supercooled droplets is not too rapid, each droplet freezes on to the ice particle almost immediately after the collision. So quickly does the freezing take place that the droplet preserves its shape and adheres as a tiny spherical pellet of ice, the resultant deposit being white and brittle because of the air occluded among the ice pellets. (This sort of thing may occur at ground level in a supercooled fog, when there grows out to windward of projecting objects the white deposit known as rime.)

But when the number of water droplets swept up by the falling ice particle is great, then the cooling processes of conduction and evaporation to the surroundings cannot take place rapidly enough to remove the latent heat of fusion and so to allow rapid freezing. The water of the droplets then spreads round the ice particle, freezing only very slowly to give a deposit of clear ice. (This process occasionally takes place at ground level when supercooled raindrops spread over and freeze on the ground, coating everything with a glass-like deposit, glazed ice, that completely disorganises transport).

What has just been described is of course the genesis and growth of the hailstone. When a large hailstone is sectioned, alternate rings of white and clear ice are often revealed. (use a razor blade cooled to freezing point by placing it in a heap of hailstones, then cut a fresh hailstone, holding the cold blade in your gloved hand). This structure may be readily explained by the descent of the growing stone through cloud conditions successively favourable to the accretion of rime and glazed ice. Ring structure seldom occurs in winter hail, for then the main deposit is likely to be rime and the resulting stone is entirely white and soft.

It had been thought that the ring structure could be explained only by assuming that the stone falls in the region outside the main up-current and then is whisked swiftly upwards again when caught in the violent upsurge at the base of the cloud. The rime deposit was supposed to form at high levels and the clear ice nearer the base, so that the number of excursions through the cloud would be given by the number of white or clear rings, much as the age of a tree is got by counting the annual rings.

(b) If large numbers of ice crystals appear among the water droplets, then evaporation from the droplets and sublimation on to the crystals occur as in (a), but the rate of crystal growth is now much reduced, for each can get only its share of the limited supply of water available. It may happen that every droplet has completely evaporated, while the crystals are still quite small; indeed so small that they remain in situ as part of the cloud. The presence of crystals in large numbers has caused this portion of the cloud to become entirely transformed into an ice cloud. Presumably this is the cause of the rapid glaciation of the cloud summit that is often followed by the development of an anvil, consisting entirely of ice crystals (Inset, No. 5).

So the initial development of precipitation in the cloud apparently requires the presence of a moderate number of ice crystals relative to water droplets. A large concentration would merely transform the cloud from water to ice. There is experimental evidence that a rapidly-growing ice particle may shed small ice splinters, each of which will grow in the same manner as its parent. A chain reaction of this kind may obviously play an

important part in causing multiplication of the precipitation elements. The position within the cloud of the region where precipitation begins may be determined from the ground or an aeroplane, using centimetric radar. The first rain echo is usually obtained from middle levels of the cloud and its top proceeds to extend upwards to about the level where the temperature is -30°C . The rain is seen to fall out of the base of the cloud about ten minutes after the first echo is received, and thereafter the top of the echo descends.

THE LIFE CYCLE OF THE CLOUD

A very thorough investigation of the life cycle of the cumulonimbus cloud has recently been made in America by the cooperation of the U.S. Weather Bureau, Air Force, Navy, National Advisory Committee for Aeronautics and the University of Chicago. Simultaneously and independently similar results were obtained in a smaller English study. The primary purpose of the work was to gain information which would assist pilots in flight through thunderstorms. In 1946 in Florida and in 1947 in Ohio, a close network of ground stations equipped with recording anemometers, hygrothermographs, barographs and rain gauges was set up. Observers serviced the stations, travelling through the network in jeeps. Within the network were radiosonde stations with direction-finding equipment and outside it were radar wind-finding stations.

When a thunderstorm approached and passed over the area five aeroplanes flew simultaneously horizontally through the storm at heights of 5,000 ft., 10,000 ft., 15,000 ft., 20,000 ft., and 25,000 ft., recording the strengths and positions of up- and down-currents, temperature, precipitation and the cloud extent. Each aircraft could be identified and traced by radar during its passage through the storm, and instructions from the flight controller on the ground passed to the pilots.

Within the cumulonimbus, regions of vigorous turbulence from 3 to 7 miles in width and then smooth conditions extending over about a mile were successively encountered. Simultaneously the surface stations recorded heavy rain directly under the

turbulent regions. It was immediately evident that the cloud consists of several clearly-defined convective systems or cells, separated by quiet, though cloud-filled, air. Further investigation showed that each cell undergoes a definite life cycle, beginning with a cumulus cloud within which the air movement is upwards throughout, while there is horizontal convergence into it at all levels (Fig. 9). When the summit reaches the level where ice crystals appear, the latter grow quickly, as previously explained,

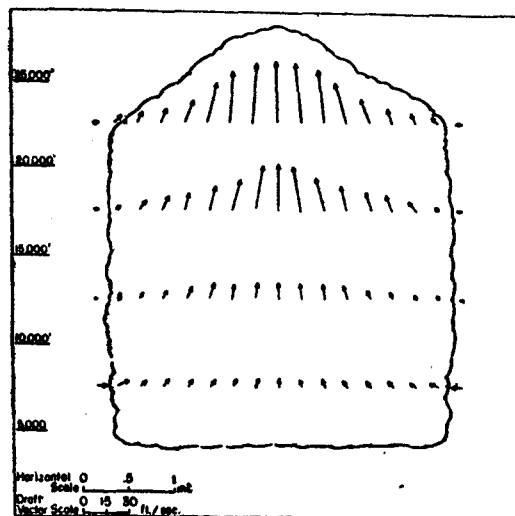


Fig. 9. Distribution of vertical currents in the cumulus stage. (By kind permission of the Editors of Weather)

and soon precipitation occurs through the weaker part of the up-current. The cloud has now passed into the *mature stage* of development (Fig. 10). The rain, though at first confined to quite a small region, is sufficiently heavy and concentrated to drag the surrounding air down with it, turning what has been an up-current into a down-current. This air brought down by the rain is kept cool during the descent by the continuous evaporation into

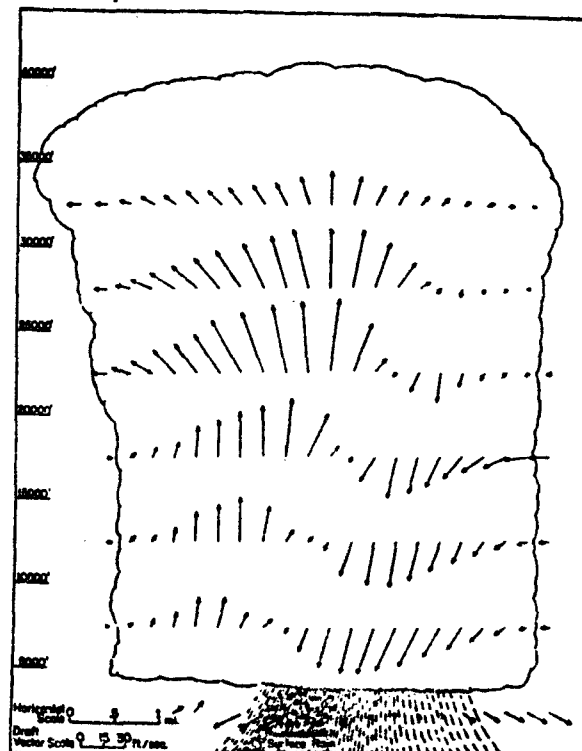


Fig. 10. The mature stage, showing rain and downdraft. (By kind permission of the Editors of Weather)

it from the raindrops, and is responsible for the well-known happenings in a thunderstorm – the sudden squall that accompanies the onset of the downpour of rain and the very rapid drop in temperature. Incidentally, it supplies considerable kinetic energy to the storm in addition to the more obvious source of energy release in the ascending damp air. The outspreading of the cold air over the ground is accurately traced both by the arrival of the squall and by the sudden drop in temperature recorded at the ground stations under and around the cell. Balloons show

that this outflow from the mature cell takes place up to heights varying between 2,000 ft. and 5,000 ft., convergence occurring above. It will be observed (Fig. 10) that vigorous ascent of air still continues outside the down-current in the lower part of the cloud and throughout the whole cloud top.

The rain becomes light as the down-currents, while losing their vigour, steadily extend to the remaining portion of the lower part of the cell, which now passes into the third and final phase of its cycle, the *dissipating stage* (Fig. 11). The air continues to ascend

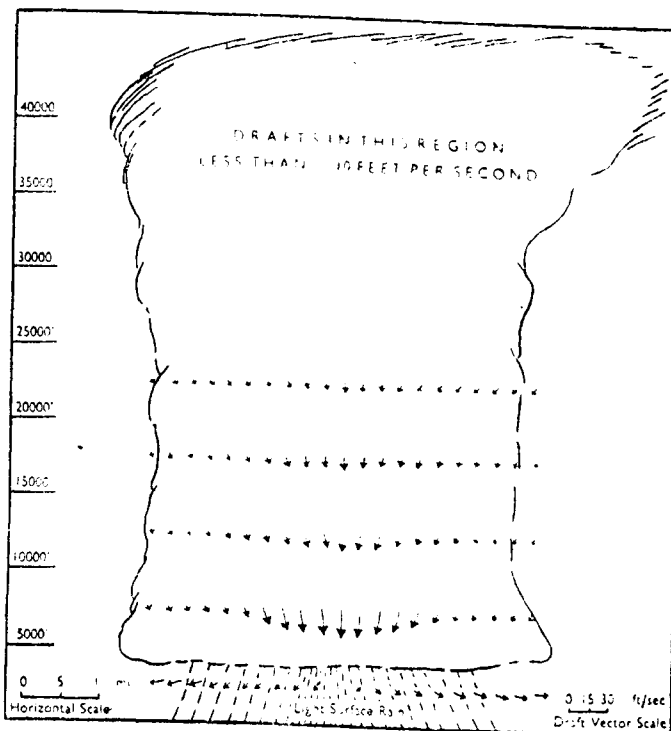


Fig. 11. The dissipating stage, showing light rain and downdrafts in the entire lower part of the cell. (By kind permission of the Editors of *Weather*)

at the summit of the cell to form the ice anvil, but soon the cloud dissipates from below, leaving only scattered remnants at middle and high levels.

A fresh cell frequently develops in the region between two fairly close and fully developed cells, whose cold outflows appear to trigger the initial upcurrent of warm air, by forcing it up where they meet.

The main aim of the investigation, to assist safe air navigation through thunderstorms, was achieved. Hazards from turbulence, hail and lightning were found to be at their maximum around 15,000 ft., just the height at which modern aircraft with pressurised cabins are most frequently flown. The experience of the pilots of the project, who flew many times through the most vigorous of thunderstorms without accident, suggests that safe navigation can be ensured if radar may be developed to allow cloud regions giving echoes to be avoided. Using 10 cm. radar, turbulence outside the cloud echo zones was found to be slight compared to that within them.

REFERENCES

- The report on the American work appears in a paper by H. R. Byers and R. R. Braham in the *Journal of Meteorology* (1948), Vol. 5, p. 71.
- Summaries may be found in
- (a) *Weather* (1949), Vol. 4, No. 7, p. 200 and No. 8, p. 244 by H. R. Byers.
 - (b) *Science Progress* (January 1950), No. 149, p. 78, by P. A. Sheppard.
- An account of the English work appears in *Nature* (1949) 163, 728.
- Two relevant papers appear in the *Quarterly Journal of the Royal Meteorological Society* (January 1950), Vol. 76 - on p. 52. 'The composition of coagulation elements in cumulonimbus' by F. H. Ludlam; on p. 59 'The nature of ice-forming nuclei in the atmosphere' by B. J. Mason.
- This last paper summarizes the results of recent experimental work on the initial processes of ice formation in the atmosphere.

SEMICONDUCTORS

J. T. KENDALL

1. INTRODUCTION

Semiconductors is the name given to a class of crystalline solids with rather unusual electrical properties. Formerly it could be used rather loosely for any substance with an electrical conductivity intermediate between that of metals on the one hand (*conductors*) and insulating materials on the other (*non-conductors*). Such diverse materials as moist wood and paper, electrolytes, iron oxide, coal, arsenic and old lace might be included in this category. During the last two decades, however, our knowledge of the solid state of matter has progressed to such an extent that it is now convenient to limit the term *semiconductors* to a fairly small class of solids which are insulators at the absolute zero of temperature but conduct electricity by the passage of electrons at ordinary temperatures. This definition excludes ionic conductors, such as wet paper or hot glass, and metals which merely conduct badly, such as arsenic or bismuth. Indeed, of the substances previously mentioned, only iron oxide would now be termed a true semiconductor.

In order to understand how the remarkable properties of semiconductors arise it is necessary to have a background knowledge of the various types of crystalline solid and how electrons can move through them. We shall therefore digress a little into a brief consideration of the structure of solids.

2. STRUCTURE OF SOLIDS

A crystalline solid consists of an assemblage of atoms held rigidly together in a regular array. In a perfect single crystal of silver, for example, we would find silver atoms at the corners of what is called a face-centred cubic lattice.* This is a common

* See 'Aspects of Crystal Chemistry' in *Science News* 15, page 51 et seq.

type of structure among metals; but there are many other ways of arranging atoms in space, some of which are quite complicated if many different kinds of atom are involved, e.g. in crystalline silicate rocks. The important thing is, however, that the structure is regular throughout space and is rigid. We must now consider what forces hold the atoms in their fixed positions.

We may distinguish four types of inter-atomic binding in the

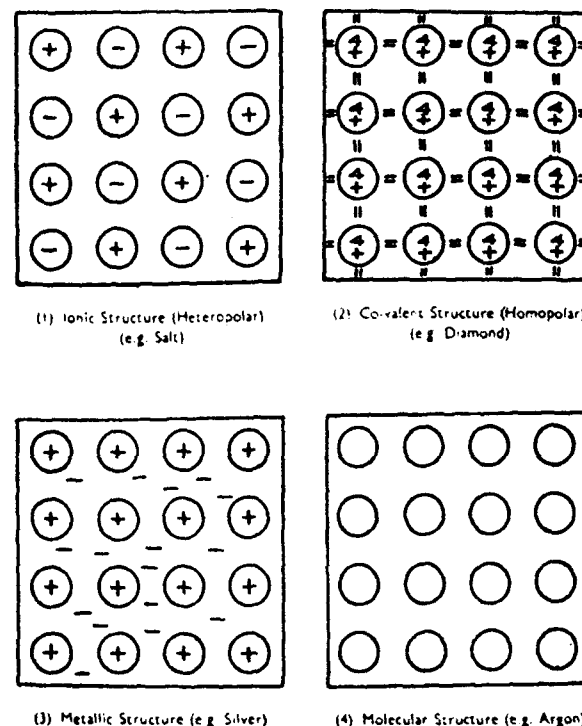


Fig. 12. Types of binding solids

solid state: 1. *Ionic Binding*, in which electrons are transferred from one atom to another, thus forming charged ions which are then held together in a crystal lattice by electrostatic attractive

forces; 2. *Co-valent Binding*, in which electrons are shared by adjacent atoms, which we may therefore consider as positive ions held together in a lattice by a *fixed* cement of electrons; 3. *Metallic Binding*, in which electrons are lost by all the atoms, which we may therefore again consider as a lattice of positive

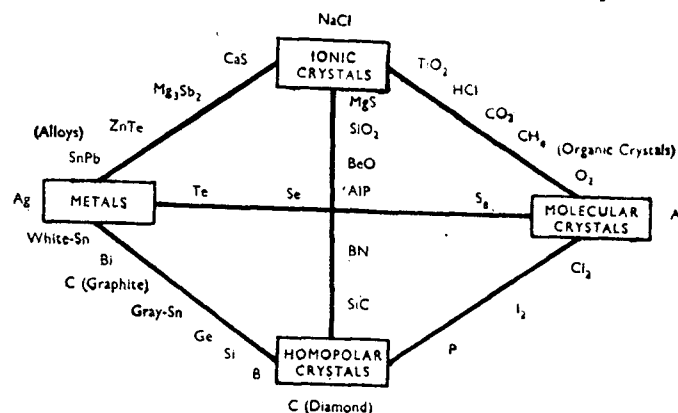


Fig. 13. Classification of crystalline solids

ions, but held together by a *mobile* cement of electrons. In this case, in contradistinction to the second, the electrons are not associated with particular ions, but form an atmosphere of mobile electrons; 4. *Molecular Binding*, in which no electrons can be made available and neutral atoms or molecules are therefore held together only by residual non-electrical attractive forces. The various types of binding are illustrated diagrammatically in Fig. 12. Typical examples are given of each type of structure, but in the case of the majority of solids it is not possible to classify them rigidly, and they usually exhibit characteristics of more than one type. Some distinctive properties of the four solid types are listed in Table I, and the scheme of classification is shown pictorially in Fig. 13.

3. ELECTRICAL PROPERTIES OF SOLIDS

We are particularly interested here in the electrical properties of solids, and it is easily seen how the different types will vary in

TABLE I: Distinctive characteristics of solid types

1. Ionic	2. Co-valent	3. Metals	4. Molecular
Insulator	Insulator	Conductor	Insulator
Hard and Brittle	Very Hard and Brittle	Hard but Malleable	Very Soft
Soluble in water	Insoluble in water	Insoluble in water	Insoluble in water
Medium Melting Point	Very High Melting Point		Very Low Melting Point
		High Specific Heat	Low Specific Heat
Transparent		Opaque	

this respect. Remembering that an electric current consists of the motion of charged particles (i.e. either ions or electrons), we can see that the metallic structure will give rise to good electrical conduction, since there is available a copious supply of highly mobile electrons. The other three structures, on the contrary, will not exhibit electrical conduction, i.e. they will be insulators, since electrons are not available in the ionic and molecular structures, and are not mobile in the co-valent structure. At ordinary temperatures the ions which are available in the first three structures are not very mobile; but at higher temperatures ionic conduction can become important in the ionic solids. We shall not, however, consider this process here, and we can therefore assume that only the metallic structure gives rise to an electrically conducting solid.

It is possible, however, by various means for a small number

of mobile electrons to be introduced into ionic and co-valent solids. Such substances, which in the pure state and at low temperatures are insulators, but can be made electronically conducting by suitable treatment, are true semiconductors. It is not generally possible to introduce a large number of mobile electrons, as in metals, so that the magnitude of the conductivity is several orders of magnitude smaller than for metallic conduction.

4. INTRINSIC SEMICONDUCTORS

There are various ways of so modifying the structure to produce mobile electrons. First, we may simply heat the solid. For a few semiconductors (there are only three certainly known at present - silicon, germanium and lead sulphide) this results in what we may call *intrinsic* semiconduction even at ordinary temperatures. The process is possible because, as previously stated, real substances cannot be rigidly classified as only one of the four solid types, and in these particular cases silicon and germanium (normally classified as co-valent) and lead sulphide (normally classified as ionic) have a little of the metallic character in their actual structure. In other words, the electrons are not *quite* held immobile between ions in the co-valent bond or not *quite* completely held by the positive ion in the ionic bond; but, as the temperature is raised a little, they have a tendency to form a bond of a metallic character in which the electrons are free to move. The number of such bonds formed will depend on the temperature, in an exponential relationship. It has recently been found that even a typically molecular crystal, such as the organic phthalocyanines, can exhibit electronic conductivity in this way. At very high temperatures almost all solids will have this tendency, but in only a very few is it apparent at easily realizable temperatures. No doubt there are a number of such semiconductors to be discovered in addition to the three mentioned above.

5. IMPURITY SEMICONDUCTORS

A more practically important process for introducing mobile electrons is to alter the structure locally by the addition of impurities or by a slight excess or deficit of one kind of atom in a

compound. This is called 'impurity semiconduction' and is shown by many oxides, sulphides, selenides, tellurides, and elements such as silicon, germanium, selenium and tellurium. In the few cases where 'intrinsic semiconduction' occurs as well, the impurities greatly increase this conductivity, especially at lower temperatures. The mechanism of impurity semiconduction is not difficult to understand, and we shall consider a few typical examples. It will be realized, of course, that impurity semiconductors are also insulators at the absolute zero of temperature; and, as for intrinsic semiconductors, thermal energy is required before electrons can be freed.

First - impurities in germanium are shown in Fig. 14. Normally the atoms in solid germanium are arranged in a co-valent crystal lattice in which all of the electrons are therefore immobilized in

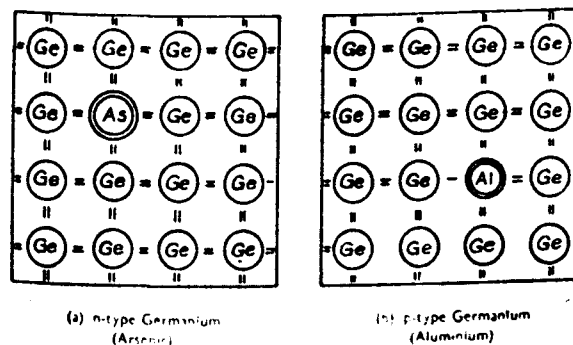


Fig. 14. Substitutional impurities in germanium

pairs. However, in Fig. 14a an arsenic atom, and in Fig. 14b an aluminium atom, has replaced a germanium atom in the structure. Semiconductors are substances in which it is chemically possible for a small amount of such replacement to occur. In *diamond* such replacement of the carbon atoms by others is not possible, and diamond does not therefore become a semiconductor (except temporarily under the influence of high energy particle bombardment, e.g. in a cyclotron). Now the arsenic atom contains five valency electrons, the aluminium atom three,

and the germanium atom four. Since the crystal structure requires four electrons per atom, replacement of germanium by arsenic introduces one extra electron which, not being needed for inter-atomic binding, is free to move. Replacement by aluminium results in a deficit of one electron in the structure. It is a rather curious thing that the absence of an electron from the structure can give rise to electronic conduction. The reason is simply that the absence of an electron imparts a certain mobility to the neighbouring electrons. Normally they cannot move because there is nowhere to move to, the whole structure being complete and perfect; but the presence of the 'hole' allows one of the neighbouring electrons to move in occasionally, and in fact the electrons around the hole will be oscillating in and out continually with a frequency depending on the temperature, and the hole will be effectively spread out over several atomic diameters. (The actual effective diameter will depend on the dielectric constant of the germanium.) Similarly the extra electron associated with the arsenic atom will be effectively spread out over an area.

The motion of holes and electrons in an electric field is shown in Fig. 15. The motion of the hole is equivalent to a general

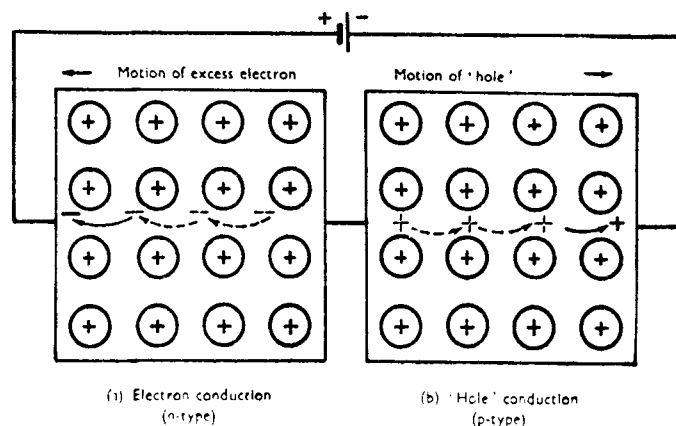


Fig. 15 Motion of holes and electrons

movement of the structural electrons in the reverse direction. It is quite convenient, however, to consider the hole simply as an electronic positive charge. On this basis the motion of holes and electrons can be treated similarly, and one speaks of *hole*, *deficit* or *p*-type conduction on the one hand and *electron*, *excess* or *n*-type conduction on the other. Other co-valent elements which are semiconductors behave similarly.

Our *second* example of impurity semiconduction will be concerned with the metal oxides, which are primarily ionic solids.

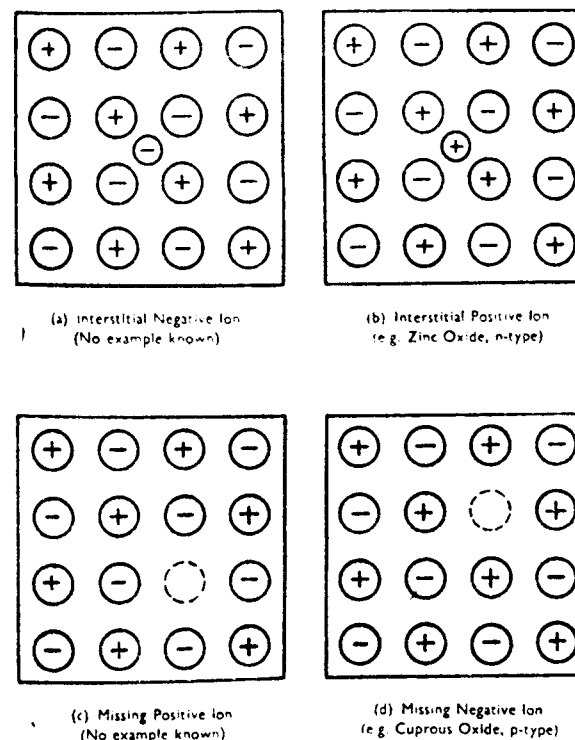


Fig. 16. Deviation from exact composition in an ionic lattice

Here we do not have the addition of extraneous impurity atoms, but simply a slight excess of oxygen or the metal. In the first case we have *oxidation* (*p*-type) semiconductors and in the other *reduction* (*n*-type) semiconductors. If the pure oxide is first prepared it will be an insulator, and will become a semiconductor by oxidation (heating in oxygen or air) in the first case, or by reduction (heating in hydrogen or other reducing gas) in the second. The amount of deviation from the theoretical formula of the metallic oxide will be very small, and may not even be detectable chemically. The occurrence of semiconduction is indeed an exceptionally sensitive means of chemical analysis; a film of semiconducting oxide or sulphide can be made to change its conductivity violently by quite small changes in the gases to which it is exposed.

One can imagine various ways in which the slight deviation from exact chemical composition can occur, and these are shown in Fig. 16. They are either excess positive or negative ions in interstitial positions, or vacant lattice points of either polarity. In practice only two of these cases seem to occur, namely (b) interstitial metallic ions, and (d) missing metallic ions. The first leads to *n*-type semiconductors, e.g. zinc oxide, and the second to *p*-type semiconductors, e.g. cuprous oxide. Most substances exhibit only one or the other type, but a few can be either *n*- or *p*-type according to the conditions - e.g. uranium oxide and lead sulphide. The latter is a particularly interesting case, since it also shows intrinsic semiconduction. The dependence of conductivity on deviation from exact chemical composition is shown for lead sulphide in Fig 17.

6. CHARACTERISTIC PROPERTIES OF SEMICONDUCTORS

(a) Magnitude of Conductivity

As shown in Fig. 18 the electrical conductivity of solids varies over the enormous range of at least 55 orders of magnitude. In passing it is amusing to note that the best metallic conductor, silver, is in fact usable as a material for insulating superconductors (e.g. tin at very low temperature), since it does not exhibit the phenomenon of superconductivity. Most semi-

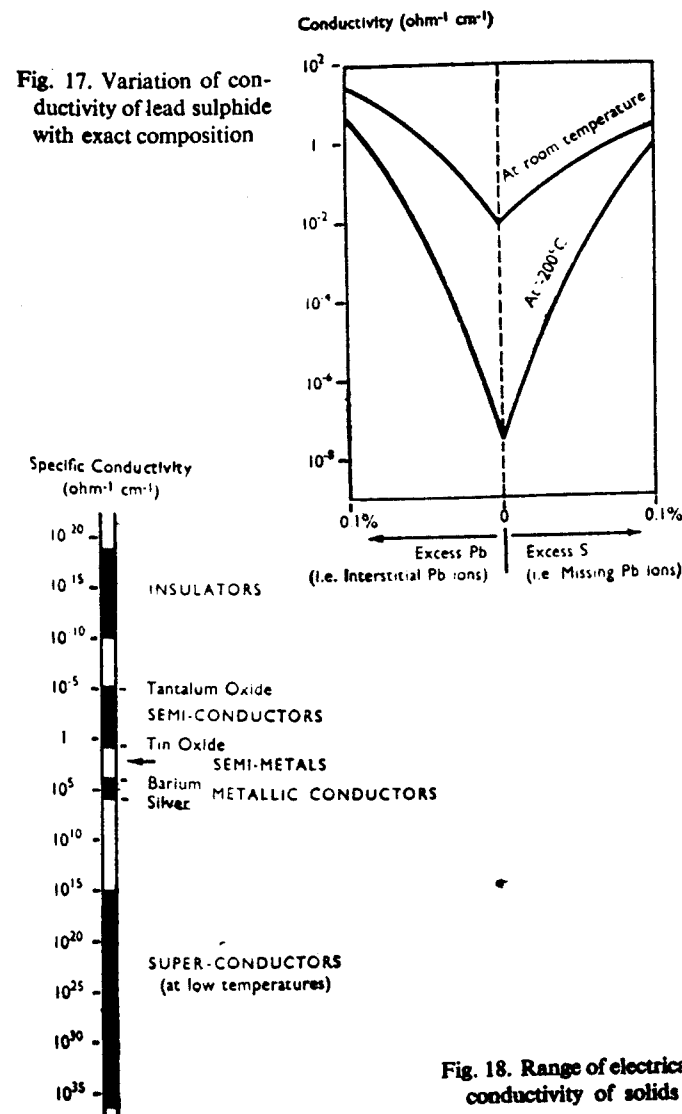


Fig. 18. Range of electrical conductivity of solids

conductors have room-temperature conductivities of between 10^{-8} and $10 \text{ ohm}^{-1}\text{cm}^{-1}$.

(b) *Dependence of Conductivity on Exact Composition and Impurities*

This is very characteristic of semiconductors, and leads to large differences in apparently similar specimens and also to great instability in their properties. In practice many semiconductors are prepared by a heat-treatment process, and very slight variations in the heat-treatment can produce large changes in electrical properties. Manufacture of semiconductors is consequently almost an art rather than a science, for it is not always easy to discern just which manufacturing factors are having effect. Similarly, changes can take place while a semiconductor is in use, particularly if it is subjected to heat.

As an example of the very small quantity of impurity required to influence the electrical properties of a semiconductor it was recently found that a certain batch of germanium, though apparently free from impurities when analysed spectroscopically, still had too high a conductivity for a particular application. The problem of what was causing the contamination was solved by a radio-active tracer technique, which showed that a minute amount of arsenic was left in the germanium by the purification process being used. A slight improvement in the process removed this residual arsenic – even though it had not been spectroscopically detectable in the first place! – and resulted in germanium of sufficiently high resistivity.

(c) *Temperature variation of Conductivity*

Metallic conductors have a small negative temperature coefficient, that is, as the temperature is raised their conductivity slowly decreases. On the contrary, semiconductors have a large positive temperature coefficient of the form shown in Fig. 19. The reason for these quite different behaviours can be seen by considering for a moment the equation for conductivity:–

$$\sigma = e.n.b.$$

where σ is the conductivity, e is the electronic charge, n is the number of free electrons, and b is their mobility. In both metals

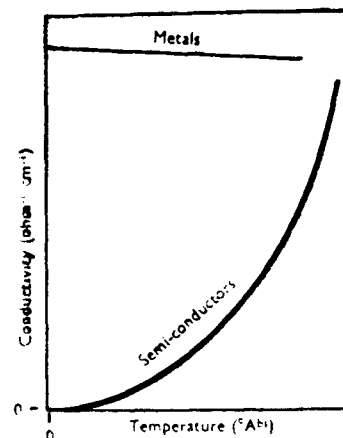


Fig. 19. Variation of conductivity with temperature

and semiconductors e is constant and b decreases slowly with temperature. This decreasing mobility is due to the fact that as the temperature is increased the atoms of the crystal lattice vibrate about their mean positions with greater and greater amplitude. They thus effectively occupy more space and make it more difficult for a moving electron to thread its way amongst them. In metals the number of free electrons, n , is also constant (since *all* the structural electrons are free) so that the slowly decreasing mobility, b , accounts for the small negative temperature coefficient. However, in semiconductors the number of free electrons, n , varies very rapidly with temperature, being zero at 0°K . and increasing exponentially with temperature according to the equation: –

$$n = \exp. [-\epsilon/kT]$$

where ϵ is the 'activation energy' required for the formation of free electrons, k is Boltzmann's constant and T is the temperature on the absolute scale. This variation in the number of free electrons completely swamps the change in their mobility, and for semiconductors we therefore have a very large positive temperature coefficient of almost exponential form.

The large change in the conductivity of semiconductors with change in temperature is made use of in the device known as the 'thermistor', which consists of a small length of a semiconductor (e.g. uranium oxide) the conductivity of which can be used as a measure of its temperature. The device can be made very small and extremely sensitive. It is especially useful for measuring temperatures between about 100°C and 400°C. At high temperatures the device would change its conductivity permanently owing to slight alteration in its composition.

(d) Thermo-electric Effect

Semiconductors have an extremely high thermoelectric e.m.f. (this is the voltage which appears across a pair of junctions between different conductors when the junctions are maintained at different temperatures; it depends on the temperature differential, the mean temperature, and the nature of the conductors). For metals, voltages of a few micro-volts per degree are obtained; while for semiconductors voltages of 300-1,000 micro-volts per degree are usual. The direction of the thermoelectric e.m.f. varies with the nature of the semiconductor, *n*- and *p*-types giving e.m.f.s in opposite directions.

Little use has yet been made of these high thermo-electric voltages for measurement purposes owing to the difficulty, common to all work with semiconductors, of obtaining exact reproducibility. However, there is the interesting future possibility of constructing a generator of electricity based on direct conversion of thermal to electrical energy via the thermo-electric effect. Such a generator has, in fact, been constructed in the laboratory; but its efficiency is not quite good enough to allow it to compete with the present-day steam turbo-alternator. The difficulty is that the thermal conductivity of a semiconductor is much too high in proportion to its electrical conductivity. If the Wiedemann-Franz Law held good (namely that the ratio of electrical conductivity to thermal conductivity is constant for electronic conductors) then a thermo-electric generator would certainly be possible, and moreover could have an efficiency much higher than that of the present-day thermal-

mechanical-electrical energy conversion. Perhaps such a semiconductor awaits discovery, and, if so, it will certainly be one with a high electrical conductivity as well as a high thermo-electric constant.

(e) Hall Effect

If a magnetic field is set up at right angles to a conductor carrying an electric current a voltage is produced in the direction at right angles to both the current and the magnetic field. This is known as the Hall e.m.f. (from its discoverer E. H. Hall, 1880). The effect is much larger for semiconductors than for metals, and use has been made of it for the measurement of magnetic field strength, a thin crystal of germanium being used.

Measurement of the Hall constant of a semiconductor is extremely useful from the theoretical point of view, since it can be used to calculate the number of free electrons present. For an *n*-type semiconductor the Hall constant, *R*, is given by:-

$$R = - \frac{3\pi}{8} \cdot \frac{b_1}{en_1}$$

while for a *p*-type semiconductor,

$$R = + \frac{3\pi}{8} \cdot \frac{b_2}{en_2}$$

(where the subscripts 1 and 2 refer to electrons and 'holes' respectively). By a simultaneous measurement of the conductivity, we can combine one of these equations with the previous one, $\sigma = enb$, and thus evaluate separately both *n* (number of free electrons) and *b* (mobility). It should be noted that the sign of *R* can be used to indicate whether the conduction is due to electrons or holes. For an intrinsic semiconductor (in which the current is carried by both electrons and holes simultaneously) the Hall constant is given by

$$R = - \frac{3\pi}{8} \cdot \frac{1}{en} \cdot \frac{b_1 - b_2}{b_1 + b_2}$$

and this may have positive or negative values according to the relative mobilities of electrons and holes.

(f) Contact Rectification

The contact between a semiconductor and a metal, or between two semiconductors, does not normally behave as an ordinary resistance obeying Ohm's Law (*i.e.* Current proportional to applied Voltage); but the relationship of current to voltage is usually of the form shown in Fig. 20 (in which current is pro-

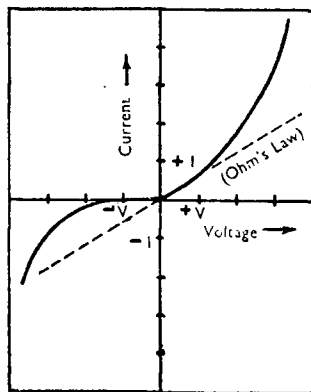


Fig. 20. Relationship between current and voltage at a semiconductor contact

portional to some power of the applied voltage). Moreover the electrical resistance of the contact not only varies with applied voltage, but is usually different in the two opposite directions. Sometimes this condition is carried so far as to amount to virtual unidirectional conductivity, or 'rectification'. For example in Fig. 20, rectification will occur for applied voltages less than V .

It is not easy to arrive at a completely satisfactory explanation of this phenomenon, which is so characteristic of semiconductors. Even now the theoretical interpretations of N. F. Mott and W. Schottky are incomplete and not wholly satisfactory. In spite of this lack of precise knowledge very great practical use has been made of contact rectification.

So-called 'plate' rectifiers are used for converting A.C. to D.C. current, and quite large powers can be handled. The type using selenium is well known, and consists of a layer of selenium

deposited on a metal plate, with another metal plate pressing on to the selenium layer. As in all rectifiers, there must necessarily be *two* metal-semiconductor contacts in series, and it is therefore necessary to arrange the mechanical construction of the system so that one contact has a much higher electrical resistance than the other. Rectification then takes place at the high-resistance contact.

Another well-known type of plate-rectifier, discovered by Grondahl in 1926, uses cuprous oxide as the semiconductor. The surface of a copper plate is oxidized by heating in air to form cuprous oxide, and a second metal plate, usually lead, is pressed on to the oxidized surface. In this case rectification actually occurs at the Cu/Cu₂O boundary, where there is a layer of cuprous oxide having the exact composition Cu₂O, and therefore having a very low conductivity. Nearer the free surface the oxide contains a 'deficit' of copper ions and therefore becomes a *p*-type semiconductor. The direction of rectification at a contact can always be predicted, since electrons pass more easily from an *n*-type semiconductor to a metal, or from a metal to a *p* type semiconductor, than in the reverse directions. Thus in the cuprous oxide rectifier the direction of easy flow of electrons will be from the copper base to the cuprous oxide, *i.e.* with the copper base negative.

So-called 'crystal' rectifiers have been well known since the early days of radio. In this case one of the two metal-semiconductor contacts is made very small (and hence of high resistance) by constructing it in the form of a 'cat's whisker' pressed lightly on to the crystal, the other contact being made of large area. Under these conditions rectification occurs at the 'cat's whisker' contact. Early wireless enthusiasts will remember that a lot of time and patience could be consumed in endeavouring to find sensitive spots on a crystal surface, and this is just another manifestation of the extreme dependence of the electrical properties of semiconductors on exact chemical composition. As the chemical composition, or distribution of impurities, varies slightly over a crystal surface, so the electrical properties vary enormously.

One of the popular semiconductors used in early 'crystal sets'

was silicon carbide (also known as carborundum). It is noteworthy that this substance can exhibit *n*- or *p*-type conduction according to its colour, green crystals being *n*-type and black crystals *p*-type. Such crystals therefore rectify in opposite directions. Silicon carbide is a covalent substance, of similar structure to diamond but with half the carbon atoms replaced by silicon. The reason for its semiconductivity is, however, not yet known, and may be due to the presence of extraneous impurities such as aluminium or iron atoms, or simply to a slight excess of silicon or carbon atoms in the structure, the one leading to *n*-type and the other to *p*-type conduction.

Within the last few years crystal rectifiers have been developed, using silicon or germanium, which have better rectifying properties than the earlier types and – in striking contrast – almost perfect stability to mechanical vibration. These devices are being used in great numbers for rectifying (or ‘detecting’) very high-frequency electric waves such as are used in radar.

(g) Conductivity of Powders

From the previous paragraphs it will be appreciated that while a single crystal of a semiconductor will behave as a normal resistance obeying Ohm’s Law (current proportional to resistance) the contacts between crystals will not behave normally. This results in anomalous effects when a current is passed through a powder which has been pressed together to form a ceramic solid body. Such a composite material will not exhibit rectification, since on the average there will be as many contacts rectifying in one direction as the other, but it will exhibit departures from Ohm’s Law. In other words we will have a curve for V/I similar to the one shown in Fig. 20 but symmetrical about the origin, positive and negative branches being identical.

Practical use has been made of such ceramics, whose resistance varies with applied voltage, in such applications as surge-diverters (particularly for protection of electrical installations from lightning). Materials are made (an example is the material known as ‘metrosil’, made from powdered silicon carbide pressed together with clay) for which the electric current varies

as about the fifth power of the applied voltage. When a high-voltage surge, such as a lightning stroke or switching surge, reaches such a material, it will behave as a comparatively low resistance and will thus *divert* the surge. On the disappearance of the voltage surge the material will regain its former condition of high resistance. In actual practice, *e.g.* for the protection of an electrical transmission line, several discs of this material would be placed in series with a spark gap across the line.

(h) Photo-conduction

The normal means of freeing electrons in semiconductors is by absorption of thermal energy, that is by raising the temperature. However, for a few substances another special process is available, namely the absorption of light-quanta. Each light-quantum absorbed can release an electron, and the substance then becomes conducting, *i.e.* it exhibits photo-conduction. In this case the temperature of the material does not appreciably affect the number of photo-electrons produced, and therefore photo-conductivity, in contrast to the normal semiconductive process, has a positive temperature coefficient similar to that for metals. The reason for mentioning the phenomenon here is that many semiconductors are also photoconductors when illuminated with light of a particular wave-length. Examples which are important from the practical point of view are the sulphides, selenides and tellurides of zinc, cadmium and lead.

7. LIST OF THE MORE IMPORTANT SEMICONDUCTORS

Table II lists some of the semiconductors which have been investigated in detail, but is by no means exhaustive.

8. THE CRYSTAL TRIODE

The latest addition to the peculiar properties of semiconductors is that known as ‘transistor action’, which was discovered at the Bell Telephone Laboratories in 1948. Use has been made of this property for the subsequent development of the crystal triode, and there is little doubt that further development will greatly increase the importance and usefulness of this device, which in

TABLE II: Semiconductors

Elements	B (?), Si, Ge, Se, Te..	
Oxides	<i>n</i> -type	Al_2O_3 , Bi_2O_3 , CdO , Ce_2O_3 , CuO , Nb_2O_5 , Ta_2O_5 , TiO_2 , WO_3 , ZnO , $(UO_2)_2$, U_3O_8 , $MgTiO_3$, $BaTiO_3$, Fe_2O_3 , SnO_2 , V_2O_5 .
	<i>p</i> -type	CoO , Cu_2O , FeO , NiO , UO_2 .
Sulphides, Selenides & Tellurides	<i>n</i>	Ag_2S , PbS , CdS , ZnS , SnS , $PbSe$, $PbTe$.
	<i>p</i>	PbS , CuS_2 .
Halides	CuI	
Carbides & Nitrides	SiC	

principle can be made to carry out all the functions of thermionic radio valves.

There are various types of 'transistor'; a well-known one is illustrated in Figs. 21 and 22. Two small 'cat's whisker' contacts (known as the *emitter* and *collector*) are made very close together on a block of *n*-type germanium, to which a low-resistance *base* contact is also made. Biasing voltages are applied as in Fig. 21 so that the emitter is *positive* and the collector *negative* with respect to the germanium. Since germanium is an intrinsic as well as impurity semiconductor the normal flow of electrons towards the *emitter* is supplemented near the contact point by a flow of 'holes' in the opposite direction. The number of 'holes' so injected into the germanium will vary according to the bias or signal voltage impressed on the emitter. These holes will penetrate only a short distance into the body of the germanium since they will eventually be neutralized by the free electrons there. However, they will produce a positive space charge in the germanium near the emitter, and thus more

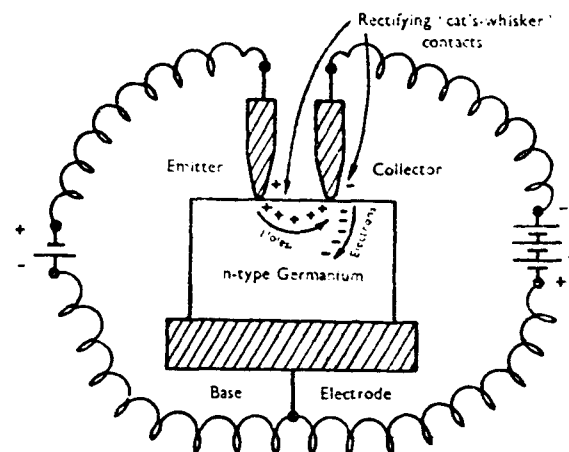


Fig. 21. Transistor mechanism

electrons will be attracted into this space, the conductivity of which must therefore increase. If the *collector* point is placed near enough to the emitter (less than 0.1 mm. away) it will be within this region of enhanced conductivity; and moreover the normal flow of electrons away from the collector (which flow is small since the collector is biased to give rectification) will be supplemented by a flow of holes. The *collector* will in fact collect the holes emitted by the *emitter*, and any signal voltage applied to the emitter will result in an amplified signal at the collector.

The device can be made to function as an amplifier, and some typical circuits are shown in Fig. 22. The performance data for these circuits are summarized in Table III. It will be seen that maximum power outputs (with tolerable distortion) are at present very small compared with those obtainable with thermionic valves. The data given refer to average crystals of germanium.

Other forms of transistor are being developed. For example the crystal may be constructed in the form of a thin wedge, the emitter and collector contacts being placed on either side. This avoids some of the difficulty associated with accurate positioning of the 'cat's whisker' contacts. A rather different type of tran-

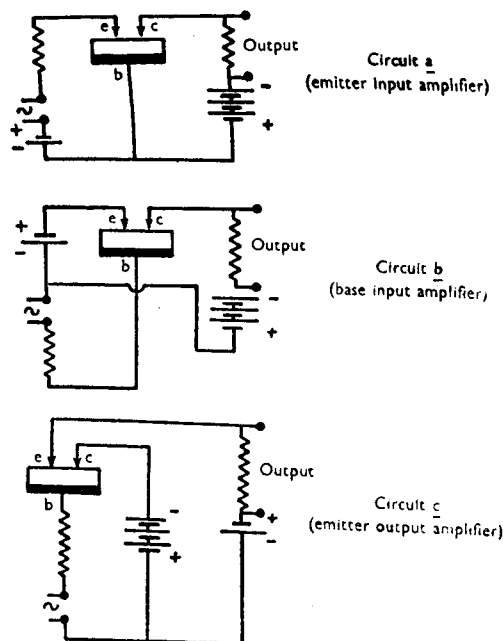


Fig. 22. Transistor amplifier circuit

sistor is illustrated in Fig. 23. In this, known as the filamentary type, holes from the emitter are carried along the length of a narrow germanium film and modify the current at the collector.

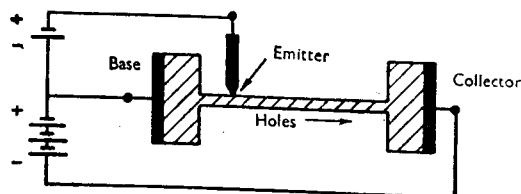


Fig. 23. Filamentary transistor

Owing to the comparatively large distance apart of the collector and emitter contacts there is, in this type of transistor, an appreciable time interval between the injection of holes at the emitter and the consequent change in the collector current.

TABLE III: Performance Data of Transistors

	Circuit a	Circuit b	Circuit c
Input Impedance (ohms)	700	5,000	20,000
Output Impedance (ohms)	30,000	10,000	500
Power Gain (decibels)	16	23	16
Voltage Gain	40/1	20/1	less than 1/1
Current Gain	approx. 1/1	10/1	40/1
Power Output (watts)	2×10^{-2}	2×10^{-2}	10^{-4}

9. CONCLUSION (AND APOLOGY)

Informed and alert readers of the preceding pages will perhaps have noted that some sort of account of semiconductors has been given without once mentioning such advanced concepts as Wave Mechanics, Brillouin Zones, Energy Levels, Fermi-Dirac Statistics and all the other paraphernalia of modern atomic physics. This has, as may be realized, been done with a purpose in mind - which was to write about semiconductors in such a way as to be intelligible to non-physicists, and even to those physicists (if such there be) who do not really understand the jargon of modern physics. In doing so the picture has necessarily become a little blurred and incomplete; but the author hopes, and believes, that it is none the less accurate as far as it goes. In spite of this simplification, the subject of semiconductors is a somewhat intractable one, and the author cannot hope to have made himself quite intelligible to everyone. To those of his readers who may remain alert but still uninformed the author offers his sincere apologies.

THE INSTRUMENTATION OF THE LIVING BODY

J. W. S. PRINGLE

THE term instrumentation is used by engineers to describe the way in which a piece of machinery is made to yield information about the various processes going on within it. The driver of a motor car, for example, who wishes to keep within the 30 m.p.h. speed limit, uses his speedometer to give him information which he would not otherwise be able to obtain, and regulates his driving accordingly. Other instruments, such as the petrol gauge, give him different information, on which he does not normally take any action, but all of them are serving the same purpose, to tell him more about what is going on in the machinery of the car than he can perceive without these aids.

In any complicated piece of machinery a lot of things happen which are of no immediate interest to the person using it. No housewife wishes to know the speed at which the electric motor is turning when she is using a vacuum-cleaner. She assumes that the designer has taken care to ensure that the motor regulates itself to a speed suitable for the purpose for which the machine is intended, namely collecting dust. In order to use it correctly she needs to know merely whether it is working or not, and for this she uses another of its properties – the noise it makes – as an indication by which to control the way she uses it. Strictly speaking this sound-production is equivalent to an instrument, and the fact that the information comes by way of the ear rather than the eye is irrelevant. The engineer would probably consider this a point of academic interest, but we shall see that it cannot be ignored when we are thinking not only about the machinery of the vacuum-cleaner, but also about the machinery of the living creature, the housewife, who is using it.

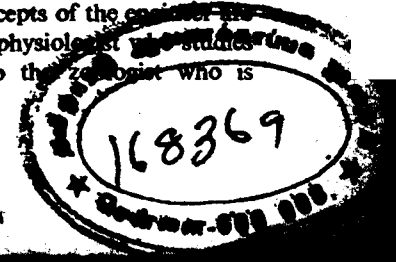
Even in the restricted sense of the engineer, there is, of course,

a very wide range of instruments which may be used for the purpose of giving information, ranging from simple pressure gauges indicating the head of steam in a boiler to the elaborate devices by which the pilot of an aircraft gains information about his position in relation to the ground stations of a radar navigation system. The interesting thing is not the variety of instruments which are possible, but that out of the wealth of detail which the designer has produced to meet the demands of the modern age there have emerged a number of general principles of instrumentation, which are now available as a guide to the engineer and as a subject for study by the scientist.

The purpose of all instruments is to convert a measure of some feature of the functioning of the machine into a form in which it can be perceived by the senses of man, to be used by him as a check on its correct operation or as a guide to his own immediate movements in control of the machine's working. The instruments serve as a channel through which the human controller inserts himself into the mechanism. A pilot flying in dense fog uses the instruments on his dashboard as a guide to the movements he makes in controlling the aircraft. At the touch of a switch he can disconnect himself from the machine and bring into action instead an automatic pilot which uses the same information as he was deriving from the dials in front of him, and whose motors replace his muscles in operating the aircraft controls.

The channel through which information passes from the machine to the human brain includes, besides the instruments on the panel, the instruments in the body of the observer – his eyes, ears, and the sense organs of his hands, feet, joints and the rest of his anatomy. These take the information from the dials and the other indications given by the machinery and pass it on through his sensory nerves to his brain, whence issue the orders to the muscles which move the controls. The principles of the science of instrumentation apply as well to these sense organs as to the instruments proper, and the concepts of the engineer are becoming of increasing interest to the physiologist who is studying the working of the sense organs, to the biologist who is

Am 1, 246
N50-12



concerned to explain how they came into existence, and to the psychologist who needs to know exactly what information arrives via the nerves at the brain whose working he is trying to understand.

This article tells something about the working of the sensory 'instruments' by means of which our brains and the brains of other animals are able to derive information from the world which surrounds them. The machinery of the living body differs in many important ways from the machinery to which we are accustomed, not least in that it is basically a chemical rather than a physical machine. But one of the most interesting developments in modern science is the way in which similarities in general principle (mechanisms) are being discovered between the operation of very different types of machinery. This study of methods of operation, called 'cybernetics' by the group of American workers led by Norbert Wiener, has already gone a long way also in Britain.

FUNDAMENTALS OF SENSORY PHYSIOLOGY

Information passes from one part of the body to another mainly through nerve fibres, which have evolved in higher animals an accurate method of sending messages which is considerably faster than that found in any of the other functional systems (such, for example, as the 'hormones' or chemical messengers). When the phenomena under consideration are rapid it is therefore safe to assume that account need be taken only of events in nerves.

The unit of function in the nerve fibre is the impulse, a self-maintaining disturbance which passes over the whole surface of the fibre much as an explosion wave travels along a train of gunpowder. It appears to be a universal rule of sensory physiology that *intensity* is represented in the nervous part of the communication channel as *frequency* of nerve impulses, *quality* being distinguished by the living body by means of a diversity of nerve endings, each sensitive to only a limited range of stimuli. Thus, of the complete range of physical phenomena occurring in the outside world, each nerve ending with its associated sense organ

reacts only to a part: the elements of the retina of the vertebrate eye to light waves in a particular wave-band arriving from a particular direction relative to the orientation of the eyeball, the elements of the semi-circular canals of the ear to angular (turning) accelerations of the head in a particular plane, and the fine hairs on an appendage (called the cercus) of the cockroach to mechanical vibrations in the air in a particular low-frequency range, including frequencies too low to be audible as sound. To the zoologist it is apparent that the nature of the selection made by the sense organs of the body from the infinite range of possible stimuli is determined by natural selection. Those sensitivities are preserved and developed which confer on the animal an advantage in the struggle for existence. It follows that the sensory equipment of a creature such as an insect, living in a very different environment from that of man, may involve a selection from the available stimuli different from that to which we are accustomed in ourselves.

The study of animal evolution shows us that there has been a progressive *extension* of the range of stimuli to which the organism as a whole reacts, and a simultaneous *reduction* in the range of sensitivity of the individual sense organ. This means that the higher animals have more sense organs and more sensory nerves. Thus *Paramoecium* (the microscopic slipper animalcule) responds hardly at all to the general mechanical features of its environment. The jelly-fish *Aurelia* has an accurate representation of its position relative to the vertical through its gravity-sensitive statocysts, but is almost indifferent to rotations. The dog-fish, by means of its semi-circular canals, adds a measure of angular accelerations, and the fly goes one better, being provided with a sense organ (the haltere or balancer) sensitive to angular rotations using the principle of the gyroscope. The reduction in the range of sensitivity of the individual sense organ may be illustrated by reference to a type which does not occur at all in man – the strain-sensitive nerve endings found in the skeleton of insects. These respond to distortion of the skeleton as the joints of the legs are moved, and are similar in function to the strain-gauges which the engineer sticks on the surface of aircraft wings

to measure the forces present during flight. The primitive form of this nerve ending, found in caterpillars, shows no discrimination of the direction of strain; that is, it is excited whichever way the leg is bent. The most highly developed type accurately selects one directional component from the complex of stresses present in the skeleton and gives information about the magnitude of that alone. The advantage of this increasing complexity of instrumentation as Evolution proceeds is presumably to be sought in the increased information which is thereby made available to the animal about the changes in the environment which affect it, so that all the time its methods of finding food and avoiding enemies are improving.

Some of the most interesting aspects of sensory physiology come to light only when we examine in more detail the way in which the information is transmitted by the sensory nerves. This has become possible through the development of electrical methods for recording the passage of nerve impulses from the sense organs to the brain. In engineering terms, the functioning of the living body can be analysed, and the role of the sensory instruments isolated from the operation of the whole of the organic mechanism. An example of such an investigation is shown in Fig. 24. In this experiment the stimulus (continuous line) was a

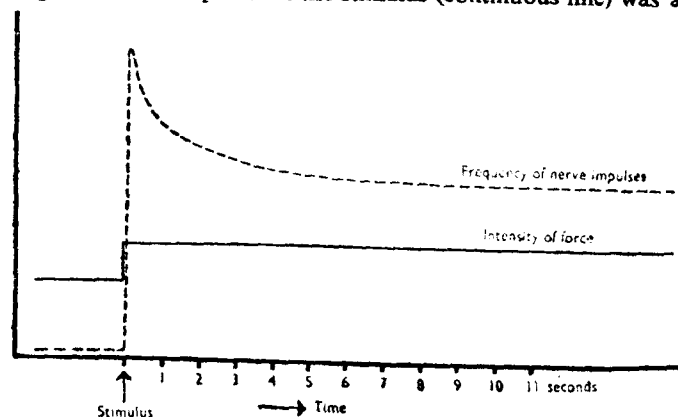


Fig. 24.

change from one intensity to another in the force applied to a joint of an insect appendage, and the graph (dotted line) shows the frequency of impulses recorded in a nerve arising in one of the sense organs sensitive to the strains produced by the applied force. Although the force, after its initial change, remains constant, the frequency falls off for several seconds and becomes constant only at a value much lower than the maximum which it reached at the beginning of stimulation. This decline in frequency when a stimulus is maintained is known as *adaptation* and is a very general property of sense organs, though it varies in degree from one type to another. In some light-sensitive endings in animal eyes it is both slow and slight, so that the frequency of impulses produced by a prolonged illumination is not much less than that present at the beginning of a flash. In touch endings, on the other hand, it is usually rapid and complete. In this latter type of sense organ a constant stimulus fails to maintain the discharge of impulses in the nerve, and a continually changing stimulus is needed to keep up a constant excitation. We are all aware of this from the fact that the sensation of touch at the finger-tips lasts for only a short time except when the object or the finger is moved.

Unlike the majority of engineering instruments which pass on to the observer a measure of the instantaneous value of the particular feature, such as force, pressure, or the like, for which it is designed, sense organs transmit a measure made up of a combination of the instantaneous value and of the rate of change. This is the first point to be noted.

The other important generalization was discovered by physiologists even before the coming of electrical methods of analysis, and is contained in the so-called Weber-Fechner Law. This law, which was first derived from experiments on human subjects, states that the minimum perceptible change of intensity of a perceived phenomenon is proportional to the intensity already present. In other words, for us to recognise that a sound has changed, it requires a greater change of intensity if the sound is loud than if it is soft. Expressed differently, the law describes a logarithmic relationship between the intensity of the phenomenon

and the effectiveness of excitation. It is found to hold good over a limited range of intensities for the human eye and ear, and the more recent electro-physiological researches of Hartline and Graham on the eye of the king-crab, and of Matthews on the tension-sensitive endings in frog muscle have confirmed it, again over a limited range, for individual sense organs. We are probably safe in assuming that it is the general rule for the sensory instruments of the body to measure not the intensity of a phenomenon in the outside world, but approximately the logarithm of that intensity. This is the second point to note. Since the experiments from which the conclusion is drawn involve determinations of the amount of change and not the rate of change, it is a feature of the physiology of sense organs quite distinct from adaptation, which was our point number one.

THE LANGUAGE OF MATHEMATICS

To carry any further the comparison of sense organs with man-made instruments we now have to introduce the language of mathematics. To some extent this has already had to be done in talking about a 'logarithmic relationship'. Many people are unnecessarily frightened by mathematics, when in fact they could find it a most useful way of thinking, especially when, as in this article, we are concerned with processes or mechanisms.

The mathematical ideas which must be introduced are the concept of the 'function' and the concept of 'operations' and the 'operator'. Let us consider these in turn.

In mathematical statements we often require to express concisely the way in which one phenomenon is related to another. For example, suppose we wish to study the rate of filling of a bath. We know that if the tap is not touched after it has been turned on, the water comes out at a steady rate, and the bath fills up. The level of water in the bath is a 'function' of time, since as time changes the level also changes. Similarly the rate of flow of the water is a 'function' of the setting of the tap, and the level of water in the bath is a function of the rate of flow. One phenomenon may quite well be a function simultaneously of more than one other phenomenon. Thus, if the setting of the

tap is varied, the level of the water is a function both of time and of the setting of the tap, and we can calculate the level of water only if we know both the way in which the setting of the tap has been altered and how long the water has been running. It is convenient to write these statements in a form of mathematical shorthand; we state that

$$L = f(S, t)$$

where L is the level of water in the bath, S is the setting of the tap, t is time and $f()$ stands for the statement 'is a function of'. The comma means 'and'. It is unnecessary and may be very difficult to state what is the nature of the function in any particular case.

An 'operation' in mathematics is a word which means 'a process of making a change in a regular manner'. Suppose that, having completed our experiment on one bath, we now substitute another bath, bigger than the first, and fit to it exactly the same tap. Again the level of water will be a function of the setting of the tap and of time, but it will be a different function. For the same tap settings after the same time the level of water will be lower. The substitution of one bath for another has altered the mathematical expression in a regular manner and we may write

$$L_2 = K[f(S, t)]$$

where $K[]$ is shorthand for the effect of a particular 'operation' performed on the function $f(S, t)$. K is the 'operator'.

Mathematicians find this type of shorthand useful because of certain rules which have been found to hold good when several operations are performed one after the other. If K_1 is one operator and K_2 is another operator, then for certain types of successive operation the final result of the two operations is the same as that of a single operation with (K_1 multiplied by K_2). The reader can check that this is so by considering the operators 'twice' and 'three times'. This rule obviously leads to a considerable saving of time and effort in working out the final result of several operations.

The operation we have considered, that of substituting one bath for another, is a simple case, and K will probably be a

simple numerical constant if the two baths are sufficiently alike. The same type of shorthand can be used if the operation consists, for example, in pulling out the plug, though in this case the operator will be not a constant, but a process which changes the nature as well as the value of the function $f(\cdot)$. Within limits, however, the same reasoning can be applied to successive operations of this more complicated type.

Let us now see how the mathematical shorthand helps with our problem of the instrumentation of the living body. All the phenomena in the outside world with which the body is concerned are liable to change; it is, in fact, because they change that the body is concerned with them, and has to be able to react to them. We can state this fact concisely by writing

$$I = f_1(t)$$

where I is the intensity of the particular phenomenon measured by the sense organ we are considering, and $f_1(t)$ stands for a certain function of time whose nature need not be defined. Similarly, since the message from the sense organ to the brain about the phenomenon takes the form of a certain frequency of impulses in the nerve, and this also changes if the intensity of the phenomenon changes, we can write

$$F = f_2(t)$$

where F is the frequency of nerve impulses and $f_2(t)$ is another function of time. We have already been able to find certain generalizations about the properties of sense organs, or, in other words, we have seen that the change from I to F is a regular change. The function $f_1(t)$ is therefore changed by sense organs into the function $f_2(t)$ in a regular manner, and we can talk about the operation of changing $f_1(t)$ into $f_2(t)$ and write

$$f_2(t) = O[f_1(t)]$$

where O is an operator.

There would be no point in this form of mathematical shorthand if each sense organ had an O which was different to that of every other sense organ. But the mere fact that generalizations are possible shows that the O s are not all entirely different, though they may vary slightly. The behaviour of the instruments

of an engineer can, by exactly the same reasoning, also be expressed as operators, and it is in the comparison of the operators of sense organs and the operators of instruments that the interest of this type of investigation really lies.

THE NATURE OF THE SENSORY OPERATOR

All that has been said so far is perfectly safe, and all biologists would agree with it. It is when we come to try to determine the nature of the sensory operator O that difficulties arise. Biology, like any other science, progresses by means of experiments, and the difficulty with experimental results is that they are approximate results, accurate only within the limits of 'experimental error'. Mathematical statements, on the other hand, are exact statements, and every time a scientist expresses his experimental results in the form of exact mathematical equations he is taking the dangerous step of ignoring experimental error. It is a very different thing to say that

$$f_2(t) = O[f_1(t)]$$

and to define the nature of O . However, it is by hazarding guesses at the nature of mathematical relationships that advances in science are made, and we will therefore proceed to do this.

Two separate things are known about the nature of O , contained respectively in the two generalizations already described; namely, adaptation and the Weber-Fechner Law. Mathematically inclined readers may like to know that the property of adaptation is approximately described by the operator $\frac{1 + A/D}{1 + T/D}$

where A is a ratio, T is a time constant, both different for each sense organ, and D is the differential operator which changes a quantity into a rate of change of that quantity. The Weber-Fechner Law describes an operator 'log', or, more simply, a logarithmic relationship. Readers who dislike mathematics may ignore these statements and accept instead an assurance that these particular operators are not peculiar to sense organs, but are very commonly found in a wide variety of engineering mechanisms, particularly those concerned with control systems. They are not, however, often found in engineering instruments,

for the very good reason that the engineer wants the instrument to pass on the particular information in which he is interested more or less unchanged. If his instruments had the complicated operators of sense organs, he would find it very difficult to use them to discover what was going on in the machine.

We have now reached a somewhat surprising conclusion. We know that information about what is going on in the world outside the living body reaches the brain mainly through nerves, and is carried in the nerves only in the form of a frequency of nerve impulses. It now appears that the message which goes up the nerves does not tell the brain what is the intensity of a particular phenomenon, but gives it instead a message related to the intensity of the phenomenon in a complicated way, the change being described by two operators. I have not tried to combine the two into a single operator, as I might have done, because it is not at present quite clear how they should be combined; more experiments are needed. But it makes it look very strange that we appear to be aware of the instantaneous value of many of the changing features of our surroundings, when in fact our brains are receiving information about something quite different. It is equally strange that animals appear to react to the instantaneous state of features of their environment, as, for instance, when a swallow picks a fly off the surface of a pond. Unless the brain of the swallow is getting information about where the fly is, how does it know where to catch it?

The solution to this problem is not, as one might be inclined at first to think, that the results of sensory physiology are wrong. They may be inexact and incomplete, but they are not wrong enough to get over the difficulty. Part of the answer, at least, can be found by looking at the physiology of the rest of the living body. Let us consider first the animals other than conscious man.

The reactions of animals to changes in their surroundings have been studied a great deal by biologists interested in animal behaviour. The change is called a stimulus and the response is usually measured as some form of muscular movement. The simplest experiments are those concerned with what are called

'reflex arcs' (Fig. 25) in which the communication channel from the brain (4, Fig. 25) through the motor nerve (5) and muscle (6) to the response (7) is studied along with the channel from the stimulus (1) through the sense organ (2) and sensory nerve (3) to the brain. When both stimulus and response are phenomena

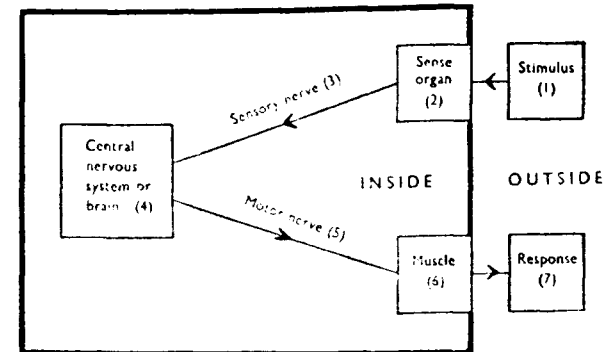


Fig. 25.

which can be observed outside the animal there is no need for the elaborate techniques of the electro-physiologist, and many reflexes have been so described. There is no doubt that in many cases the stimulus-response relationship is superficially simpler than those that we have found in sense organs.

The same process of analysis used in sensory physiology can, however, be applied to the other parts of the reflex arc. By electrical methods impulses can be produced in motor nerves at a constant frequency and measurements made of the resulting changes of tension or length in a muscle. Since both these phenomena are functions of time, the operator of muscles can be worked out just like the operator of sense organs. Similarly by a comparison of the frequency of impulses in sensory nerves with those in motor nerves we can arrive at the operator of the central nervous system. The nerves themselves complete the arc by supplying an operator which allows for the time taken for the impulse to get from one end to the other, and by combining all

these operators in the right way we should get a single operator describing the change from the $f(t)$ of the stimulus to the $f(t)$ of the response.

This full analysis of a single reflex arc has not yet been carried out in an exact manner, but the plan of action is clear. Parts of it have been done on one reflex and parts on another, and we can already see the way it is coming out. Fig. 26 shows the relation-

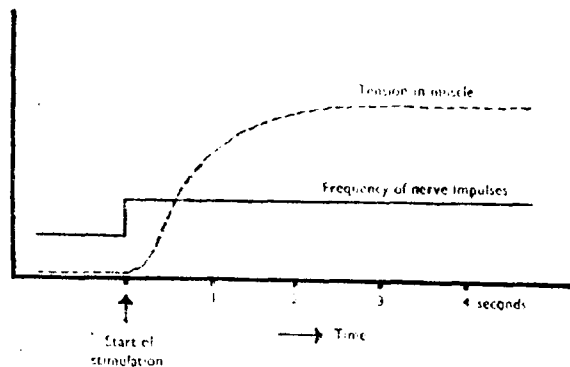


Fig. 26.

ship between the frequency of impulses in a motor nerve in a crab leg and the tension in the muscle to which the nerve runs, as determined by C. F. A. Pantin. Just as in the experiment to which Fig. 24 refers, although the frequency of impulses, after its initial change, remains constant, the tension goes on changing for several seconds, and only later reaches a steady value. An operator can again be worked out which describes this relationship, and the interesting thing is that this operator is nearly the 'inverse' of the operator of sense organs. That is, if the sense organ changes the message in one way on its passage into the internal communication channels, the muscle changes it back to its original form on its way out again. It is not surprising, therefore, that when we look only at the whole animal from the outside, we remain unaware that these complicated changes are going on at all.

It is when we think about our own human brains that the observations described in this article take on such a strange appearance. What the environment does to us and what we do to it may be related in a simple manner, just as they are in some animal reflexes. But the message our brains receive is not simple, and we have to try to find out why this elaborate sequence of operations has been evolved by the living body, when to all appearances it would have been equally good to use some of the straightforward methods of the engineer.

One thing can be seen at once. The nerves of our bodies, unlike the wires of a piece of machinery such as a wireless set, take an appreciable time to transmit a message from one part of the body to another. The quickest movement a man can make with his hand in response to a flash of light occurs about 1/5th second after the flash. Engineers now know that one of the most troublesome things to have to allow for in automatic machinery is a delay in the control channel. When delays are present, the whole mechanism can very easily go into uncontrollable oscillations. They get over the difficulty by including in the mechanism a device known as a 'lead network', whose working is equivalent to an operator of the type found in the adaptation property of sense organs; the oscillations then stop. The living body has evidently discovered this trick and puts this property into the very first component part of all its reflex arcs in order to compensate for the delay in the nerves.

But this is not all. The operator of the muscle is of the same type as that of another component part of the mechanisms of the control system engineer: this time of a 'lag network' which makes changes in the output from the component lag behind changes in the input. If there is a 'lag network' in a control system, the 'lead networks' can be adjusted to balance this lag; they can also introduce more 'lead' than is needed to compensate for the delay in the communication channel. In such a system the response can be made to depend not on the present but on the future value of the stimulus, and this is obviously an advantage to an animal, since it enables it to take action in time to meet a situation instead of waiting for it to happen. No cricketer would

be able to catch a cricket ball if his reactions were controlled only by the stimulus he received when the ball actually touched his hands. He uses his eyes, and reacts not to what he sees but to a future value of the phenomenon, namely the position of the cricket ball.

This type of analysis of the working of living bodies does not, of course, tell us anything definite about the nature of our conscious sensations. Consciousness is a thing which physiological science cannot study, because it cannot be measured. All the scientist can do is to measure the phenomena occurring in conscious beings and the phenomena produced by them. We can in this way discover a lot about the messages which reach the brain where consciousness is supposed to reside, and I have tried to show in this article that these messages are not messages about the intensity of the phenomena at the moment when we think the measurement is being made. The message is subject to an operation in the sense organs which changes it more nearly into a message about the logarithm of what the intensity will be a short time ahead if it goes on changing in the same sort of way that it has been doing in the immediate past. Why this message makes us feel consciously that we are travelling through time at the same instant as the phenomena in the outside world remains a mystery. It would be much more reasonable if it made us feel we were a bit ahead.

RESEARCH REPORT

A. W. HASLETT

DATING THE PAST

DATES are to a record of past ages as is the surveyor's chain to the measurement of distances. Both serve a double purpose: first, to provide a standard, so that measurements can be expressed in recognisable terms; and, second, to make comparisons possible between parts of the world which cannot otherwise be related. Within the period covered by historical records, dates present only minor difficulty. There is comparatively slight difficulty also over the longest periods, those between the formation of the oldest rocks, around 2,000 million years ago, and the last few million years; for these are adequately recorded by the slow breakdown of uranium into lead. Most of the difficulties lie in the interval between about one million years ago and the beginning of history – the date of which naturally varies between one part of the world and another. The ideal state of affairs, for this period as any other, would be to have one or more absolute methods by which materials of all kinds could be accurately dated. In practice, there are limitations both on the kinds of materials which can be dated, and also on the conditions in which dating is possible, so that comparative methods are still necessary. The object of this note is to bring together two recent examples in which progress has been made, one by absolute methods, and the other by comparative.

The absolute method has been developed by Professor W. F. Libby of the University of Chicago, and was mentioned by J. T. Kendall in a note published in *Science News* 10. It depends on the fact that radioactive carbon (carbon-14) is continuously formed in the earth's atmosphere from the impact of cosmic radiation from surrounding space; that the rate of breakdown of this

radioactive carbon is slow enough to allow it to become evenly distributed between the atmosphere, plants which absorb it from the atmosphere, and animals which feed off plants. In figures, the 'half-life' of radio-carbon is about 5,700 years. This means that every 5,700 years the initial radioactivity of any specimen is reduced by half: for example, at the end of 11,400 years, radioactivity will have fallen to one-quarter of its initial value. Any organic material, and particularly wood, thus carries the means within itself by which it can be dated within the interval from about 1,000 to about 25,000 years ago.

There are various assumptions in the method. One is that there has been no long-term variation in the intensity of cosmic radiation, since this would affect the initial level of radioactivity. Another is that the radioactivity of the specimen, as now determined, represents the surviving activity from its initial state; in other words, there must have been no contamination of the specimen with carbon of more recent 'origin'. It has therefore been important that the method should be adequately tested on specimens of known date, before using it on unknown specimens as a method of dating.

The first such test reported was on wood from the tombs of two Egyptian Pharaohs, Zoser and Sneferu, dated to about 4,600 years ago. One more recent was on part of the cedar of Lebanon used as deckwood in the funerary boat of another Pharaoh, Sesostri III. This was supplied by the Chicago Natural History Museum, its age being given as 3,792 years. A further test was on fragments of pine from the floor of a Syro-Hittite palace, with a presumed age of 2,624 years. Professor Libby's measurements gave dates which agreed in each case with those quoted, within the probable error of his method. This was estimated at 450 years, 400 years and 150 years in these three cases. In addition, there was one earlier and less accurate estimate - which also agreed within the probable error - and as a check of a different kind heart-wood has been measured from a giant sequoia tree which was felled in 1874 and, from the tree ring method, is given an age at the time of measurement of 2,928 years. This, too, showed good agreement. With this encouragement, the use of the

method is now being extended to more remote periods for which dating by other methods is not available.

A considerable number of specimens have already been examined. The results have either been summarised by Professor Libby at informal meetings attended by archaeologists and others interested, or supplied as provisional values to those whose specimens have been accepted for testing. Few have yet been published or supplied for publication. Among them are wood from a pre-pottery site in Mesopotamia (age, 6,900 years); from a campfire site in Nebraska, 40 ft. below present ground level (age, 8,500 years); and from the period corresponding with the last advance of the ice in Wisconsin (age, 11,500 years). These are preliminary figures. A single English specimen has also been dated from the period following the last retreat of the ice, but has not been published at the time of writing. A limitation to the method is that 2 oz. weight of wood or charcoal are at present required. A point in its favour is that, since it is the *disappearance* of radioactivity which is measured, contamination of whatever kind must lead to under-estimates of age and not to over-estimates.

The second method to be described is based on the examination of cores obtained from ocean-bed sediments. It is a comparative method, as it stands, enabling climatic changes to be correlated as between one part of the world and another; but is capable of being linked to another form of radioactivity determination, which would give values in years to these comparative records of climatic change. The periods covered are of the order of a million years.

The method consists in doing sample counts, at different depths in these ocean-bed cores, of the different species of Foraminifera which they contain. These are small single-celled organisms which live in the surface layers of the oceans, and then sink to the bottom where their 'tests' or skeletons accumulate as sediment. They have remained substantially unchanged over long ages of geological time, and the present distribution of species can be readily correlated with the temperature of the water in which they live. This has been done lately by Mr C. D. Ovey of the British Museum. He has also applied statistical tests

to determine the number of Foraminifera which should be counted in each 'population' if reliable indications of climate are to be obtained. The answer is that not fewer than 1,000 should be counted in each case. The method is therefore relatively laborious, and the more so as it requires to be applied to 'cores' obtained from as many different parts of the oceans as possible. When this has been done, it will become possible to obtain a complete picture of climatic changes over the world as a whole; and to connect up the sequence of European ice-ages, which is comparatively well worked out, with corresponding changes in other continents. Finally, in conjunction with the radioactivity method already mentioned, it should be possible to attach approximate dates to all stages of this world climatic sequence. This, if accomplished, would effectively bridge the gap between the radio-carbon method, already mentioned, and the older uranium-lead method which covers the period from the oldest rocks down to a few million years ago.

Arnold and Libby: *Science*, 1949, 110, 678.

Wiseman and Ovey: *Proceedings of Geologists' Association*, 1950, 61, 28.

Cornwall: *Archaeological News Letter*, 1950, 2, 177.

WORLD GRAVITY SURVEY

One of the less obvious gains from the use of air transport in research is that the great reduction in travel time which is thus made possible may alter the character of the observations which are possible with a particular instrument. This has happened lately with comparative measurements of the earth's gravity. These can be made by two different methods, and for different purposes. The older and more laborious is to record the time of swing of a pendulum of known length, or to compare the rate of swing of the same pendulum in various positions. It was by this method – through the variation in time-keeping of a pendulum clock – that the French astronomer, Jean Richer, discovered in 1672 that there was a variation in the strength of gravity as between Paris and French Guiana. The difference amounted to 2 minutes 28 seconds per day, and it was interpreted by Newton

as being due, in part, to the bulging of the earth round the equator due to its spin. Such measurements, carried out on a world-wide basis, provide one method by which information on the shape or 'figure' of the earth can be obtained. This, in turn, is of practical importance for the linking up of surveys, made for ordinary mapping purposes, in different parts of the world. However, the pendulum method of comparison is clearly inconvenient, and the number of reliable interconnections which have been established in this way is therefore limited.

The second method is equivalent in principle to measuring the extension of a spring from which a constant mass is suspended. It is both quicker than the pendulum method and can be made more sensitive. It is the method used in the local type of gravity survey which is now widely employed in prospecting for oil. The interest, for this purpose, is in relatively small and local differences in the strength of gravity, due to underlying rock formations, the absolute accuracy of the measurement being of small importance. Instruments of this type are known as 'gravimeters'. Readings with them can be taken in five minutes with a scale accuracy of 1 part in 100 million. But until 1947 there was no such instrument which was adequately compensated against temperature changes; and, even when this difficulty had been overcome, there remained the further uncertainty that any instrument which relies on springs for its accuracy of measurement is likely to show a slow drift with time owing to changing mechanical properties.

The new development is the recent demonstration by Dr G. P. Woollard, of the Woods Hole Oceanographic Institution, Massachusetts, that given fast enough transport an instrument of the 'gravimeter' or spring-balance type can be successfully adapted to the needs of a world gravity survey. The principle is merely to measure the rate of drift of the instrument immediately before each journey, and again immediately afterwards. The average of the two can then be applied to the journey itself, provided that this is short in time. Preliminary trials were carried out during July and August, 1948, with the co-operation of the Geophysics Branch of the U.S. Office of Naval Research.

Observations were made in Washington, San Francisco and Hawaii; and, the results being encouraging, a further world flight was undertaken in September and October of the same year. Comparisons were made at 33 points for which pendulum observations were available, including the primary 'pendulum stations' of 11 countries.

The results have provided rather good evidence for the validity of the method – provided that it is checked, at all stages, by as many pendulum comparisons as may be already available. The evidence presented is of two kinds. First, as between the National Bureau of Standards in Washington, and the National Physical Laboratory at Teddington in England, there is already available a comparison by the pendulum method carried out in 1940. Four separate comparisons, using different flight routes, were made by the gravimeter method between the same two stations. Their average departure between the mean value of the four was less than 1 part in 4 million; and there was agreement with the pendulum method to within 1 part in 2 million. The second method is based merely on internal consistency. Returning to the same station, after whatever intermediate visits, the same value for local gravity should be obtained. Judged by this test, the worst 'loop' was one from Tripoli back to Tripoli, with intermediate observations at Khartoum, in Egypt, and in Saudi Arabia. The accumulated error on this trip amounted to a little less than 1 part in 1 million.

There is no suggestion that the new method has on this account superseded the pendulum as a source of information. Its value seems rather in the quick linking up of gravity surveys which have not before been reliably connected, and in the extension of gravity measurements to parts of the world which have not before been covered.

Woollard. *Geophysics*, 1950, 15, 1.

THE EARTH AS A TIME KEEPER

Gravity surveys assume a distribution of the earth's mass which is fixed and constant. Evidence that this is not strictly the case has

lately been obtained by H. F. Finch from analysis of the apparent performance of about a dozen of the best quartz crystal clocks used in the Greenwich time service during the period 1943 to 1949. The effect is to show a small seasonal variation in the earth's time-keeping, judged by absolute standards. The overall variation is about 0.12 seconds, and is not quite symmetrical in its distribution. If the earth and a quartz clock are regarded as starting together in February, then by about the beginning of June, the earth would be 'slow' by about 0.06 seconds. From then on, the earth would gain, and by October – about 4 months later – would be fast to the same extent. Such an effect had been suspected for some years from the work of N. Stoyko in Paris, but rather as something probable than as a fact which must be accepted. It implies an annual transfer of mass relative to the axis about which the earth rotates. There is a well-known lecture demonstration which illustrates the principle. The apparatus required is merely a light turn-table, and a human subject who will allow himself to be spun on it. Starting with his arms at his side, he is told to raise them to shoulder height at full extension, and then to lower them again. The effect of the first movement is to slow down his rotation on the turn-table; that of the second to speed it up again. The most likely causes of similar variations in the case of the earth are, first, a seasonal movement of air as between polar regions and equatorial; and, second, the melting of ice over the Antarctic continent during the southern summer. Meteorological estimates, made in Belgium, are stated to be in fair agreement with the observed change in the earth's rotation.

It may be noted that, as time measurement is now defined, it would be the quartz clocks which were in error and not the earth: for it is the earth which provides the standard. This is the time which is needed for most purposes. But in the most accurate laboratory work, the difference could be significant – as indeed follows from the fact that we have already time-keepers capable of showing it.

J. Brit. Astronom. Ass., 1950, 60, 120.

CLIMATIC CHANGE

A discussion on climatic change, including the suggested current climatic change, was included in the centenary celebrations of the Royal Meteorological Society, held in Oxford and London earlier in the present year. The broad fact that the climate of temperate regions in the northern hemisphere has been showing a tendency to become warmer during recent years has already become familiar. We have been told, among other effects, of retreating glaciers and of changes in the summer feeding grounds of North Sea fish. There is also other evidence from a slight apparent rise in sea level over a period of years. However, there are two main scientific questions which have remained unanswered. The first is whether this present change, which so far as it goes is undoubted, is to be regarded as the beginning of a long-term climatic change, or merely as a short-term fluctuation which is likely to be reversed within comparatively few years. This can hardly be answered except by time. The second is whether the change applies to the earth as a whole; and, in particular, whether it includes the Antarctic continent, where the progressive reduction in the ice cap might be expected to have an accelerating effect on the climatic change, of which it was itself one symptom. The results of a new analysis, bearing on this second question, were presented by Professor H. C. Willett of the Massachusetts Institute of Technology during the discussion already mentioned. His figures were based on average winter and average annual temperatures grouped in 10 degree zones of latitude, and taken as far as possible from stations for which continuous 50-year records are available. He placed the main emphasis on a progressive gradation in the extent, and even the direction, of change from the North Polar region southwards. As an indication of scale, he estimated that since 1885 there had been an average rise in world temperature of about 0.5°C , and in world winter temperature of about 1.2°C . He described the change as 'rapid' within the Arctic during the last 20 years, and as 'quite rapid' within the zone extending from 60° to 70°N . Further south, across the equator and down to about 30°S , he

found the trend to warmer temperature to be still continued, but less strongly. On the other hand, south of about 40°S , his figures showed the trend, if any, to have been in the reverse direction; with the greatest apparent rise shown by a single station in the Weddell Sea, representing the Antarctic. It was pointed out in informal discussion that the latter might be affected by changes in the local distribution of sea ice. But the broad effect was to leave the position that, whereas evidence for an improvement in climate in the northern hemisphere was soundly based, there was at least no positive evidence for such a trend in the case of the Southern hemisphere as a whole.

LOW TEMPERATURE PHYSICS (See Inset, No. 7-13)

It has been the pride of low-temperature physics to have discovered a peculiar world of its own, of which explanations have remained uncertain except in general terms. The first peculiarity to be discovered was the total disappearance of electrical resistance in metals, so that a current, once started, in general persists indefinitely. A second, discovered shortly before 1939, is the existence in helium II of a liquid so unusual in its behaviour as to justify the description of matter in a 'fourth state' – neither solid, ordinary liquid, nor gas. Its properties were described in an article published in *Science News* 10. It is formed by transition from normal liquid helium (helium I) at a temperature 2.19° above the absolute zero of temperature (about -273°C). It cannot be frozen unless compressed. It climbs as a thin film out of beakers – the property of 'superfluidity'. It is a superconductor of heat, so that the slightest local heating causes a spectacular 'fountain' effect through fine capillaries or their equivalent (Inset, No. 11). And heat, instead of spreading through it normally, is transmitted as a form of wave disturbance – 'second sound' – of which the speed can be measured. There is no further connection with ordinary sound than that this, too, is a wave disturbance in a material.

Further measurements of the speed of 'second sound' have been made lately at lower temperatures than before by J. R.

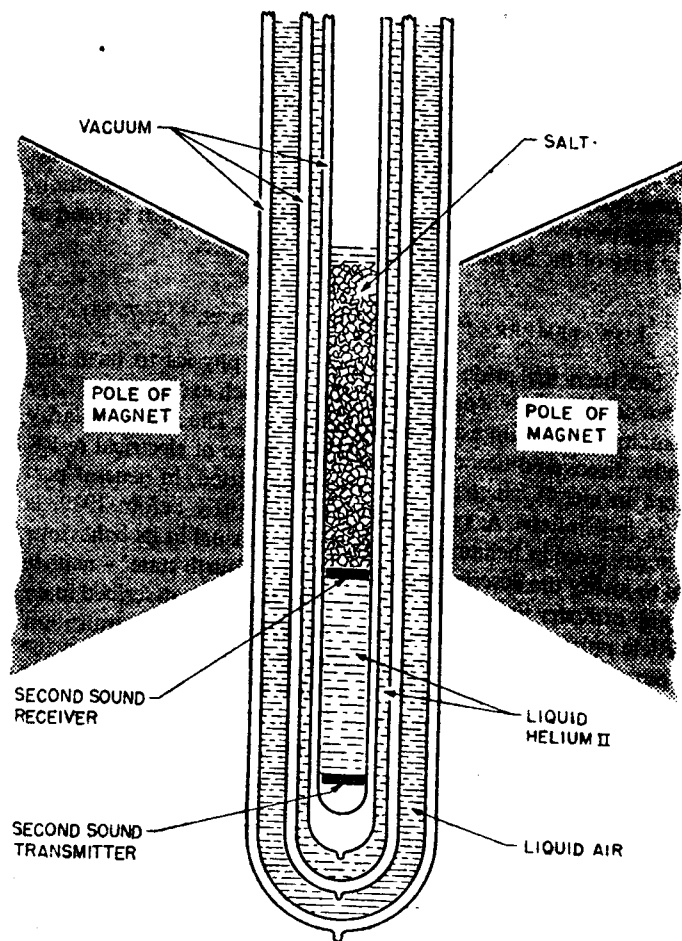


Fig. 27. The apparatus for measuring second sound in liquid helium II (see also Inset, No. 13).

Pellam and R. B. Scott, of the National Bureau of Standards, Washington. They are of interest, both as providing one more effect to be explained and, as the Bureau suggests, in relation to existing theories. Taken by themselves, they support one theory in particular – that put forward by Professor L. Landau, of the Institute for Physical Problems of the Academy of Sciences of the U.S.S.R., in 1941. A more conservative view is that the theoretical position is still very open, and that the main value of the Bureau's work is that it opens up new lines for experiment.

The earliest theory of helium II was suggested by Professor F. London, now of Duke University, North Carolina, in 1938. His proposal was that it should be treated as a form of simplified gas, for which the necessary mathematics had already been worked out. This idea was developed by Professor Laszlo Tisza, of the Massachusetts Institute of Technology, in the light of known effects, and was used by him to predict a further effect – that of 'second sound' – which had not at the time been discovered. The characteristic of this theory is that it applies only to helium atoms of mass 4 – which form the great bulk of natural helium – and not to helium atoms of mass 3 which are present to the extent of about 1 in 10,000.

The significance of this limitation is that experiments have lately become possible, both with small quantities of pure helium-3 prepared by nuclear transmutation, in the first instance in an atomic pile, and also with mixtures of helium-3 and helium-4 in which the normal proportion of helium-3 has been many times increased. Although not yet conclusive, the broad effect of these experiments has been to strengthen the London-Tisza theory – at least to the extent of suggesting that the peculiar properties of liquid helium are characteristic of helium-4 atoms. On the other hand, there has not yet been enough helium-3 for the only final test – which would be a full investigation of the properties of pure liquid helium-3 down to the lowest temperatures attainable. Meantime, as already mentioned, an alternative theory had been developed by Professor L. Landau in Russia. This was based on the application of quantum theory to the movement of liquids, and in

particular to the travel of wave disturbances. Like Tisza's theory, it drew in part on experiment. Also like Tisza's theory, it predicted the effect of 'second sound', which was confirmed by a Russian experimental physicist, V. Peshkov, towards the end of the war. The position up to this stage was thus that either of these theories could claim support from experiment. But Landau's theory, unlike Tisza's, drew no distinction in kind between the behaviour to be expected from helium-4 atoms and those of helium-3. Finally, a more general application of quantum theory to liquids has been made by Professor Max Born and Dr H. S. Green at Edinburgh, in which the peculiarities of liquid helium are derived as a special case. As in Landau's theory, there is no distinction in kind between helium-4 and helium-3 atoms; but there is no connection, otherwise, between these theories.

The Bureau of Standards measurements (see Inset, Nos. 12 and 13) were undertaken as a test of predictions made specifically by the Tisza and Landau theories in the region of temperature below about 1.3° Absolute. At higher temperatures – between 2.19° and 1.3° Absolute – the predictions of 'second sound' given by these two theories are indistinguishable by experiment. Below 1.3° , exactly opposite effects had been predicted: by Tisza, a progressive decrease in speed, but by Landau a sudden rise in the rate of transmission. The Bureau of Standards measurements show a minimum speed of transmission of 60.4 ft per second at about 1.1° Absolute, rising smoothly to 111 ft. per second at a temperature between 0.5° and 0.7° Absolute, the lowest temperature at which measurements of the speed of 'second sound' have yet been published. On the particular experiment there has thus been a direct confirmation of Landau's prediction, as against Tisza's, and these facts must clearly be covered in any complete theory of low-temperature effects. But in relation to the general background which has been given it still seems possible, and even probable, that there is no single theory which can be regarded as established.

CHEMISTRY AND HEREDITY

Apart from pigments, it is doubtful if any simply inherited quality has been connected up to now with a corresponding chemical difference which can be detected chemically. Such a connection is strongly suggested by the work of Professor Linus Pauling and others, at the California Institute of Technology, on a blood disease of Negro races known as sickle cell anaemia. This is a disease which has been known since 1917. It shows itself in two grades – the one a harmless abnormality; the other, a chronic anaemia severe enough to prevent many who show it from surviving to maturity. The abnormality is found in about 8 per cent of American Negroes, and in higher proportions in African Negroes, ranging up to about 19 per cent in Nigeria and the Cameroons. It shows itself as a change in the shape of the red blood cells, or a proportion of them, when their oxygen supply is reduced. In place of the usual disc shape, thinnest in the middle, there is a variety of shapes of which the sickle or crescent shape gives the condition its name. This can be demonstrated experimentally by taking a drop of blood, sealing it with vaseline under a cover to keep out oxygen, and then watching for abnormal cell shapes at intervals up to 72 hours. Under bodily conditions, there is sufficient deprivation of oxygen in the return circulation through the veins to cause about 1 per cent of 'sickling' in those with the abnormality. In those with the disease proper, the proportion in the veins reaches 30-60 per cent. The red cells have also a shorter life than usual in those with the disease, and this is the immediate cause of the anaemia. But apart from this, the essential difference is the extent of oxygen deprivation needed to produce a given proportion of sickling. This more drastic condition has been reported in the proportion of 1 case with the disease to every 40 who show the abnormality; but, because of early deaths, it is likely that the proportion has been underestimated. The genetics of the disease have been studied lately by Dr J. V. Neel of the University of Michigan. From 13 cases of the disease, in which both parents showed the abnormality, he concludes that both effects are due to a single gene which,

if inherited from one parent only, shows itself in the abnormality, and if from both in the disease.

Professor Pauling and his colleagues in California have reached the same conclusion independently, but from chemical evidence. Their method has been to prepare solutions of haemoglobin from the red cells of normal individuals, and from those with the abnormality and with the disease; and to apply to these solutions a technique of electrical separation introduced by Professor A. Tiselius of Uppsala. The result has been to establish a difference between the haemoglobins of normal individuals and those with the disease, that of individuals with the abnormality appearing as a mixture of the two. The indications are that both types of haemoglobin exist together in individual red cells of those with the abnormality, rather than some cells containing only one type and some the other.

Pauling, Itano, Singer and Wells. *Science*, 1949, 110, 543.
Neel. *Science*, 1949, 110, 64.

INSECT TRANSPORT

From an early stage in the history of air transport, the possibility was recognised that disease-carrying insects might be conveyed from one country to another in this way, one of the more alarming prospects being that yellow fever might be introduced from Africa into India. Because of the possibility of insect carriage, the spraying of long-distance aircraft has become a routine measure and, in relation to the size of the problem, seems on the whole to have achieved its purpose. None the less, there are occasional examples reported of unwanted introductions. A recent case was the introduction of the common wasp into New Zealand. This is believed to have been by air in 1945 through an aerodrome near Hamilton in the Waikato district of the North Island. In spite of an organised campaign, using both cyanide and D.D.T., the wasps had established themselves within four years over a 200 miles range from north to south, and of more than 150 miles from east to west. At this stage, it is unlikely that eradication will be possible, and it will be a question of

keeping the wasps away for as long as possible from the main fruit-growing district of the South Island.

A transport problem of another kind has been encountered in East Africa, where the spread of tsetse flies by railway trains has been studied on the main line between Mombasa and Nairobi by E. Aneurin Lewis, Chief Field Zoologist of Kenya Colony. The problem is an old one in the sense that, whereas the railway was completed some 50 years ago, references to the conveyance of tsetse by trains go back to 1910. It is new in the sense that the extent of fly movement from this cause has for the first time been studied numerically, and that the control problem presented must apply with variations to many other areas of tropical Africa now being opened up by road or rail transport. It is not a question of extending the range of the tsetse in the ordinary sense – for, where the country suits them, they are already in occupation – but of the periodic introduction both of the flies and of the livestock diseases which they carry to stock farming areas outside the 'natural' limit of the fly belt. The course of the railway is unfortunately well suited to such movements. It is 330 miles in length, and in its middle section traverses the main tsetse belt of the colony. There is thus the possibility of the conveyance of the fly in either direction, and both up and down trains were examined during a 12 months period during their passage through at least two stations, near the beginning and end of the tsetse belt. On passenger trains, a total of 5,671 flies were caught, of which 228 contained viable pupae ready to be deposited. For goods trains, the total was 20,709. With the daily train length soon to be increased to 1,700 ft., it is obvious that their proposed 'deffying' would demand an expenditure of time and insecticide out of all proportion to that needed in the case of air transport. It is proposed that two stations should be established for this purpose, and preliminary evidence suggests that there is an equal case for the treatment of road vehicles.

Finally, an interesting case has been reported from Zanzibar of the transport of one insect by another. The suggestion, as yet unproved, is that a virus disease of cloves is carried by a species of scale insect which in turn is dependent on a red tree ant for its

protection and spreading. The ants forage over the clove tree in which they nest and on the surrounding ground, and guard their territory. They carry insects of many kinds back to the nest. In particular, they 'cultivate' the scale insect. Not only does this reach enormous numbers in the nest, but also strong infestations are established in the young terminal twigs. These external colonies are invested in fine silken tents, thought to protect the scale insects against at least two parasites. Scale insects and ants thus move together - with the pace, which is a slow one, set by the ants. The problem has been studied by Dr F. J. Nutman and Dr F. M. L. Sheffield for the Clove Research Scheme, Zanzibar, Dr Sheffield having been seconded for the purpose from the Rothamsted Experimental Station, Harpenden. The picture which they suggest fits well with both a sporadic form of outbreak, in which only a few trees are killed, after which the outbreak dies out; and also with the usual form in which spreading continues more or less regularly from an initial centre. Within the past 10-12 years, it is believed that at least half the clove trees on Zanzibar have died from the disease, the neighbouring island of Pemba being also affected.

Tsetse: Lewis. *Bulletin of Economic Entomology*, 1950, 40, 511.

Cloves: Nutman and Sheffield. *Annals of Applied Biology*, 1949, 36, 419.

SENSORY PITS

Wartime interest in the detection of infra-red radiation, as a means of night observation, has had repercussions already on biological research. Its most obvious application is as a means of observation. But it has led also to some new thinking on the interpretation of sensory organs. Before the war, for example, it had been shown that the facial pit of the pit vipers (*Crotalinae*) was a heat-sensitive organ enabling them to detect warm-blooded animals in the dark. More lately, M. J. Block of the Engineer Research and Development Laboratories, Fort Belvoir, Virginia, has pointed out the close similarity of these pits to the pneumatic type of radiant energy detector in which there is an

outer cavity which serves as a 'collector' of radiation; and, separated from this by a membrane, an air space which expands when heated, as when radiation is absorbed. It is pointed out that the 'pit' organ, too, is provided with a membrane and inner cavity, and that it is sensitive to gusts of air as well as temperature, which suggests that it is the movement of the membrane which gives the sensory response.

Independently, and some months earlier, it has been suggested by G. R. M. Grant of the University of Queensland that the sensory pits of various insects, including bees, may act also as receptors of infra-red radiation. Physical analogies are drawn in this case from radio, with the result that the interpretation offered is in terms of resonant frequencies. If this is correct, it might be possible to confirm it. On the other hand, it would be simpler and perhaps more plausible if the same type of explanation could be adapted to the two cases.

Block. *Nature*, 1950, 165, 284.

Grant. *Proc. Roy. Soc. Queensland*, 1949, 60, 89.

YOUR GALVANIZED IRON ROOF

D. J. SWAINE

INTRODUCTION

FOR more than a hundred years galvanized iron sheets have been used as the roofing material for countless houses in every part of the globe. It is not generally realised that approximately half of the world production of zinc is used for coating steel. How is this zinc applied to the steel sheet, why does it protect the underlying steel and what happens during weathering? This and other questions will be discussed now.

THE MANUFACTURE OF GALVANIZED SHEETS

There are a number of methods of applying zinc to steel, but since most sheets are coated by the hot-galvanizing process, this one will be considered. The basic principle of this 'hot-dip' method has not changed since the early 19th century, although various modifications have been introduced in the interim. The sheet (formerly wrought iron, but now plain carbon steel) is cleaned by pickling it in dilute sulphuric or hydrochloric acid or by sand blasting. It is then passed through a layer of zinc-ammonium chloride flux and into a bath of molten zinc. The sheet emerges with its coat of zinc and the familiar 'spangled' appearance. These 'spangles' have no practical value, although some buyers seem to associate them with good quality. Finally the sheet is corrugated.

THE STRUCTURE OF THE COATING

If one cuts a piece out of a galvanized sheet (made by the 'hot-dip' process) and examines it under a microscope at right angles to the zinc surface, the structure of the coating is seen (Fig. 28). Between the outer zinc layer and the underlying steel sheet there

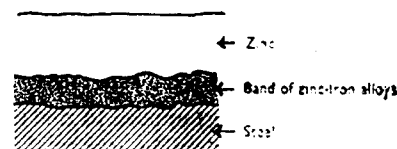


Fig. 28. Structure of galvanized sheet coating

is a band of zinc-iron alloys of complex composition. This is not surprising, since some reaction between zinc and iron, as well as diffusion, would be expected to occur at the temperature of galvanizing (about 450°C). The nature of these alloys has been in doubt for some time, but it has now been established by X-ray analysis that three alloys are present, namely, $\text{Fe}_2\text{Zn}_{11}$, FeZn_7 , and FeZn_{13} . If a small amount of aluminium (0.15 – 0.50 per cent) is added to the galvanizing bath, the resulting zinc coating has good bending properties. This is because the brittle zinc-iron alloy layer is not formed under these conditions.

THE COMPOSITION OF THE COATING

Besides zinc, the coating includes the zinc-iron alloys, and there may be some impurities present also, e.g., iron, lead and tin. It is not surprising that a small amount of iron occurs in the coating, since the base metal is steel, and the pot in which the sheet is galvanized is also made of steel. Lead and tin are added to the galvanizing bath, the former to collect dross (insoluble zinc-iron compounds) and to protect the pot from overheating, and the latter to promote the formation of 'spangles'. The analysis of the coating of a modern galvanized sheet gave the following results:

zinc	95.8 per cent
iron	3.0 „ „
lead	0.9 „ „
tin	0.3 „ „

There is disagreement as to the effect of these impurities in the zinc coating on its properties of resisting attack by the weather. It would seem that they only have a minor effect.

THE DETERMINATION OF THE COATING WEIGHT

Probably the most important test applied to zinc coatings is the determination of the coating weight, i.e., the total weight of zinc per unit area. The method laid down in British and American Standards consists in treating a weighed, measured piece of galvanized sheet with hydrochloric acid containing some antimony trichloride (this minimises the attack of the acid on the steel base material). Then the specimen is washed, dried and reweighed, the difference in weight being the weight of zinc in the coating. A roofing sheet might have a coating weight of about 2 oz. per square foot of sheet, which means that the zinc layer is nearly two thousandths of an inch thick on each side of the sheet.

CORROSION PRODUCTS

When the galvanized sheet is exposed to the weather, chemical reactions will take place between the zinc coating and water, oxygen, carbon dioxide, sulphur dioxide or other substances that may be present, resulting in the formation of whitish material on the sheet. Hence, the nature of the corrosion products, which are the substances formed by these weathering reactions, is dependent on the atmospheric conditions. For example, water, oxygen and carbon dioxide are present in the atmosphere and result in the formation of zinc oxide and basic zinc carbonate on the galvanized sheet. In addition, in industrial areas, where the air contains sulphur dioxide, zinc sulphate may be formed. The mode of formation and the exact nature of the corrosion products are not definitely settled, but X-ray diffraction analysis has shown that zinc oxide, zinc carbonate, basic zinc carbonate and other basic zinc salts are present in the corrosion products on a galvanized sheet.

It has been suggested that a film of zinc oxide is formed on the sheet as it leaves the molten zinc bath after being coated with

zinc. This film gives the sheet some protection, but eventually water and carbon dioxide act on the zinc oxide, causing the formation of basic zinc carbonate (Fig. 29).

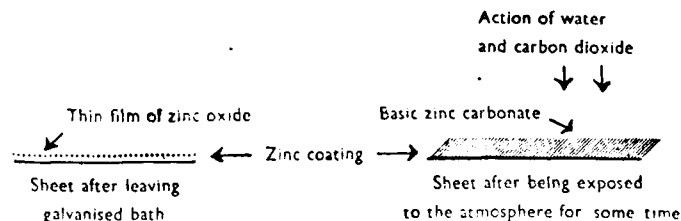


Fig. 29.

Since a zinc oxide film cannot form zinc hydroxide, the older theory, that zinc hydroxide is formed initially and slowly changed into zinc oxide and basic carbonate, is no longer tenable.

The corrosion rate for city air is about three times, for sea air two to four times, and for industrial air about six times that for rural air. This is to be expected, since sea air contains some salt spray and industrial air has a relatively high concentration of sulphur dioxide and carbon dioxide. It has been shown that the corrosion rate is greater during the night than during the day, which is probably due to the condensation of dew, which has a relatively high concentration of industrial gases.

HOW A ZINC COATING PROTECTS STEEL UNDER ATMOSPHERIC CONDITIONS

So far, no reason has been given for the choice of zinc as the coating material for the sheet. The answer to this question will be found in the following discussion. One must consider two conditions:

- (a) while the surface of the steel is completely covered with zinc or zinc-iron alloys;
- (b) when the underlying steel is exposed, as occurs when the galvanized sheet is deeply scratched, thereby removing some of the zinc near the scratch.

In case (a), the zinc gives the underlying steel mechanical protection, which is good, because zinc (under normal atmospheric conditions) is about fifteen times less corrodible than iron.

In case (b), the sheet is also protected to a large extent, because zinc has a greater tendency to go into solution than iron has, when the zinc and iron are in contact.

Hence the steel is protected at the expense of the zinc, which forms the corrosion products.

CONCLUSION

Summarising this simplified account, one sees that the humble galvanized sheet has quite an interesting life history. The steel with its protective zinc coating is placed in position on a roof. Slowly the weather causes the formation of corrosion products on the sheet, and eventually the steel is exposed. One very important factor has been omitted, namely, the use of paint as a further protection. Occasional painting of galvanized sheets will prolong their life and give them a more pleasing appearance. But this is another story ...

HYDRAULIC MODELS

BASIL STANDING

By a 'model' is generally understood a reproduction of some original device, the dimensions in the reproduction being obtained by decreasing the dimensions of the original, or prototype, in some definite proportion. Thus there are the scale-model locomotives beloved of schoolboys, the model steamships, and, more recently, the model, or midget, racing cars, which are run in a circular path and tethered to the centre during the run. Note, however, that various qualities of the performance of the original are not reduced in the same ratio as the dimensions. The midget racing cars are approximately 1/20 of the size of the average full-scale racing car, but the speed of the two is roughly the same.

Hydraulic models are used to give information, at best quantitative and at worst approximately qualitative, about some original phenomenon of water flow, and as such are generally models in which the reduction of all the dimensions in a strict ratio is not of primary importance. What is more important is that a predetermined ratio should exist between certain *qualities of performance* in model and prototype, so that the effects in the model should be very closely equivalent to those in nature. To put it another way, the model is to be similar, in a dynamical sense, not in a geometrical sense, to its prototype. For this similarity the necessary conditions are:

- (1) The paths followed by corresponding particles shall be geometrically similar, and
- (2) corresponding forces shall bear a constant ratio to one another.

Thus if the force due to surface tension in the model is 1/5 of that in the prototype, then the force in the model due to say viscous resistance must also be 1/5 of its counterpart in the

original. The speeds of flow in model and prototype at which this constancy of force-ratio holds, for all the various forces involved in the phenomenon, are called the 'corresponding' speeds, but it is not generally possible to arrange the conditions of operation of the model so that the ratio is identical for all the sets of forces. In other words, there is no single 'corresponding' speed.

All models are concerned with resistances to flow, and two particular kinds of force are of primary importance:

- (a) the viscous forces, or forces arising from the viscosity* of the liquid, which are most important at low speeds when the flow is slow or sluggish, and
- (b) the gravity forces, which are important at higher speeds.

A theoretical analysis shows that dynamical similarity between the viscous forces is obtained when

EQUATION 1

$$\frac{v_m l_m}{V_m} = \frac{v_p l_p}{V_p} \quad \dagger$$

where v_m , v_p are the corresponding speeds in model and prototype, l_m and l_p are the values of some characteristic dimension in the system, e.g. the diameter of the pipe when flow in a pipe is being considered, or the depth of flow in an open channel, and V_m and V_p are the kinematic viscosities (or viscosities divided by the densities) of the fluids in the two systems. Condition (b)

* A scientific definition of viscosity is no help in understanding how these forces arise, but if the reader can imagine the flow of a very viscous liquid, e.g. syrup, and mentally compare it with the flow of a less viscous or 'thinner' fluid, e.g. water, the influence of the viscosity on the flow may be realised.

† The quantity $v l / V$ is called the Reynolds Number, after Professor Osborne Reynolds. The reader with some scientific knowledge will realise that it is dimensionless.

demands that

EQUATION 2

$$\frac{g l_m}{v_m^3} = \frac{g l_p}{v_p^3}$$

where g is the gravitational acceleration. Thus, if both equations are to hold, then the fluids used must be such that

EQUATION 3

$$\left(\frac{V_m}{V_p}\right)^2 = \left(\frac{l_m}{l_p}\right)^3$$

Consideration of the space in the laboratory generally fixes the ratio $\frac{l_m}{l_p}$. For example, if the space available for a model of an estuary half a mile wide is 20 feet, the horizontal scale will be 1:132, or, as it is more convenient to work in round figures, 1:150. The model would then require to be tested using a fluid of which the kinematic viscosity was $1/\sqrt{3375000}$ of that in the prototype, a virtual impossibility. The ratio between the speeds, in this case the velocities of flow of the two liquids, would be $1/\sqrt{150}$, and complete dynamical similarity at this speed ratio would be obtained if the liquid possessing the required kinematic viscosity could be found. It might now be thought that the two conditions can never be fulfilled simultaneously and that therefore complete dynamical similarity is impossible. In practice, however, the problem is simpler than it looks at first sight. The two kinds of forces do not give rise to effects of the same magnitude. For low velocities and thick liquids the flow is generally such that the paths described by the various particles of liquid are straight lines or smoothly curved lines – the flow is viscous or streamline, and the viscous forces are responsible for the way in which the motion proceeds. For higher velocities the motion of the elements of fluid can now be described as haphazard, the motion is now called turbulent, and viscous forces produce negligible effects by comparison with those of the gravity forces. For the streamline type, only equation (1) need hold, and for the turbulent type, only equation (2). In general the motion of water in nature, in a

river for example, is turbulent, and equation (2) is easily satisfied and dynamical similarity attained. But the neglect of the viscous forces, in the determination of the velocity and scale ratios by equation (2), does not always mean that their effects are not present or can be neglected, and the conversion of the model results to those for the full-size system may not be just a mere increase by a known amount. In other words, the viscous forces may cause 'scale-effects' in the model which prevent a simple conversion of the model results into full-size results. One way in which the scale-effects may be investigated is by carrying out tests on models to various scales; e.g. the Upper Narrows Dam on the Yuba River in America was reproduced to scales of 1:25, 1:40, and 1:100. The information from the tests on a number of models, carried out at corresponding speeds, will then show whether scale-effects will have to be allowed for in determining the performance of the prototype.

Models of sluice-gates, pipe works and so on, are conveniently made to a definite scale in all three dimensions. In models of rivers or estuaries the vertical scale is generally bigger than the horizontal, or, more simply, the heights and depths in nature have been exaggerated by some definite amount in the model. The reasons for this are not easy to explain without mathematical reasoning, but in the first place it will be realized that if the horizontal scale is, say, of the order of 1:10,000, then to the same scale it will only be possible to work to 1/20 of an inch in reproducing some outstanding feature of the prototype, i.e. a feature of just over forty feet in height will be the smallest capable of reproduction in the model. The main reason, on theoretical grounds, for an exaggeration or distortion of the vertical scale, is that the exaggeration makes it possible for ordinary roughnesses of the sides and bed of the model water course to be sufficient to give the requisite scaled-down friction-loss, or loss of head due to friction, along the channel. Without the distortion, the sides and bottom of a model channel conveying water would, for scales of the order of 1:2,000 and above, have to be very highly polished, a condition which would be difficult to maintain in an apparatus using water, which easily collects a surface film of oil,

dust, etc. Another advantage of using a distorted scale, especially in a small model, is that the flow is more likely to be turbulent, and the effects of viscous forces less important, when the water channels are deeper, which is the direct result of the distortion. A final reason for using models having an exaggeration of vertical dimensions is that the bed-material is much more easily moved in such models than in models having identical vertical and horizontal scales. The theory underlying the movement of the bed-material is complicated and not yet in an exact form, and no experimental results of universal applicability are available. It can, however, be pointed out that in hydraulic models used to investigate a problem such as coast erosion or beach accumulation, the choice of bed-material to be used is the most difficult question that faces the investigator. It will be realised that it is not simply a matter of scaling down the size of the bed-material in nature, since a fine sand such as occurs in an estuary would scale down, at say 1:5,000, to a very fine dust that would behave most unsuitably in a model.

Turning aside now from the theoretical aspects, it is interesting to consider the progress made in the use of models since they were first applied to the solution of hydraulic problems. Professor A. H. Gibson ascribes the first use of a model to a M. Fargue in France in 1875, and points out that, although all conditions were arbitrary, i.e. no definite relationships held between the dimensions, rates of flow and times in the model, and those in nature, the model did give results of considerable practical value. The first real application of scientific principles to the design of hydraulic models was by Professor Osborne Reynolds in 1885. The models were tidal ones, and represented a part of the river Mersey. The scales were: horizontal, 1:31,800 and 1:10,600; and vertical, 1:960 and 1:396 - i.e. the vertical distortions were about 33 and 26.

The British Association, before whom Reynolds laid his results in 1887, appointed a committee to investigate the action of waves and currents on river estuaries, and Reynolds continued his work, which led to the committee reporting in 1891 that the use of models could be applied with confidence to the investiga-

tion of practical problems, provided that the fundamental theory was carefully considered. Reynolds himself considered that no works of any cost in a harbour or estuary, e.g. training walls or extensive dredging, should be carried out without a preliminary model investigation.

This work by Reynolds inspired the next important tidal model, one of the Seine estuary, carried out this time by a widely-experienced consulting engineer, Mr L. F. Vernon-Harcourt. The results were communicated to the Royal Society in 1888. One interesting feature was this investigator's examination of the suitabilities of quite a wide range of materials for simulating the movement of the bed of the estuary. Reynolds had used sand, but Vernon-Harcourt tried a number of sands, powdered pumice, flowers of sulphur, pulverised coke, fuller's earth, lupin seeds, coffee grains, fine sawdust and powdered Bath-brick. All these materials, other than the sand, possessed certain deficiencies and the best material was found to be a fine sand from Clapham Common. The next important model was one constructed by a Mr G. E. W. Cruttwell in 1903, to examine proposals for improving the navigable channels in the Thames. There is some doubt, however, as to whether the velocities used were large enough to cause movement of the bed-materials, and it is very probable that an increase of the vertical distortion would have remedied this deficiency.

Coming now to more recent experiments, an interesting investigation has been carried out as to the most suitable location for the dredged channels in the harbour at Bombay. A subsidiary use for the model was to investigate the effects of the discharge of raw sewage into the harbour, a scheme having been put forward which assumed that the whole area of the harbour, about 120 square miles, would serve to dilute the effluent enough. The model, however, showed that the sewage would tend to accumulate in a particular portion of the harbour, this result forming the basis of the successful opposition to the scheme by the harbour trustees. It is extremely interesting to note that at this time (1932) the opinions of most eminent civil engineers were divided as to the value of models.

Probably the most outstanding model investigation yet undertaken, and the largest attempted at its time in 1926, was the tidal model of the Severn estuary constructed by Professor A. H. Gibson at Manchester University. The scales were 1:8,500 (horizontal) and two vertical scales, 1:100 initially, and a later one of 1:200. The work was done on behalf of the Government, and the object of the tests was to examine the probable effects of a proposed tidal barrage on the estuary. This barrage was to utilise the tidal power available in the estuary, and the Severn is probably the best site for this in the whole of the British Isles, because of the great range of tide, there being approximately 40 feet difference in level between high and low water of an ordinary spring tide at Avonmouth. It is impossible to give even a brief summary of the vast amount of valuable data obtained during this investigation, which lasted over several years. The interested reader should refer to the two reports published. Since this time models have been made in almost all countries of the world, and there is no doubt that the conflict of opinion as to the value of model investigations has now finally disappeared. Civil engineers are unanimous that models can give at least qualitative results of great value, and there is the economic advantage that the relatively small cost of a model may be many times smaller than the cost of errors in works erected without previous model tests.

Models may be made of any hydraulic project, for example, those portions of a river which are non-tidal. One of the earliest of such models was a 1:250 scale model of a harbour near Vienna on the Danube, carried out at Dresden in 1898, and since then very many more have been constructed. The latest of these is one of the Mississippi begun in 1945. Small portions of this river, an extremely fruitful source of problems, with its unstable flow, had been constructed previously, but this new model is to embrace the whole of the river basin and its tributaries. The scales are 1:2,000 and 1:100, and the estimated cost is \$3,000,000. This work comes under the direction of the U.S. Waterways Experiment Station at Vicksburg, Mississippi, which is a branch of the Corps of Engineers, U.S. Army. A very complete investigation has been carried out at this establishment on the factors

affecting meandering of rivers, the results obtained being in some degree applicable to the control of the meanderings of the Mississippi, the largest example of such a phenomenon in the world.

Meandering is the name given to the state of affairs shown in Figs. 30a, b, c, where an initially straight length of river develops into a series of loops, which may or may not become cut off to form ox-bow lakes. The portion of the river Mersey (in Lancashire) above Flixton Bridge is a fine local example of meandering and cut-offs.

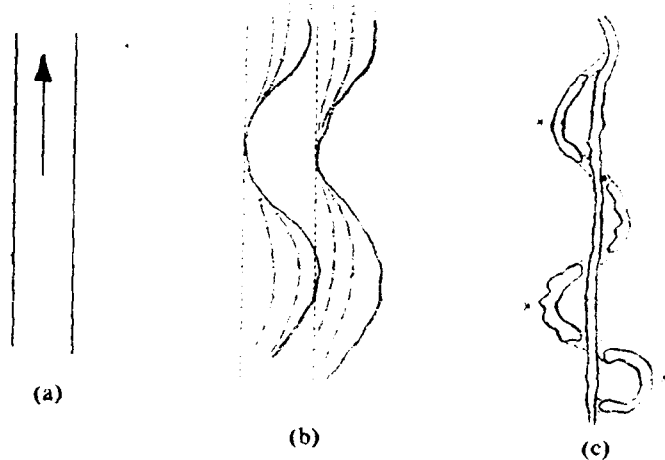


Fig. 30. (a) Straight flow (b) Successive stages in development of loops (1, 2, 3) (c) Final stage after loops have been cut off, forming ox-bow lakes X.

Weirs, spillways, sluice gates and dams have also been tested. One such investigation concerned the overflows from the Ladybower reservoir in Derbyshire, the form of which is shown, very diagrammatically, in Fig. 31. Here the best form of bell-mouth might be imagined to be one having smooth interior surfaces (a), but model tests carried out by Professor A. H. Gibson showed



Fig. 31. Shape for reservoir overflow

that the stepped form of surface shown at (b) was an improvement on the smooth surface from a hydraulic point of view, besides being of a form more easily constructed.

How are these models made? In the first place each is essentially a watertight box in which the estuary or other feature is moulded. The upper edge of the box is carefully levelled to provide a datum from which depths can be measured. The solid bed is moulded to what are called templates, Fig. 32 showing a typical arrangement, the material used being broken stone or

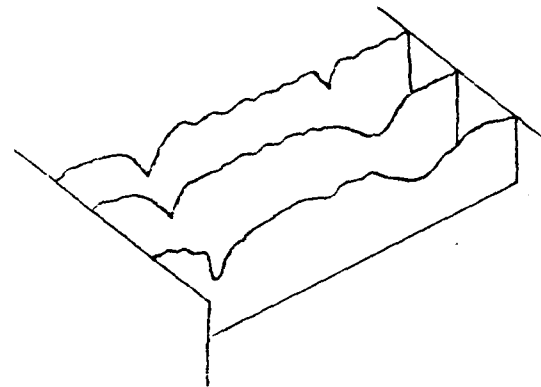


Fig. 32.

gravel, which is faced over finally with cement-sand mortar. The mobile material is now added and is moulded to other templates which rest on the upper edges of the model (Fig. 33).

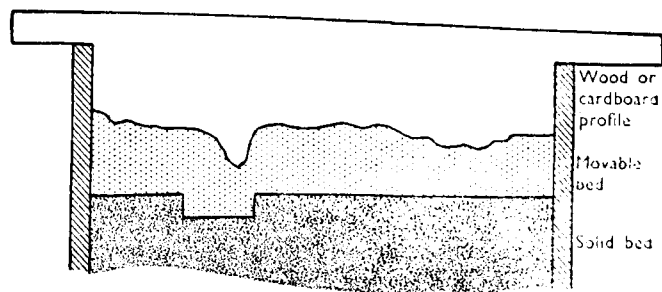


Fig. 33.

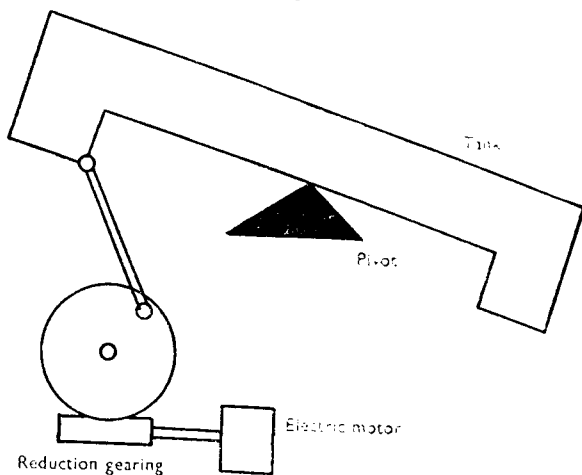


Fig. 34.

Tides are generated in a variety of ways. The simplest, but of use only on small models, is some mechanical device for tilting the model at definite times, shown diagrammatically in Fig. 34. For larger models an end tank is fitted to the model, and a plunger is arranged, by suitable mechanism, to descend into the tank, and by displacing water cause a flood tide. As the plunger is

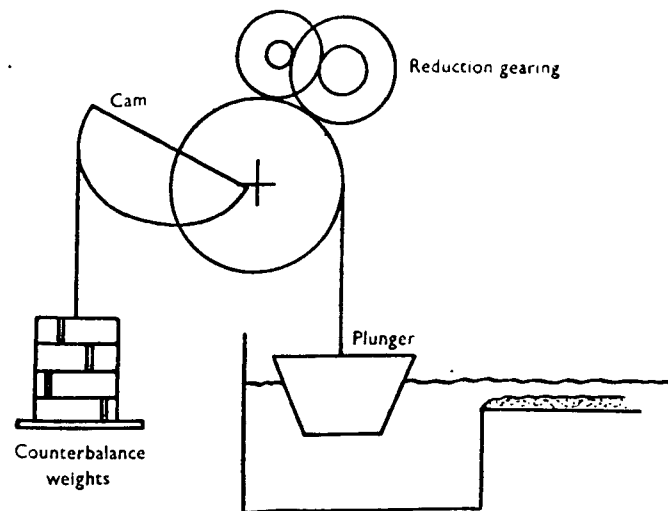


Fig. 35. The cam automatically adjusts the leverage of the counterbalance weights as the plunger enters the water and loses weight by immersion.

withdrawn an ebb tide results. The general arrangement is shown in Fig. 35.

The effect of strong winds on water surfaces may be reproduced by using small fans. Silt brought down into an estuary forms a special problem in itself. One method adopted has been to feed the silt in at the upstream end of the river by a belt conveyor. Another way is to provide enough material in the end tank, and to keep the water agitated so that it takes with it a certain percentage of silt, Fig. 36.

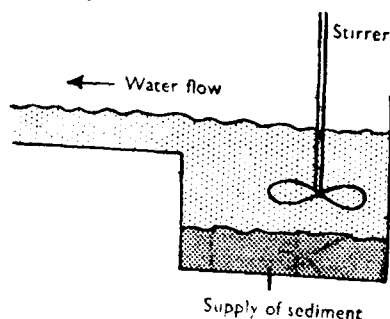


Fig. 36.

One action, which has to be allowed for in the determination of the size and quantity of silt to be used in the model, is the action of water containing dissolved salts, i.e. sea-water, in causing the silt to settle at a rate different from that in fresh water.

The above account is, of necessity, only a brief outline of some of the theoretical and practical aspects of hydraulic model investigations. For the scientifically qualified reader there is a book, *Scale Models in Hydraulic Engineering* by Professor J. Allen, which gives a complete account of model theory and technique, and which makes fascinating reading, even for the layman, in the accounts of various investigations. The layman will also find Professor Gibson's biography of Osborne Reynolds (*Science in Britain* series, published by Longmans) extremely interesting.

THE CARBON ARC

HUGH SODEN

How many of the millions of people who visit the cinema each week give a thought to the source of light which makes the projection of the film possible? – except of course when it goes out, to the accompaniment of derisive claps from the audience.

The universal source of light in the cinema is an arc struck between two carbon rods, which in most theatres are copper plated and about the diameter of a pencil. In the low intensity lamp used in some smaller cinemas the carbon rod is somewhat larger in diameter, shorter, and has no outer coating of copper.

MANUFACTURE OF ELECTRODES

The materials usually employed are lampblack and coke flour, mixed with a tar or pitch binder in heated mixers. The resulting material is forced by hydraulic pressure through dies of the required diameter.

The extruded rods are formed with a central hole for the insertion of a core as mentioned later. The 'green' rods are then bundled together and baked in gas or coal-fired ovens to over 1,000°C when the binding agent is converted into coke, leaving rods of homogeneous carbon which are then ready for the cutting and finishing processes.

Though to the uninitiated one carbon rod looks like another, the proportions of the ingredients and the size of the particles of the powders used are carefully controlled according to the use to which the resulting electrode is to be put.

PHYSICS OF THE ARC

To strike an arc two carbons connected to a source of current are brought together, end to end, to complete a circuit. They are then separated by a distance of a quarter of an inch to an inch.

At the instant of separation there is a high concentration of energy at the final point of contact, and this volatilizes the electrode material, thus providing a path of ionized vapour. The flow of current through this vapour heats the ions to a very high energy state, and this energy is continually dissipated as radiation. The radiation is not only visual but has important infra-red and ultra-violet constituents as well. Different types of radiation for different purposes are obtained by placing suitable materials in the central core.

TYPES OF ARC

The *low intensity* arc employs a carbon having a small neutral core. This contains a little alkali salt, and its function is merely to improve steadiness of operation. This type of arc is usually operated on direct current, and the gas column between the electrodes is only faintly luminous. Most of the light comes from the positive crater, - the end of the positive carbon (Fig. 37).

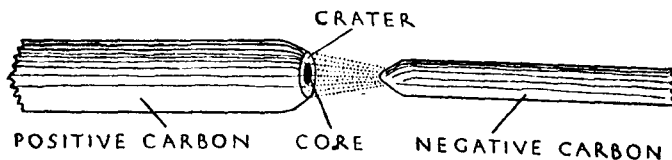


Fig. 37

If a low intensity arc is operated inside a glass globe with a restricted supply of air, the atmosphere inside becomes rich in carbon monoxide and nitrogen. Consequently, the cyanogen band of radiation with its peak intensity in the near ultra-violet is very much strengthened, and this arc is particularly effective for photographic purposes and in the production of blue-prints.

The low intensity arc finds a specialised application in spectroscopy with plain carbon electrodes of a high degree of purity. The arc struck between these electrodes, on which traces of an unknown material are placed, produces the characteristic spectrum of the unknown and leads to the detection or identification of its constituent elements. The carbon used in the

electrodes for this work is one of the purest elements yet made on an industrial scale.

The brightness of the low intensity arc is limited by the sublimation temperature of carbon to about 180 candles per sq. mm. The brightness of the crater of the *high intensity* arc on the other hand is of the order of 400 - 1,200 candles per sq. mm. and evidently has its source elsewhere than in the solid tip of the incandescent carbon.

A diagram of a high intensity arc is shown in Fig. 38. The high current density, which is of the order of 100 amps per sq. cm., causes the formation of a cup-shaped depression in the

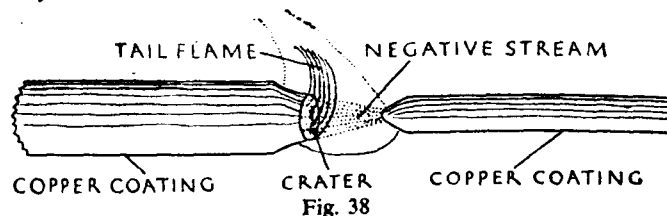


Fig. 38

crater of the positive carbon, the core of which contains easily ionisable cerium salts. About 30% of the light of the arc comes from a small tail flame formed by the gases from the core, but the majority of the light has its origin in an incandescent ball of these gases held in position in the crater by the negative arc stream.

Metals of the cerium group of rare earths are also used in the cores of electrodes producing a *flame-arc*. With this type of arc, the gas column between the electrodes is the main source of light, relatively little coming from the carbon tips. This type of radiation can be made to give a close approximation to the spectral distribution of natural sunlight, and consequently this carbon is used in 'sun-ray' lamps to duplicate the beneficial effects of sunlight on the human body. For the same reason this type of arc is used in accelerated-testing machines which determine in a short time the resistance of dyed fabrics, protective coatings and other materials to the effect of natural sunlight and exposure to weather.

If the core of a flame electrode contains iron, the resulting arc is rich in ultra-violet radiation and is used in machines to increase the vitamin-D content of milk. Ergosterol present in the milk is converted by the action of the light to vitamin-D and other substances.

Flame carbons also find application in the various processes of still and motion-picture photography, especially in colour photography where the preservation of the correct colour values is of major importance.

To obtain a bright, narrow, beam of light, a source of high brightness over a small definite area is needed. This condition is admirably fulfilled by the carbon arc and larger high intensity carbons are employed in searchlights.

It has been mentioned that high intensity carbons used for cinema projection are copper-coated. This is to increase the electrical conductivity of the electrode, which is fed with current by a clamp at the back end. Searchlight carbons on the other hand, carrying currents of well over 100 amps, are unplated, the current being fed through jaws just behind the burning crater. The electrode is fed mechanically through these jaws as it is volatilised away, and is rotated at the same time to maintain a symmetrical face on the crater.

The accompanying table shows the intrinsic brilliance of the various types of carbon arc compared with that of tungsten filament lamps and of the sun.

SOURCE	Candles per sq. mm.
Tungsten Lamp 40 watt vacuum clear	2.06
Tungsten Lamp 900 watt gas-filled clear	26.6
Sunshine flame arc	8
Positive crater - Low Intensity Arc	155-180
Positive crater - High Intensity Arc	380
Positive crater - High Intensity Arc (Rotating)	400-800
Sun at zenith	1,650-1,800

A brilliance in excess of 2,000 candle power per square millimetre has been obtained under special conditions in the laboratory.

ARTIFICIAL GEOLOGY

LESLIE BOR

GEOLOGY is a science essentially founded upon the study of rock structures and textures as seen in the field. Nevertheless, geologists frequently resort to experiment to find out what might happen under various physical conditions in nature. Clay blocks, for example, have been used to learn how plastic portions of the earth's crust behave when subjected to various stresses. Artificial silicate work, in which rocks and minerals are created under known conditions, is very important in elucidating the natural history of rocks. Whereas the earliest phases of geological enquiry were largely analytical, the synthetic approach to the earth's history is nowadays very prominent.

For this reason, geologists are interested in the products of, and the processes taking place within, industrial furnaces. Metallurgical, glass, coke and cement furnaces experience during their working life temperatures at least close to the temperatures of subterranean rock reactions. In addition, the linings of industrial furnaces are made from a wide variety of minerals. The bulk of refractory ware is manufactured from silica and alumino-silicates to give bodies varying in content from over 99% silica down to 20% silica and over 70% alumina. Magnesia, dolomite, chrome, spinel, and many other minerals may be used for special refractories needed to resist the severe conditions found in some furnaces. During furnace operation the heat and chemical action of the molten metal or glass invariably corrode the refractory walls. This, troublesome to technicians and expensive to industry, gives the geologist the opportunity to learn something about the mechanism of geological activity, for a striking similarity exists between the manner in which slag attacks its confining refractory walls and the activity of molten rock in the earth's crust. Molten slag

creeps into the microscopic pores of refractory ware and causes chemical changes which are manifest in the assortment of minerals crystallizing out when a blast-furnace, for instance, is allowed to cool down. Many of the geological phenomena seen in large blocks of the earth are thus displayed in a single corroded refractory brick.

A blast-furnace is nothing more than a large vertical cylinder, lined with refractory bricks, in which iron ore, limestone, and coke are burnt at temperatures above $1,200^{\circ}\text{C}$. During the period of intense heating, the metal is extracted from the ore and the residual impurities (mostly silica, magnesia, and alumina) are taken up by the lime to form a slag. This separation of ore into metal and slag is a second example of a geologically interesting process, since it is analagous to the earth's primary differentiation into a metallic core and a silicate crust. There are several features common to the everyday event in the blast-furnace and the historically unique differentiation of the molten earth. The laws of crystal chemistry, for instance, govern the distribution of chemical elements within the solidifying fluids of both earth and blast-furnace.

ABOUT OUR CONTRIBUTORS

LESLIE BOR was born in 1926. He studied geology at the Universities of Manchester and New Zealand, and is now engaged in mineralogical research at Cambridge.

MAX BORN, F.R.S., has been Tait Professor of Natural Philosophy at the University of Edinburgh since 1936. He was educated at Breslau, where his father was a professor in the University, and also at the universities of Heidelberg, Zürich, Göttingen and Cambridge. After various appointments he was a professor at Göttingen, 1921-33, and subsequently Stokes lecturer in Applied Mathematics at Cambridge, 1933-36. He is the author of about 240 scientific papers, and of eleven books, including *Atomic Physics* and *The Restless Universe*. He is a keen musician.

ALBERT EINSTEIN, now aged 71, has been Professor of Theoretical Physics and member of the Institute for Advanced Study at Princeton (New Jersey) since 1933. Before that he was Director of the Kaiser Wilhelm Institut für Physik, and Professor of Physics in the Prussian Academy of Sciences at Berlin. He has been twice married and has two sons.

CYRIL SANKEY FOX was trained as a Mining Engineer at Birmingham University, where, after the usual practical work underground in collieries in the Midlands, he was appointed a Lecturer in the Mining Department. He preferred geology and accepted an appointment in the Geological Survey of India, and eventually became Director of this department. He is essentially an engineer geologist with a particular flair for water-supply and related problems - soil erosion, irrigation, removal of alkali, prevention of water-logging, and problems of silting. He is perhaps best known for his wide knowledge and great experience in geological problems relating to mineral deposits, and their exploration by mapping, boring, and geophysical methods.

A. W. HASLETT is a Cambridge graduate, who has made scientific journalism his career. He is Editor of *Science Today*, and author of a number of books, of which *Science in Transition* was recommended by the Book Society. He is a member of the *Topic for Tonight* panel of the B.B.C.

JAMES T. KENDALL is a Science graduate of Cambridge University. In 1939 he started working in the research laboratories of Metropolitan-Vickers Electrical Co. Ltd at Manchester, and is now a Leverhulme Research Fellow at Imperial College, London.

L. S. B. LEAKEY was born in Kenya in 1903 and took his degree in Archaeology and Anthropology at Cambridge in 1926. From 1926 onwards he has devoted his whole time to the study of problems relating to early man in East Africa. Starting with a study of the Pleistocene he and his colleagues extended their work backwards in time to the Miocene, and their discovery of Lower Miocene apes of several genera has opened up a new field and thrown much new light on the problems of the early evolution of the stock from which apes and man have evolved.

JAMES PATON lectures on thermodynamics and meteorology in the Department of Natural Philosophy of Edinburgh University. He is especially interested in atmospheric optical phenomena and is engaged in research on aurora and luminous light clouds. He is Scottish Secretary of the Royal Meteorological Society and a member of the Committee of the Meteorological Office, Edinburgh.

J. W. S. PRINGLE was born in 1912 and educated at Winchester and King's College, Cambridge, of which he became a Fellow in 1938. During the war he worked on radar with Telecommunications Research Establishment. He returned to Cambridge in 1945 as University Lecturer and Fellow of Peterhouse, where he is now Tutor. He has published papers on nerve and muscle physiology, mainly of insects, and is a keen glider pilot.

HUGH SODEN was born in 1901 and educated at the Municipal College, Portsmouth. He is a science graduate of London University, and for the past eight years has been on the staff of the Research Laboratories of the General Electric Company. One of his great interests is in the History of Science, especially the eighteenth-century Birmingham men, and the Lunar Society.

BASIL STANDING served in the R.A.F. during the war, and since then has been a lecturer in Civil Engineering at Manchester University. He is carrying out research on various problems of fluid flow, and on hydraulic models.

D. J. SWAINE is a member of the staff of the Chemistry department of Melbourne University, of which he is a graduate. He is at present on leave to work in the Department of Spectrochemistry at the Macaulay Institute for Soil Research. His chief recreation is bushwalking.

BOOKS RECEIVED

EVOLUTION by Chapman Pincher (*Reason Why* series: Herbert Jenkins, 1950), 7s. 6d.

This book follows a well-beaten track through the evidence for evolution, and after an introduction to Mendelian genetics, discusses also the mechanism of evolution, and its present course. It is written as the answers to a series of questions, which makes it somewhat disjointed, and illustrated by the author's drawings, which unfortunately are rather poorly reproduced. Since the pages are also overcrowded with a small size in print, the general appearance of the book is discouraging. The treatment seems adequate, if not exciting: Mr Pincher is evidently a better sprinter than long-distance runner.

GLOSSARY

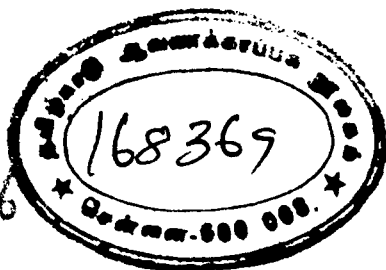
GEOLOGICAL AND PREHISTORIC PERIODS: Authorities differ on the precise time-scale, and on the exact point at which one period ends and the next begins, decisions made more difficult by the complexity of relating discoveries made in vastly different parts of the world to a common scale when there are few apparent connecting links. But a roughly satisfactory scheme runs like this (most recent on top, most ancient below):

<u>Geological Epoch</u>	<u>Approximate Age</u>	<u>Archaeological Period</u>
(recent)	5×10^3 years	metal age neolithic mesolithic
Upper Pleistocene	15×10^3	palaeolithic Magdalenian Solutrean Aurignacian Mousterian
Middle Pleistocene		Levalloisian (second Ice Age) Acheulean (first Ice Age) Chellean
Lower Pleistocene	1×10^6	} (the age of mammalian evolution)
Pliocene	15×10^6	
Miocene	25×10^6	
Oligocene		
Eocene		

PALAEOLITHIC of course, is the same as Old Stone Age.

INDEX

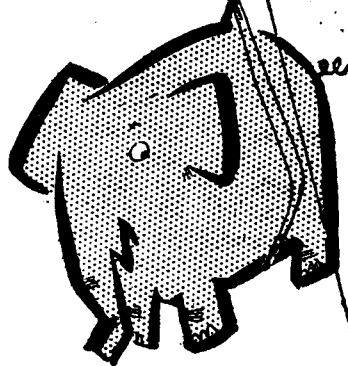
- Adaptation in sense organs, 105
- Aden, water supply of, 65
- Aeroplane studies of thunderstorms, 73
- Antecedence and causality, 13
- Atoms, forces binding together, 79
- Archaeological dating with radio-carbon, 115
- Binding forces between atoms, 79
- Causality and contiguity, 13
- Cause-effect relation discussed, 11 *et seq.*
- Climate, changes in, 122
- Cloud development, 69
- Complementarity (Bohr's ideas), 19, 27
- Conduction of electricity, 86
- Impurities and, 82
- Theory of, 81
- Contiguity and causality, 13
- Cumulonimbus, 67, 73
- Determinism examined, 11
- Electric current, nature of, 81
- Fauresmith (African) culture, 43-4
- Forces binding atoms together, 79
- Fossil apes discovered by Broom and Dart, 37
- Geological map-making, 59
- Germanium, semiconduction in, 83
- Glaciation in clouds, 69, 70-2
- Gravity, the force of, 118
- Hailstone formation, 72
- Hand-axe culture in Africa, 40-3
- Helium II, atomic nature of, 125
- Hyrax Hill neolithic culture, 47
- Ice ages and Africa, 43
- Ice in clouds, 70-2
- Insects, disease and spread by transport, 128
- Sense organs, 103, 130
- Instruments, purpose of, 101
- Kenya, Capsian folk, 45, 47
- Fossils from, 37 *et seq.*
- Lapse rate of ascending air, 67-8
- Mathematics, language of explained, 106
- Models, scientific use of, 137
- Nerves, properties of working, 102
- Oxides as semiconductors, 85-6
- Prehistory in Africa, 40
- Primitiveness of human fossils, 38
- Radar and thunderclouds, 73, 77
- Rain formation, theory of, 70
- Rectifiers, crystal, 93
- Nature and function, 92
- Rime, nature and formation, 72
- Sangoan (African) culture, 44
- Second sound, speed of, 123
- Semiconductors listed, 96
- Sense organs and radiation, 130
- Sickle cell anaemia, 127
- Skulls of apes and men compared, 38
- Solids, electrical properties of, 80, 86-9
- Atomic nature of, 78
- Straingauges in insects, 103
- Thermoelectricity, 90
- Time measurement, 121
- Transistors, 95-9
- Transvaal, fossils from, 37
- Underground water flow, 54-7, 64
- Weber-Fechner law, 105
- Wells, proper siting of, 57



Am, 246

NSO-17

the advantage. is your



with **PHYSICS** | *toda*

A monthly magazine written especially for the reader outside the field being discussed, **PHYSICS TODAY** can keep you abreast of current happenings, trends, and problems in physics. Its feature articles, book reviews, news and views are presented in non-mathematical language with emphasis on generalization and principle, not on disparate detail. The American Institute of Physics, publishers of eight technical journals, began this post-war venture in science journalism expressly to let one kind of physicist know what another kind of physicist is doing, to let any scientist know what physicists are doing, and to provide old students (with a frustrated interest in current happenings) a new insight into physics, today. The cost is \$5.00 per year (postage abroad included) and we will be happy to include you on our subscription list if you would send your name, address, and a remittance to:



PHYSICS TODAY • 57 East 55th Street • New York 22 • N.Y.