

SCIENCE
NEWS

VIII

edited by

J. L. CRAMMER



PENGUIN BOOKS

1948

ACKNOWLEDGEMENTS

For permission to reproduce the plates in the inset we are grateful to

The Visitors of the Ashmolean Museum, Oxford.

The Prehistoric Society.

The Ray Society.

University Museum of Archaeology and Ethnology, Cambridge

National Museum of Wales, Cardiff.

The Controller of H.M. Stationery Office and the Director
General of the Ordnance Survey.

The Society of Antiquaries of London.

Dr. A. Bulleid, F.S.A., and Mr. H. St George Gray, F.S.A.

Messrs. Magnesium Elektron Ltd. (plates 21 and 22).

B.T. Batsford Ltd.

Macmillan & Co. Ltd.

Methuen and Co. Ltd.

Editions A. & J. Picard and Cie, Paris.

'We regret that we omitted in Science News 7 to give acknowledgement for the photographs of the Micromotion study of the collation of documents (plates 42 and 43) to W. H. Smith & Sons, and of the Radiograph of golf club (plate 27) to Westinghouse Electric.'

Editorial

FACTS, and more facts! cry the laboratory scientists, banging their fists with approval on the bench, and jolting some apparatus in course of construction. But what's their significance? say the natural philosophers, looking up from their desks, leaving the mixture of doodles and speculation scribbled on the paper before them. And the two groups eye one another with suspicion and mistrust. Which camp you work in is a matter of temperament, and the sad truth is that so few are capable of work in both. Yet both are essential to the progress of science. All the time, the experimenter is battered by a flood of inchoate sensations arriving simultaneously through all his senses to his brain: how shall he select the significant, direct his powers of concentration and interpretation, except by reference to some theory previously acquired, and now perhaps so completely absorbed as to be a part of the unquestioned dogma of his everyday life?

Ideas we must have, and without them facts are nonsense. But ideas, like practical methods, are best out in the open where everyone can see them for what they are, with their little invalid assumptions and shoddinesses. It is our faith and our policy to emphasize the thought over the fact in the fields of science our articles review, and we shall do this more and more in later issues. In this way we hope not merely to provide a mental map on which the factual detail can later be inscribed, but further to stimulate the progress of Science itself by drawing attention to ideas of one speciality which may find a use in another, unrelated to it for often a single new intellectual approach proves fruitful over a wide area.

This issue opens with an account of how the archaeologist goes to work, and closes appropriately with a study of the

subject which has made our culture materially so different from the preceding Stone Ages—Metals and Alloys. Dr Fox, in an article which repays careful study, discusses the underlying principles in the production of alloys, and the problems of modern metallurgical research. His essay can be profitably read with the article on Metals by Sir Lawrence Bragg which appeared in our first issue. In *What the Earth is Made of*, Lieut.-Colonel Tillotson excavates a good deal deeper than the ten feet or so of surface soil which interests the prehistorian, and describes how earthquakes and artificial explosions are used to probe the composition of the underlying strata: and to search for oil. Mr Haslett discusses the origins of Coral Islands, and the light thrown on this problem by recent excavation, in his Research Report. *Viscosity*, by Dr Bell, is an elementary introduction to the modern conception of the liquid state, which is exemplified also in a note on *Glass*. *Group Psychotherapy* explains the general approach in the psychological treatment of neurosis. *Demography, Science and Administration* is a critical discussion of the study of population and its vast importance for modern government, which can only apply controls successfully to a society whose size and composition are known. Throughout, our contributors are striving to give a critical, balanced account, to reveal how the new 'facts' were obtained and what their significance is taken to be. We want to offer something more valuable than knowledge, and that is Understanding.

Scientific Prehistory

L. W. CORNWALL

What is Prehistory?

WRITTEN records of the past are the raw material of History – but what of the long career of mankind before writing was invented? That is the field of the science of Prehistory. In the Near East writings are found from as early as about 4000 years B.C., but in Britain Julius Cæsar's account of the first Roman invasion brings *our* prehistoric period to an end as late as 55 B.C.

The prehistorian compiles his account from the traces left of their lives and activities by peoples without writing. His principal helper is Archæology, the science which studies human culture through its material equipment, but, especially for the earlier part of man's story, many other sciences have an important contribution to make.

In this sense, palæontology and geology are the handmaids of Prehistory. They enable us to discover something of the biological origins of the human race and of the nature of the world in which the earliest men lived, their way of life and relation to their environment. Another contributor is comparative anatomy, which concerns itself with their rare skeletal remains and gives us some account of their forms and faces. Archæology takes a hand when their practically indestructible stone implements are in evidence and becomes the principal instrument in exploring the later periods, when, with advancing material culture, remains of the homes, villages, cemeteries, tools, weapons, ornaments and so on of ancient man become available for examination.

Many of the rarest and most instructive relics of man's remote past, such as skulls, bones and stone implements, have been discovered by chance when digging foundations

have been wattle hurdling daubed with clay for the walls. If this had been burnt the impression of the hurdle might be left for us to find. Of course there would be pottery, but many of the other household fittings and utensils must have been of wood. All this, with basketry and woven textiles, would perish, but clay or stone spindle-whorls and loom-weights would survive to testify to the manufacture and use of textiles. Nothing made of hide, horn, sinew or gut – all useful primitive materials – would remain, but bone and deer-antler implements might be preserved – blade-bone shovels and antler-picks, pins, weaving-combs, etc. – and even, perhaps, much-corroded iron tools – axes and bill-hooks. The rest of the farm implements, mostly of wood, must have rotted. Of clothing we would find only bone or metal fittings and ornaments – pierced animal-teeth, shells and beads for stringing, bronze brooches and so on. Sling-stones, flint arrowheads and metal weapons may have survived; of food, only the meat-bones and perhaps carbonized grains of corn.

In peat-bogs and other waterlogged deposits, objects of wood, basketry, textiles, leather, grains, fruits and seeds are occasionally preserved to give us a glimpse of how much is generally lost (Plate 13).

For still earlier times, when men lived in caves or in the open and had a much scantier equipment than this, there is even less to show – flint and other stone implements, bone tools, meat-bones and, perhaps, a few pierced teeth, shells and other ornaments. Even pottery, the standby of the later periods, is lacking, but the caves have yielded clay-modelled animal figures and pictures, chiefly of animals, engraved, carved and painted on the walls.

What is a 'dig' and how is it done?

(a) Where to dig.

The flint hand-axes and flakes of the earliest men, and their very rare bones, are generally found by chance in commercial excavations – gravel-, sand- and clay-pits, and,

occasionally, in limestone quarries which expose the filling of ancient caves. The bones of extinct and still-surviving species of animals among which they lived are found in the same places. This stage was the Old Stone Age or Palæolithic period. In the later part of it the climate of what is now temperate Europe became almost arctic at times. Mankind took to caves in limestone rocks and over-hanging shelters (Plate 8), and it is in, and in front of, these places that their tools, animal bones, wall-paintings and, sometimes, their own remains are found.

Later, forests covered most of the land. Middle Stone Age, or Mesolithic, men lived in the more open places – sea- and lake-beaches, in fens and bogs and on sandy wastes. Mounds of shells and rubbish, the waste from their beachcombing and hunting, yield evidence of their way of life, as well as their tools and sometimes their skeletons.

Soon we find the first settled communities tilling the soil, stock-rearing and (most important for us) using pottery. Their settlements are found on the open Chalk downs and hill-tops, associated with the long barrows, cairns and chambers of huge stones which they built as communal tombs (Plate 2).

Round barrows and other circular monuments, such as Stonehenge, belong to the Bronze Age, though it is likely that some of them, at least, continued to be used and venerated up to historic times. There are also cemeteries of cremated burials in pottery urns, of the later Bronze Age. Only very few settlements of these people are known. Apart from their graves and temples, chance-found 'hoards' of bronzes, probably the stock-in-trade of travelling smiths, tell us almost all we know of Bronze-Age equipment in this country.

Many commanding hill-tops are crowned with forts and earthworks dating from the Early Iron Age (Plate 3). We also know some habitation-sites and widespread field-systems of the time. The latter are often marked by

These principles apply to every dig and every archaeological period, with variations in plan of action for particular cases. A round barrow, or other circular structure, may be excavated a quadrant at a time (Plate 7); an area such as a cave-floor or an ancient land-surface, without visible structures, by arbitrary squares; the line of a wall or ditch plotted by short transverse trenches at definite intervals.

Interesting details, such as burials and groups of pottery or post-holes denoting wooden buildings, hearths and so on, are carefully cleaned and recorded in place, with every associated object, before anything is moved.

Finally, the dig is filled in and levelled off.

*How the fieldwork is recorded
and what the excavator takes home.*

Of no less importance than the actual digging is the recording in the field. The field-notebook contains a day-to-day account, with detailed small scale-plans and sections, of the work as it is actually done. Trenches or areas, and each distinct level in them, receive identifying symbols. Finds in them are individually or collectively marked with their appropriate locations and a reference to the page of the notebook on which their discovery is recorded. Objects other than pottery may be entered on cards with serial numbers, each with a rough drawing, measurements and description and all the relevant information as to its original situation and position.

Pottery, which may be very plentiful on later sites, is washed, sorted and selected on the site. Much of it may consist of small undistinguished sherds belonging to a common and well-known type. Once enough types have been selected, establishing every characteristic variety of ware, pot-form, finish and decoration found on the site, the rest may be discarded and re-buried when the dig is filled in. In this way careless disposal of unwanted fragments will not complicate future excavations anywhere else. Of any group of sherds which seem likely to have formed a

single vessel every one is kept, in the hope that the pot may be rebuilt or at least its shape recovered.

Photography at every stage saves pages of description in the eventual publication. More important: it is *prima facie* evidence of the original appearance of structures and objects which may have had to be moved or destroyed as the dig proceeded. Archaeological photography is a special art, beset with snags and limitations. It is well to be sure that a good negative has been obtained before the irreplaceable subject is demolished.

Samples of shells, bones, soils from the different layers, peat, charcoal, rocks and minerals (in the case of geological deposits), food-grains and industrial waste (e.g. metal-slag) may help, after examination, to fill in the picture given by the other finds. They are best taken in person by the expert who will examine them.

Deciphering the story told by the finds.

(1) *The study of types.*

Ancient man was bound by tradition, and successive generations tended to reproduce the types of implements, pottery, and so on used by their forbears. Changes, both of evolution and degeneration, took place by degrees insensible in any one lifetime. The archaeologist, therefore, bases many of his conclusions on similarities, and the indicated relationships, between known types and the particular objects he is studying. A whole assemblage of characters, common to the culture at his site and sites known elsewhere, is even more convincing evidence of relationship. Thus, a particular group of people buried their dead in a crouched attitude in round barrows, used winged-and-tanged arrowheads of flint and placed with each body a clay beaker of characteristic form with a peculiar style of comb-impressed decoration (Plate 12). This assemblage, when found at two widely-separated places, establishes a close relationship. We do, in fact, recognize the 'Beaker-Folk' of the early Bronze Age by this

constant assemblage of characters. Their remains are found all over Western, and a part of Central, Europe.

It seems as if at least the more momentous discoveries of prehistoric times – of the bow and arrow, of agriculture, of the wheel and so on – took place only once and in one region, being thence diffused by migration, trade and contact between adjacent peoples. Evidence for the contrary process of independent development, at several different

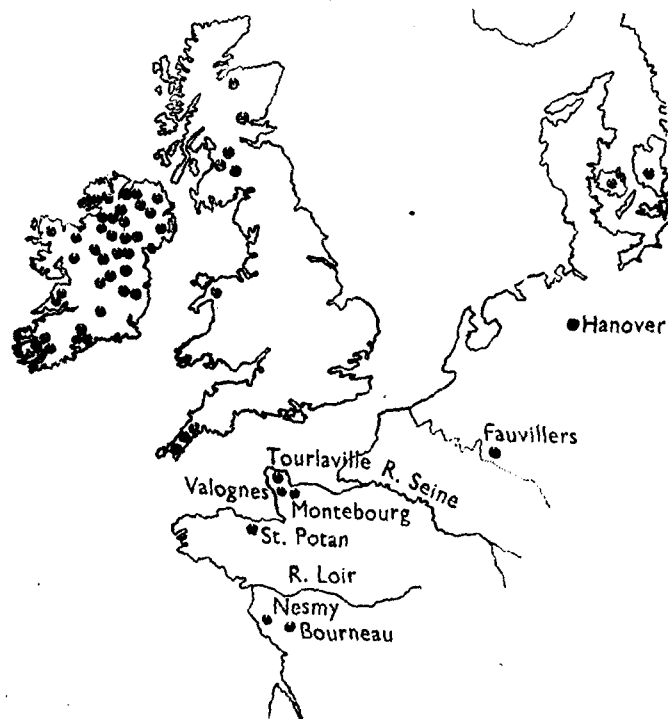


Fig. 2.—Distribution—map of gold 'lunulae' (crescent-shaped neck ornaments) of the Early Bronze Age, showing their Irish origin, outlying occurrences and predominantly maritime distribution. (Fox: *The Personality of Britain*. Nat. Museum of Wales, 1943.)

centres and times, is much more scanty, but cannot be disregarded as a possibility in certain cases.

A sudden cultural break at a site, therefore, indicates outside influences at work. This is seen by the excavator as a sharp distinction between the cultural contents of one layer and the next above it. It may not be obvious, consisting, perhaps, only of a change in the occurrence or character of a single class of objects, but when it is found it is highly significant.

(2) *Distribution.*

If the new element is already known at other sites and in other regions, the investigator will consult, or construct, a map of the distribution of that and allied objects. The map may show the greatest concentration of the foreign element in a particular region, with outlying occurrences along natural routes linking it with his area (Figs. 2 and 5). In this case the prehistorian may conclude that he has discovered the source of his intrusive objects, or, at least, of their inspiration (for they may not be identical with the originals). According to the degree of resemblance and numbers of objects involved, he must decide, perhaps with the aid of statistics, whether his finds are sporadic imports, local copies of such imports (Fig. 3) or the property of an actual body of migrants or invaders from the parent region.

Objects themselves betray something of the way of life of their owners. Hoes and sickles point to agriculture, axes and adzes to wood-working, querns to the grinding of cereals, swords to war and conquest. If discarded bones are all those of wild animals it is to be inferred that the people in question hunted, but did not rear, their meat.

(3) *Specialist reports.*

These may add very sensibly to our information. Here is an instructive example of what the palaeontologist can contribute to Prehistory.

W. Soergel studied animal bones from several sites in

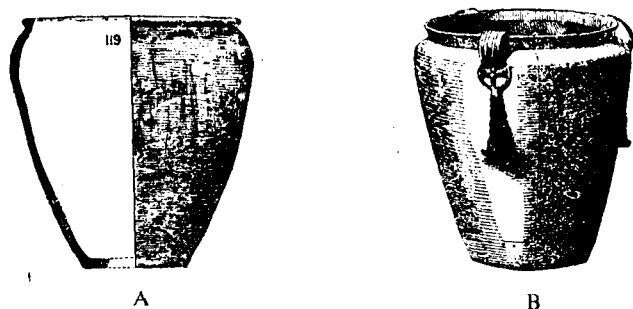


Fig. 3.—A typical Early Iron Age household pot from Maiden Castle, Dorset and a bronze 'situla' or bucket from N. Italy, of which the former is a debased form, translated from metal into clay.

(A. from Wheeler: Maiden Castle. Soc. of Antiquaries, London, 1940. B. from Déchelette: Manuel d'Archéologie préhistorique, A. & J. Picard, Paris, 1908.)

Germany, at some of which implements and actual remains of Palæolithic man had been discovered. He found that his sites fell into two well-defined groups, in one of which a third to a quarter of the remains of elephants belonged to young beasts less than 6 years old. In the other group older animals predominated, some two-thirds being over 50 years old and none under 6. Now the first group included all the sites at which man was known to have been living, as well as two, showing a small proportion of young animals, where there was no evidence, at that time, of his presence. Soergel concluded that, in those places, man had hunted the inexperienced members of the herds, for nothing would otherwise account for the difference in age-distribution. Further support for this theory came some years later, when a human skull was found at Steinheim, one of his two anomalous sites. The other remains a blank, but we look forward hopefully to hearing, one day, that the presence of man has been proved there also.

To-day we associate elephants, rhinoceroses and lions with tropical regions. The deposits just described show that

they were also numerous in western Europe when the men of the early Old Stone Age were living, though the climate

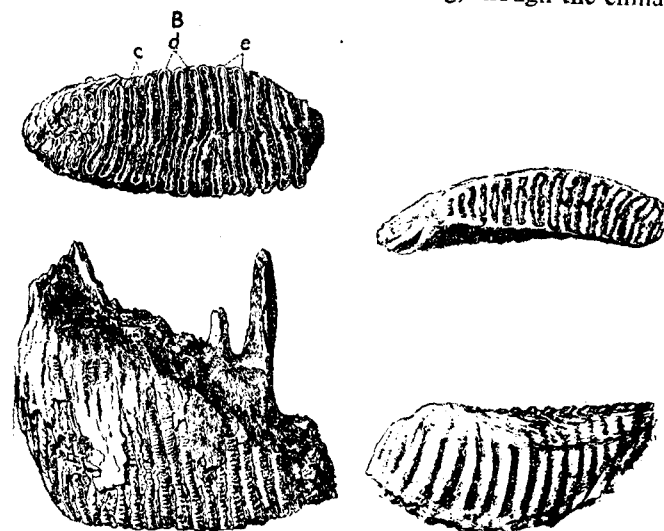


Fig. 4.—Elephant-teeth. On the left, two views of an upper molar of the Mammoth (*Elephas primigenius*), whose presence in a geological deposit denotes a cold climate for its formation. On the right corresponding views of a lower molar of *Elephas antiquus*, which lived in a temperate or warm climate.

From Zittel's Textbook of Palæontology, Macmillan.

was not very different from that of our day (Fig. 4). On the other hand, arctic conditions are indicated by the remains of reindeer, lemmings and arctic foxes from the cave-earths of southern France, belonging to later stages of the same cultural period.

Nor are bones the only indicators of climate. Sea and freshwater shells and land-snails can indicate fluctuations far less extreme than these. A few degrees of change in average annual temperature or sea-salinity, a few inches difference in rainfall, will be reflected by alterations perceptible to the conchologist in the relative numbers of the

species composing the molluscan faunas of beaches, river flood-loams and field-ditches.

The botanist is another specialist whose help is often required by the prehistorian. He can determine the species of trees and bushes from sizeable fragments of charcoal in the hearths of prehistoric sites. Quite a full picture can be made in this way of the local vegetation of the time. In a peat or a lake-silt the very wood of piles or wattle-work may be preserved for him to determine. Even wind-borne pollen of trees and plants, preserved in such conditions, can be identified under the microscope. Pollen-analysis of a level in a peat showing the presence only of pine and birch woods suggests a climate colder, at least in winter, than one with oak, elm, lime and other deciduous forest trees. This method is now being extended to include the pollen of herbs and grasses as well as of trees.

The botanist may also be able to trace, from a few carbonized grains or a grain-impression in the once-plastic clay of a pot, something of the story of prehistoric cultivated wheats and the travels of the men who sowed and reaped them.

The composite picture of animals and plants contemporary with early man may be of importance in dating, at least for the earlier periods of prehistory.

Apart from the palæontologist and the botanist, the help of the expert on soils is often needed. In later times, climate and surroundings were not very different from what they are to-day, but the soil-scientist can recognize a 'fossil' turf-surface in the section of an Iron-Age rampart which has been reconstructed in antiquity (Plate 4) as readily as he does the weathering in a boulder-clay or loess due to some thousands of years of exposure.

The petrologist may be able to say where the characteristic stone for a particular 'polished' Neolithic axe originated. That of an axe found in Hampshire must have been quarried at a particular place in Cornwall. Axes of a rock from the Lake District have been found near Barrow-in-

Furness, Birmingham, in Northamptonshire, Buckinghamshire, Surrey, Wiltshire and near Gloucester. The unavoidable inference as to the widespread ramifications of trade

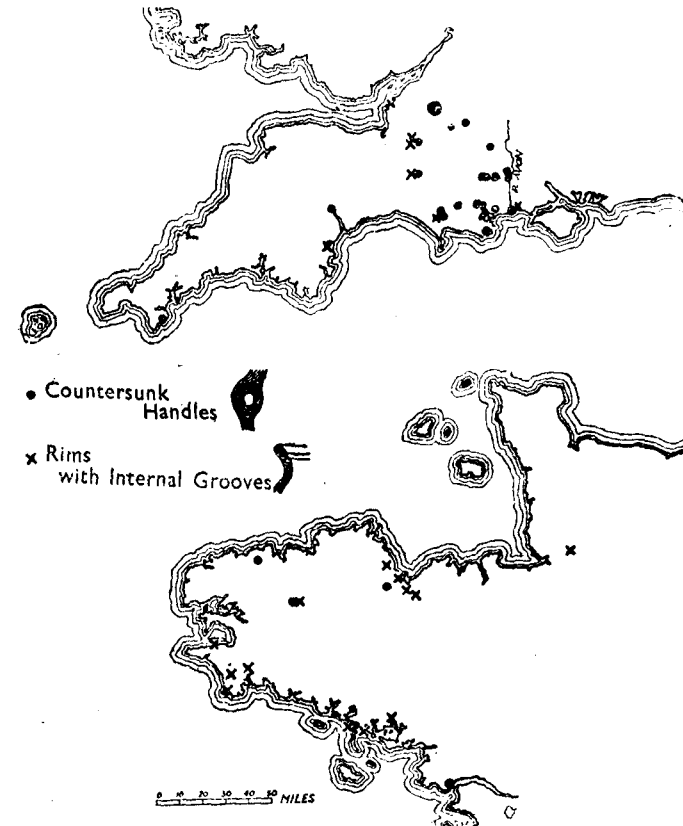


Fig. 5.—Distribution map of Early Iron Age countersunk pot-handles and internally-grooved rims, showing Breton cultural affinities and the probable points of entry into Britain with the concentration of the immigrants in Hants and Wilts.

(Wheeler: Maiden Castle. Soc. of Antiquaries of London, 1940.)

and exchange, even in Neolithic times, is most striking.

Similarly, the chemist may determine, from the composition of a far-travelled amber bead, whether it originated in East Prussia or Sicily. A mineralogist may be able to decide, from traces of characteristic impurities present, the particular ore-deposit from which the metal of a Bronze-Age spear was smelted. The student of animal bones can help to unravel the history of domestication in animals. Working with the human remains, the anatomist and physical anthropologist seek to define the racial affinities of the people concerned.

Experts' reports on materials from any one excavation are not always impressive or abounding with spectacular results. Now and again, one of them may, however, be able to say that, to the best of his knowledge, a particular object or material from a British site could only have come from (say) the Rhineland. At once a connection of some kind is set up between two remote prehistoric peoples, of which the importance to the prehistorian may prove to be very great. It is worth many inconclusive investigations to obtain even one such informative result.

How long ago was it?

Archaeological deposits consist mainly of the rubbish of human occupation and have slowly accumulated where they lie, often in more or less distinct layers. *Provided that there has been no later disturbance*, it follows that what is found deep down must be earlier than what lies above it. Very often, however, there *has been* some disturbance – as when later occupants dug pits or wall-foundations have been active. If rubbish, or when rabbits and badgers have been active. If we find, say, a Roman coin in the same layer as Neolithic pottery, there is only one possible conclusion: that the deposit has been disturbed and mixed, perhaps as early as Roman times, possibly only quite recently. Certainly the pottery is not in its original context. The latest datable object dates the layer in which it is found. If, however, we

find the same New Stone Age pottery in a layer including nothing demonstrably later in date, we may be reasonably certain that we are in a stratum which has lain undisturbed since the time when that sort of pottery was being made and used. The same stratigraphic principle (derived, of course, from geology), applies to Old Stone Age relics in geological, not humanly-formed, deposits. This principle gives us a *relative chronology*: the bare order in which prehistoric events occurred.

Within the last ten years or so, geology has developed tentative *absolute time-scales* – dating in years.

One of them is based on the known extremely slow rates of decomposition of traces of radioactive substances present in rocks and minerals. This dates the earlier geological periods.

Prehistory is, of course, only seriously concerned with the Pleistocene and Recent periods of the geologists, comprising, perhaps, the last million years of the Earth's estimated age of 3,000 million. The sub-human ancestors of man must have been in existence before this, but the first undoubted men known to us belong to the early part of the Pleistocene, say 600,000 years ago.

Working on geological evidence in the Alpine region, Penck and Brückner constructed a curve to show the growth and retreat of four main Pleistocene glaciations and to indicate their probable duration and extent.

Now, astronomers, for their own purposes, have calculated for the last million years the effect, on the amount of solar radiation received at the Earth's surface, of cyclic changes in the complicated motions of our globe. Some geologists interested in the Pleistocene Ice Ages have noted a degree of correspondence between these results and their own which seems to exceed that attributable to pure chance, on a basis of mathematical probability. Assuming some connection, therefore, but without attempting to define it, they have provisionally correlated the more detailed 'radiation-curve' and its time-scale with the glaciation-curve

(Fig. 6). The absolute durations and intensities of glaciation, and of the interglacial periods, indicated in this way, prove to be of much the same order as that suggested by Penck and Brückner on purely geological grounds. We thus obtain dates for the main events of the Pleistocene which, if not yet very accurate, are at least valuable working hypotheses and probably of the right order of magnitude.

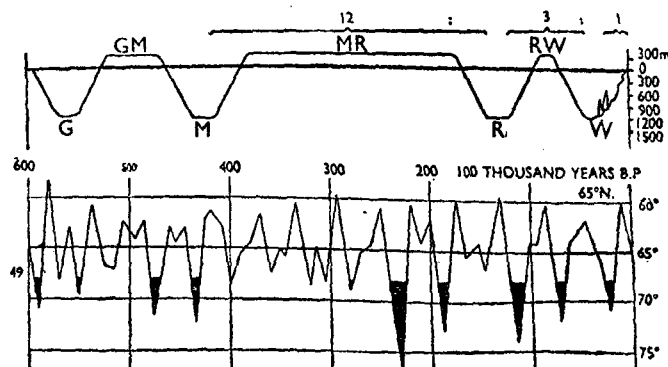


Fig. 6.—Glaciation- and radiation-curve. The upper curve is that of Penck. The axis represents the level of the present Alpine snowline with graduations above and below it in metres. The dips in the curve correspond to the four Alpine glaciations, Günz, Mindel, Riss and Würm (marked by their initial letters); the peaks to the Günz-Mindel, Mindel-Riss and Riss-Würm interglacial periods. From the degree of weathering of the glacial deposits, Penck estimated that, taking post-glacial time as unity, the Riss-Würm amounted to a period 3 times as long, Mindel-Riss being 12 times as long.

The lower curve is the radiation-curve calculated by Milankovitch in 1930. It shows, on an absolute time-scale, the variations in solar radiation received at Lat. 65°N. during the last 600,000 years. The four groups of minima shown in full black are regarded by some geologists as corresponding with Penck's four glaciations. In this case, three of them are seen to have been double and the last triple. Recent field-work tends to corroborate this theoretical multiplicity of the glaciations.

(Penck's curve re-drawn from W. B. Wright: *The Quaternary Ice Age*, 2nd ed., Macmillan, 1936. Milankovitch's from Zeuner: *The Pleistocene Period*, Ray Society, 1945.)

During the last retreat of the ice-sheets, certain banded or 'varved' clays laid down in glacial lakes have been shown to mark the seasons of retreat, by a thick layer formed during summer thaw followed by a thin band during the winter freeze-up (Plate 5). Series of these have been counted and linked up and give reasonably reliable dates for the last 12,000 years. Now, with the varves can be correlated minor climatic phases botanically determined by pollen-analysis. The settlements and implements of pre-historic man are also often accompanied by tree-pollen, so that the dates of the varves can sometimes be transferred to the human cultures.

Finally, we come to the astronomically-fixed calendars and king-lists of the earliest Near Eastern civilizations, which flourished long before western Europe emerged from savagery. Datable objects traded and conveyed thence, when found with prehistoric European remains, afford approximate historic dates for these by synchronism, ample allowance having been made for possible long delay in transit.

A striking example of the method is made possible by the occurrence, in round barrows in Wessex, of segmented faience beads, clearly imported, of which the nearest known parallels were made in Egypt about 1400 B.C. From these we get an approximate absolute fixed point for our Early Bronze Age (Plates 14 and 15).

What do we get out of it?

A great contrast between History and Prehistory is that, while the former is largely concerned with the doings of prominent individuals, Prehistory deals with the daily life of nameless ordinary people. Dynasties and policies have passed unrecorded, but we do begin to know something of the domestic arrangements of the prehistoric peasant-farmer.

Even if the story is tantalizingly incomplete, Prehistory shows something of the evolution of man, of his equipment, economy, society and ideas. Unconsciously, perhaps, but

not less certainly, we compare what we can discern of the life of prehistoric man with our own. If we are tempted to feel a little patronizing towards the achievements of our ancient forbears, we should remember that we deserve no credit for having been born in this century and consider soberly how we, as individuals, might have fared in the conditions of, say, the Old Stone Age.

Once started, knowledge and technology advance, to some extent, by their own momentum. Thus, the astronomer, the atomic physicist and the microscopist, among others, make prehistoric dating possible. Without their equipment and skill, it would still be a matter of pure speculation over most of the field of which we have knowledge.

Even to-day, when education is general and research is a field open to any trained inquirer, there are not very many original minds among our enormous populations. Is it not surprising, therefore, that the invention of so many of our basic economic activities is due to the ingenuity of prehistoric men – among them agriculture, domestication of animals, pottery, spinning, weaving, turning, metallurgy? No one will maintain that the technique of weaving was a chance discovery. Here are the working of brain and hand, the exercise of ingenuity and imagination, at their best. Though the smelting of copper ore was probably discovered by chance, how many of us, seeing a fragment of green malachite amid the glowing charcoal and finding a globule of copper among the next day's cold ashes, would connect the observations and arrive at the conclusions necessary to found thereon a metallurgical industry?

All the superstructure of civilization upon which we pride ourselves is based on some such rare exercise of intellect on the part of a long-dead anonymous artisan. Without such as these, it could never have been erected. Without their heirs, in our time, it may well collapse.

Therefore, better to understand the hothouse civilisation we have built round ourselves, we need to understand the

origins of society – the Old Stone Age family feasting behind its blazing brushwood in the cave-mouth; the Neolithic farmer and his neighbours safe with their brimming grain-pits as the winter sets in; the barbaric Belgic conqueror and his followers, lords of all the South and East of Britain before the Romans came.

History is the connecting link between their times and our own, but it is necessary to trace the devious career of man from his origins up to the opening of history, to obtain the fullest understanding of ourselves.

Eye Grafting To-day

DR MAURICE OUDOT

THERE is no question, in the present state of our techniques, of making a total eye graft, that is to say of replacing a diseased and blind eye by another healthy one to give the patient his sight back. Only a *localised* ocular replacement can be made.

The front part of the eyeball, called the *cornea*, is completely transparent, revealing behind it the black disk of the *pupil* surrounded everywhere by the membrane of the *iris* (See Fig. 7, a sagittal section of the eyeball.) Thanks to this

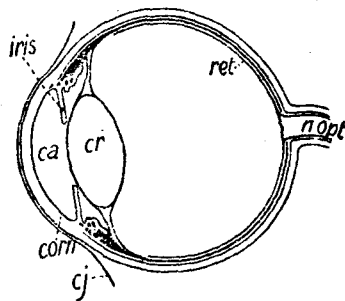


Fig. 7.—A section of the eyeball: front on the left, back on the right. *corn*, cornea; *ca*, anterior chamber; *cr*, lens; *ret*, retina; *n opt*, optic nerve; *cj*, conjunctiva.

transparency, the cornea transmits and refracts the incident light rays and so permits of vision. If, for one cause or another of an accidental or infectious nature, it becomes opaque, the light is blocked, and the sufferer complains of much weakened eyesight. And yet, behind this veil the other media and membranes of the eye remain in good shape and quite fit for the perception of images.

That is why, in those cases, unfortunately so frequent, where local and general medical treatment have failed, the only thing to do is to make a fresh transparency with the help of eye surgery. This problem has been resolved by keratoplasty, or corneal transplantation.

Although in ophthalmic surgery to-day the palm of novelty goes without question to corneal grafting, for all that it is not a recent discovery. Science means patience, and the first trials were made as long as 140 years ago. In this work, we remember abroad the names of Van Hippel, inventor of the trephine for detaching discs of corneal tissue, of Salzer, of Fuchs, of Elschnig, of Nizetic, of Imre, protagonists of new techniques, and especially the name of Filatov, who created and organised a specialised Institute for Keratoplasty at Odessa in 1927. America came into the field later, but with the immense resources at her disposal, and the drive of Elschnig's pupils, who had emigrated, she has largely made up her leeway and floods us with her publicity. On the other hand, in France for instance, in spite of the early attempts of Magitot, which go back to 1913, the operation remained a Congress curiosity for a very long time.

But a few years ago, just before the 1939-45 war, efforts increased. The Lyons school, in particular, in contact with Central Europe through the intermediacy of Switzerland, became interested and excited by the new techniques; and already, French experts such as Sourdille believe on the basis of their personal experience that the corneal graft is now a field for practical surgery, and that very probably the next few years will see operations of this kind on the increase. Theoretically, intervention is very simple. One removes an opaque segment of cornea and replaces it by a transparent piece from a healthy eye.

The *graft* is the portion of transparent cornea which will replace the opaque section of the diseased eye. It can be taken from an eye which has had to be removed from a patient for some other reason. But all authors agree in preferring a graft taken from the eye of a corpse. Enuclea-

tion of the eyeball must be made as soon as possible after death, in the first twelve hours. It is then kept in a cold room 24 hours or more, preserving it in the future recipient's blood with every desirable aseptic precaution. Unfortunately in some countries laws forbid any interference with a dead body until 24 hours after the death has been registered. This must be modified as soon as possible because of the obvious social importance of the operation.

The Americans, with that sense of size, organisation and publicity peculiar to them, have conceived of the creation of a special bureau to supply the necessary material: 'The Eye Bank for Light Restoration.' To this agency (in New York) come those who will have to have an eye removed, to offer themselves for a future corneal transplantation. Furthermore, a great publicity campaign has been undertaken to encourage people of every race and all ages to leave their eyes to the bureau in their wills. They will be examined, and tested to confirm the possibilities of future use. Recently, to the original collecting centre has been added a teaching centre which gives every ophthalmic surgeon the opportunity to educate himself in the technique of operation. In the course of its first year's functioning this body has responded to more than 3,000 requests. Without going to the full length of such an organisation, which makes us smile a little, we would find it worthwhile to collect together offers and requests for corneal transplantation.

Operative technique.

The graft can be penetrating or non-penetrating, total or partial, according as it involves all the thickness of the cornea or not, and all its surface up to the limbus (periphery of the cornea) or only a limited segment of that surface. We cannot for the moment make a success of a total graft for the whole of an opaque cornea. Actually the most frequently practised operation is a transplantation partial as regards surface - limited to a central zone 4 to 5 mm. diameter, corresponding to the visual area of the cornea;

SCIENTIFIC PREHISTORY Plates 1 - 15.



Plate 1. System of excavation in squares, leaving intervening balks as 'controls'. The East Entrance, Maiden Castle, (Wheeler: Maiden Castle, Society of Antiquaries, London, 1940).



Plate 2. A 'dolmen': a New Stone Age communal tomb, probably originally covered with a mound of earth. Pentre Ifan, Pembrokeshire. (Grimes: *Proceedings of the Prehistoric Society*, 1936).



Plate 3. An Early Iron Age hill-fort: Maiden Castle, Dorset, from the air. The latest reconstruction of the defences was to resist the invasion of the Roman Emperor Claudius in A.D. 45. (Wheeler: *Maiden Castle, Dorset*. Society of Antiquaries, London, 1940).



Plate 4. Maiden Castle: section of ditch near E. entrance. (1) Turf-line over natural silt. (2) Artificial filling during a reconstruction of the defences. (3) Hut-floors of the Early Iron Age. (4) Dump of the Roman period. (5) Post-hole. (6) Early Iron Age metalled roadway. (Wheeler: *Maiden Castle*. Society of Antiquaries, London, 1940).



Plate 5. 'Varved' clays. A section, showing unusually thick and distinct varves, at Opava, Sudeten Mtns., Czechoslovakia. (Zeuner: *Dating the Past*. Methuen, 1946).



Plate 6. Barrow-circles and other traces of soil-disturbance near Eynsham, Oxon., shown by crop-marks in an air-photograph. (Grahame Clark: *Prehistoric England*, Batsford, 1941). (By permission of the Visitors of the Ashmolean Museum).



Plate 7. A Bronze-Age round barrow excavated in quadrants. (Excavated by W. F. Grimes, Director, London Museum). (*Proceedings of the Prehistoric Society*, 1938).



Plate 8. Rock-shelter at Les Eyzies, Dordogne, France. Later Old Stone Age man lived beneath the overhang.
(Déchelette: *Manuel d'Archéologie préhistorique* I. A. & J. Picard, Paris, 1908).



Plate 9. Little Woodbury, Wilt: the site of the Iron Age farmstead and its surrounding ditch as revealed in an air-photo. (Grahame Clark: *Prehistoric England*. Batsford, 1941.) (By courtesy of the Controller of H.M. Stationery Office and Director General of the Ordnance Survey).



Plate 10. A Neolithic ditch (A) at Maiden Castle, Dorset. (B) the old turf-line, denoting the level of the surface during the time between the silting up of the ditch and the raising of the Early Iron Age rampart over it, (D). A post-hole of the palisade on the defences is seen at (C). (Wheeler: *Maiden Castle*. Society of Antiquaries, London, 1940).

Plate 11. Two of the 34 Belgic skeletons discovered at Maiden Castle, casualties in the last defence of the fort against the Romans about A.D. 44. (Wheeler: *Maiden Castle*. Society of Antiquaries, London, 1940).



Plate 12. One form of Early Bronze Age 'beaker', with characteristic comb-impressed decoration. (Clark: *Prehistoric England*. Batsford, 1941). (Courtesy of University Museum of Archaeology and Ethnology, Cambridge).



Plate 13. Basket-work, knobbed piles and worked timbers, probably belonging to rectangular buildings, Glastonbury Lake Village, Somerset. (Glastonbury Lake Village I. Glastonbury Antiquarian Society, 1911). By kind permission of the authors, Dr. A. Bulleid, F.S.A., and Mr. H. St. George Gray, F.S.A.).



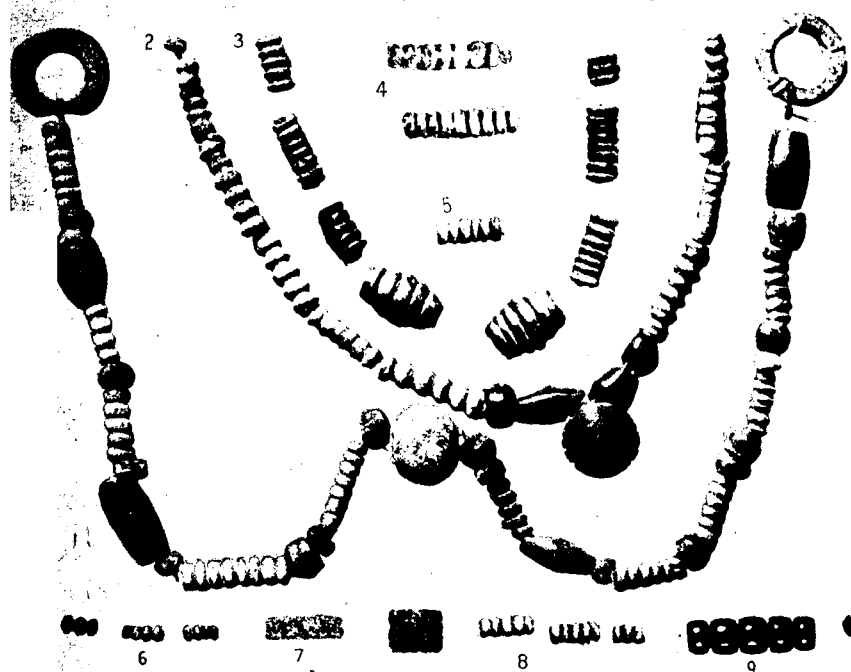


Plate 14. Faience and other segmented beads, mostly from Wiltshire. (Beck & Stone: *Archaeologia* LXXXV. Society of Antiquaries, London).

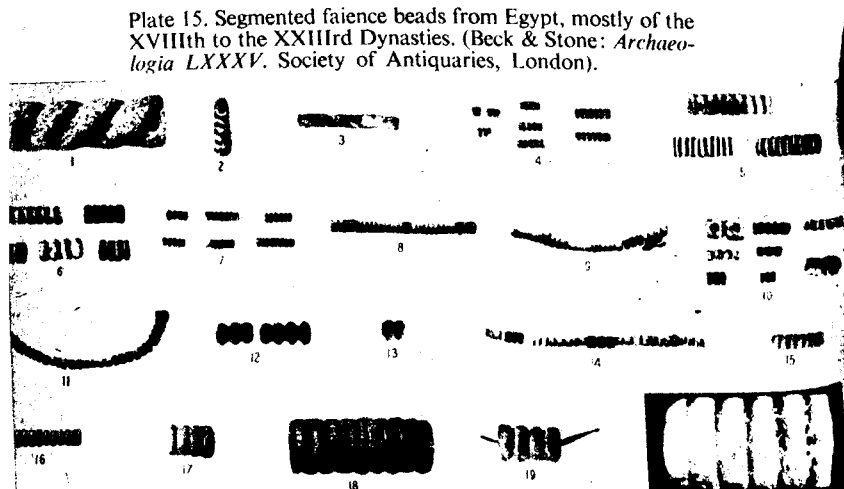


Plate 15. Segmented faience beads from Egypt, mostly of the XVIIIth to the XXIIIrd Dynasties. (Beck & Stone: *Archaeologia* LXXXV. Society of Antiquaries, London).

AUSTRALIAN SCIENTISTS' HOLIDAY EXPEDITION. Plates 16—19. A party of Melbourne scientists collected many interesting botanical, geological and ethnological specimens during a recent 5,000-mile holiday expedition from Port Lincoln in South Australia across the seldom-explored Nullarbor Plains to Perth (Western Australia). The party was led by Mr. Russell Grimwade, president of the Melbourne National Museum trustees. Specimens collected will enrich the Museum and the Melbourne Herbarium. (See page 128).



Plate 16. Sand dunes near Eucla, at the head of the Great Australian Bight, where the only vegetation is clumps of nitre bush. (Photograph by Mr. Russell Grimwade).

Plate 17. Entrance to one of the Nullarbor Caves, a geologists' paradise which, owing to its remoteness and lack of water, has been visited by only a handful of people. (Photograph by Mr. Russell Grimwade).





Plate 18. Curved stalactites composed of mineral gypsum instead of the usual calcite were obtained from the Nullarbor Caves. Such stalactites had been found previously only in caves in Kentucky and Southern France. Australian Official Photo: (Photograph by courtesy of the Melbourne National Museum).

Plate 19. A crude aboriginal chopping-axe of a type not previously found in Western Australia, of a period well before the coming of the white man to the country. Australian Official Photo. (Photograph by courtesy of the Melbourne National Museum).

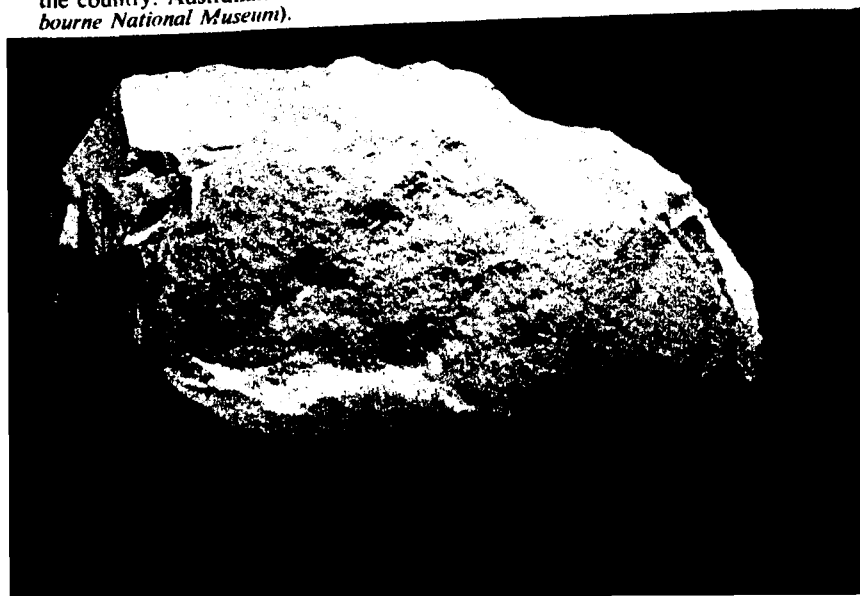


Plate 20. Photograph of horizontal strata at Filey, Yorkshire, showing layers of rock concentric with the centre of the earth, like the layers of an onion. Hence by extrapolation concentric layers deeper in the earth. Courtesy Leeds University. (See page 63. What the earth is made of)

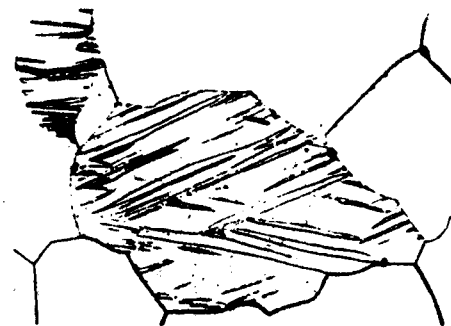


Plate 21. Pure magnesium as seen under the microscope at a magnification of 50. The parallel lines are indications of the presence of twinned crystals.

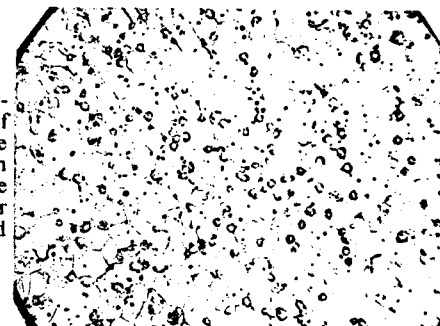


Plate 22. Magnesium containing 0.7 per cent. of zirconium. Note the change of grain size in comparison with Plate 21. (See page 135. Minor additions in metals and alloys).

CONTACT LENSES

Plate 23. Placing the plastic lens in position by means of a rubber holder. The lens is placed upside down on the holder, filled with fluid, raised to the eye as here, and when in place the holder closed, which releases the lens.



Plate 24. The plastic lens in position. Notice how it covers rather more of the eye-ball than the gap between the eyelids.



EYE GRAFTING

Plate 25. Right eye with corneal opacity before operation. (see Eye-Grafting, page 30). Notice how the pupil is hardly visible.



Plate 26. The same eye over a year after the operation in which new transparent cornea has been grafted in.



EYE GRAFTING TO-DAY

33

and *total in thickness* — involving all layers of the cornea; whence the name of *partial penetrating graft*.

Like the type of graft, the technique of implantation itself varies from author to author. We shall borrow the description of the Swiss masters, notably Franceschetti and Streiff, who have a following in France. We must consider in turn:

1. The preparation of the patient.
2. Cutting the graft.
3. Placing the graft on the diseased eye, and keeping it in place.
4. Post-operative care.

1. Preparation of the patient.

A penetrating graft always requires the most minute preparation of the recipient's eye. As well as the cases where, before the actual grafting, it is necessary to make a number of preparatory operations (for which the ophthalmologist alone can judge the need), in favourable cases also, where grafting is the one and only operation it is still very important to insist on a whole series of preventive measures.

From the first, perfect asepsis of the whole region is essential. The instillation of penicillin drops will thus be very valuable. Anaesthesia will be obtained with cocaine drops, and especially by the injection behind the eye of an anæsthetic solution of syncaïne, a compound allied to

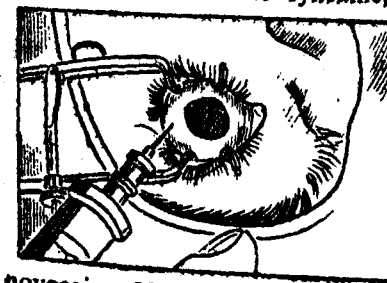


Fig. 8.—While the eyelids are held open by the retractor, local anaesthetic is injected into the eyeball. This is the left eye: note the nose on the right of the diagram, and the eyebrow below.

novocaine. Neighbouring parts (eyelids, skin) are likewise locally anaesthetised. Finally, two horizontal and two

vertical threads are stitched across the eye so that they cross and form a tiny square. At first they are left slack, but later when the graft has been put in place they are tightened and then form a little trellis to hold the graft in place. As for the risk of a rapid rise in pressure of the intraocular fluids after the operation, it is avoided by drops which lower the tension of the eyeball (eserine, etc.).

2. Cutting the graft.

Whether one uses a graft from the dead or one from the living, it is detached for preference from the enucleated eyeball: for this a trephine is used, a kind of tube with a very sharp end, the punch-movement of which is limited to about 1 mm., which corresponds to the usual thickness of

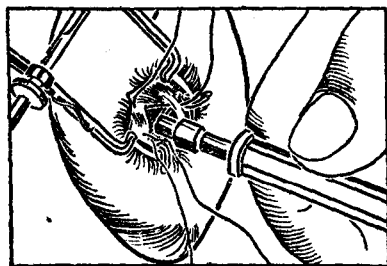


Fig. 9.—The punch removes a disc of opaque cornea.

the cornea. Its diameter varies between 3 and 6 mm. according to the size of corneal disc it is desired to cut, and which is to be taken from the centre of the cornea. Great care must be used in pushing the trepanned disc very gently out of the tube. It is placed on a little spatula, or better, on a flat circular drainer perforated with fine holes, and plunged into a glass dish filled with a physiological fluid or with serum from the future recipient of the graft.

3. Positioning.

A trephine with a slightly larger diameter is used on the diseased eye to remove an opaque disc from the centre of the cornea, taking great care to trepan absolutely per-

pendicularly to the iris, so as to obtain a complete section, clean and clear, and to limit the movement of the trephine in depth in order not to wound the lens. This is the most difficult point of the operation. To simplify it, some practitioners make use of a little wooden or metal spatula, which they slide behind the cornea into the anterior chamber of the eye. It then acts both as cutting board and protector.

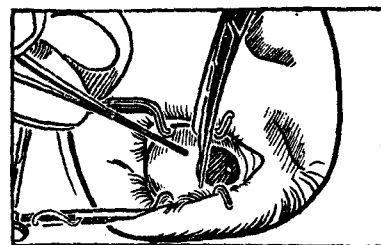


Fig. 10.—Cutting into the anterior chamber with a scalpel (right) to introduce the cutting plate (left).

The quality of the trepanning appears to be one of the essential elements in a successful outcome to the operation. Once the opaque disc is raised, and the bed clear for the graft, it is brought up on its back on a little spatula, slipped into the prepared bed and finds its right position with the help of pats from the spatula.

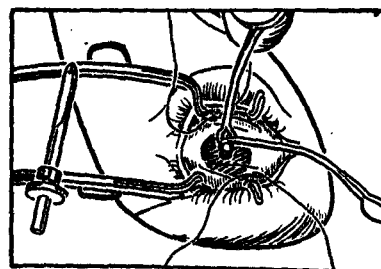


Fig. 11.—The new transparent cornea is lowered into place. Note the two sutures ready in position.

Some surgeons then apply a very fine membrane to the whole extent of the cornea, either eggshell lining or fine sausage skin, the dimensions of which are approximately those of the cornea. Then the threads are tightened (not

too much) in front of this membrane and tied. The whole keeps the graft well in place in a sufficiently firm manner.

Finally the introduction of eserine and penicillin ointment into the conjunctival cul-de-sac and a general grooming of both eyes completes the operation.

4. *Post-operative care.*

The first cleansing does not take place before the sixth day. Only the absence of ocular reactions is then confirmed, without even an examination of the graft. Before replacing the bandage, it is well to instil a few drops of eserine in oil, and a lamella of penicillin.

It is only towards the ninth or tenth day that the sausage membrane and stitches are removed, and one can make sure that the graft has remained properly in place and not become cloudy. If progress continues favourably without complications, the unoperated eye is freed of all bandages on the eleventh day and the patient allowed to get up.

In summary, it is necessary to reckon on about 15 to 20 days stay in hospital, with ten or eleven days immobilised in bed. In spite of undeniable progress in recent years, the operation remains delicate, and the results are not always equal to the trouble expended by both surgeon and patient.

What cases take the graft?

Corneal grafting is not, alas, a panacea, a remedy for all blindness. It is for a certain category of the blind: those with a clouded cornea, whether the opacity is deep or superficial, extensive or localised. The operation can thus be performed in all accidents to the cornea, e.g., burns (quicklime, cement, acid), or ulcers which have become infected and leave a residual speck or leucoma. Patients in whom, following a general infection, the interstitial layers of the cornea have become cloudy, benefit likewise; but because of operative difficulties and the ever-present risk of a hold-up, it seems wise to reserve corneal grafting in actual practice for cases where the specialist judges it to

offer excellent possibilities of improving the patient's vision, without making him run too great an operative risk. It is therefore the ophthalmic surgeon alone, after a complete eye examination, who can make the decision to operate.

Perfect cure, that is to say, regeneration of the corneal nerves into the graft, and its continued transparency, essential conditions for vision, usually require several months of convalescence and periodic examination. It is never guaranteed in advance (see plates 25 and 26).

Some grafts stay clear, others remain translucent, a certain number go wrong, and become cloudy, condemning themselves completely. The most recent figures given by very different authors offer almost identical values. About 20 to 25 per cent of cases are found in which the graft remains transparent, and 30 to 35 per cent translucent. Altogether, in 55 to 60 per cent of cases sight has been improved. These statistics are not miraculous, but they are much more than encouraging and ought to put a stop to sterile scepticism and create a real atmosphere of operative confidence. Step by step as our techniques improve, the percentage of our successes will increase. Let us offer homage to the worth of all pioneers, whatever their nationality, who by their patient labours and their faith have managed to push back the limits of blindness, and by giving sight, have given back hope and the joy of living to so many unfortunate blind people.

(translated by J. L. Crammer).

Contact Lenses

FRANK DICKINSON

CONTACT lenses have been 'in the news' again recently. They were worn by a well-known professional footballer, who paid high tribute to their efficiency.

Spectacles, always a handicap in most forms of sport, are being replaced increasingly by the newer types of plastic contact lenses, which enable the sportsman to enjoy normal vision in safety. The footballer featured in the news item is one of many who take advantage of this modern development in optical science.

The basic principle of the contact lens has been known for more than a century. Sir John Herschel, Astronomer Royal, suggested its practicability in 1827. Only in recent years, however, have the more serious limitations of this method been overcome, through experience gained in the fitting of many thousands of lenses in Europe and North America.

The lens is worn in close contact with the sclera (the white portion of the eyeball). It provides a new, optically perfect anterior surface for the cornea, without actually touching the sensitive corneal surface. This clearance of the cornea makes it possible to wear the lens without discomfort. The space between lens and cornea is filled with a fluid similar to the tears.

By means of contact lenses, focal errors of vision may be corrected with a degree of accuracy unequalled, in many instances, by spectacle lenses. This unique feature of 'contacts' is particularly valuable when the cornea has been damaged by disease or injury. In keratoconus, for example (a conical protrusion of the cornea) they offer by far the best means of visual improvement, giving almost miraculous results in many cases previously considered hopeless.

Unfortunately, problems other than those of optical correction obtrude themselves. The fashioning and fitting of the lenses demands a high degree of skill and dexterity. The delicate cornea may be injured by the wearing of an ill-fitting lens, and there are cases in which even a perfect fit fails to bestow the desired degree of comfort. The average period of use at a stretch is still somewhere between five and eight hours.

Perhaps the greatest limitation of the conventional form of lens is the gradual development of a cloudiness of vision similar in effect to that of a smoky atmosphere. The clouding, which begins in most cases after 2½ to 4 hours' use, is believed to be due mainly to the accumulation behind the lens of carbon dioxide, normally exhaled by the cornea into the atmosphere. Recent experiments by Dr Josef Dallos, one of the pioneers of contact lens technique, have proved that clouding can be largely eliminated by the provision of ventilation holes in the scleral portion of the lens.

Several distinctive methods are employed in the fitting of contact lenses. Many recognized authorities favour the moulded lens. The technique consists of the making of an impression of the eyeball, from which a positive cast is prepared, the lens being pressed from sheet plastic to conform to the curvature of the cast. Other practitioners prefer a lens of conical form, designed to fit a limited area of the surface of the eyeball. A much-favoured design in Great Britain is the ground spherical lens, originally introduced by Zeiss. All of these types, which are made from so-called plastic material (synthetic acrylic resin) permit a fluid lens-like zone between the contact lens and the cornea. Another kind of lens is inserted without fluid, being worn in direct, capillary contact with the cornea (see plates 23 and 24).

Although glass lenses are not dangerous, the majority of present-day contact lenses are made from I.C.I. plastic.

The advantages of contact lenses over spectacles include an unrestricted field of vision; security from displacement in sporting activities such as swimming, football and riding;

almost complete invisibility (an undoubted asset in social and stage life) and the visual benefits mentioned above. Among their disadvantages are comparatively higher cost, the limited period of use, and physical unsuitability of the eyes in certain cases.

Demography - Science and Administration

R. R. KUCZYNSKI

1. *The Beginnings of Demography.*

UNTIL ten years ago the word 'demography' was rarely used in this country. Population experts, it is true, had attended the International Congresses for Hygiene and Demography, and they studied the *Aperçus démographiques* published by the International Statistical Institute, but the *Concise Oxford Dictionary* still defined demography as 'statistics of births, diseases, etc., illustrating condition of communities,' while demography, as is evident from the word itself, aims primarily at a description of the population, using, of course, if available, censuses, birth, marriage, death, and migration statistics, and occasionally also statistics of diseases. However, in recent years 'demography' has been used more frequently and more correctly, and the University of London even appointed (in 1938) a Reader in Demography.

Demography, like every other science, has to do with methods and with results. A few scholars have been interested only in the development of methods by which the basic data may be correctly interpreted, and the demographer owes much to those pure mathematicians. Among demographers, who are all, of course, interested in interpreting results, a certain number have contributed to the development of methods.

The first man who ever made a demographic study was the London haberdasher John Graunt. His little book *Natural and Political Observations upon the Bills of Mortality*, published in 1662, was a remarkable achievement. His basic material consisted of the weekly London 'Bills of

Mortality' which showed baptisms by sex and burials by sex, age, and cause of death. He compiled all available Bills and analysed the content of his tables, which covered many decades, most thoroughly and with a vast amount of common sense. He even went so far as to construct a life table for London. This life table, of course, was most primitive and inaccurate – about as inaccurate, I should say, as the recent official life tables for India or for Lagos. His motive for 'deducing so many abstruse, and unexpected inferences out of these poor despised Bills of Mortality' was that 'the Art of Governing, and the true *Politiques*, is how to preserve the Subject in *Peace*, and *Plenty*.'¹¹ When Graunt published his pamphlet he had not the faintest idea that he had created a new science. Demographic data, he thought, could be useful only to the Administration. 'There seems to be good reason, why the Magistrate should himself take notice of the numbers of Burials, and Christenings, viz., to see, whether the City increase or decrease in people; whether it increase proportionably with the rest of the Nation; whether it be grown big enough, or too big, etc. But why the same should be made known to the People, otherwise than to please them as with a curiosity, I see not.'¹²

Graunt lived in an enlightened period. The great importance of his work was at once recognized both by the State and by Science. His name appears in the original list of Fellows of the Royal Society, which received its charter in 1662, and Sprat, in his History of the Society, speaks of 'the Recommendation which the King himself was pleased to make, of the judicious Author of the *Observations on the Bills of Mortality*: In whose Election, it was so far from being a Prejudice, that he was a Shop-keeper of London; that his Majesty gave this particular Charge to his Society, that if they found any more such Tradesmen, they should be sure to admit them all, without any more ado.'¹³

The original list of Fellows contained also the name of Sir William Petty, who wrote *Essays in Political Arithmetick* and

other demographic studies.⁴ An able disciple of his friend Graunt, he was both more knowledgeable and more brilliant, but his contribution to the development of methods was not important. Edmund Halley, on the other hand, who joined the Society a few decades later, was hardly interested in demography as such, and his whole demographic work covers only eighteen printed pages.⁵ His great contribution was the discovery (in 1692) of a most ingenious method of constructing a life table. He was not sufficiently explicit as to his data and the use he made of them, and his method has, therefore, been very often misunderstood. Many of his successors in various countries committed grave errors in trying to apply it, while others introduced slight improvements. But in a general way it may be said that Halley's method was for about 175 years the usual device for constructing life tables.

2. Science and administration.

England was the cradle of demography, and in the two centuries following the publication of Graunt's *Observations* England's contribution to the new science was equal to that of any other country. But the position has since changed. If a list were made of the hundred best demographers who set out on their careers during the last hundred years, it would be very difficult to find five Englishmen to include in it.

The causes of this stagnation in English demography can be easily discerned. In the first place, the number of people engaged in demographic research (full time or part time) is probably smaller now in relation to the population than it was fifty or a hundred years ago. On the Continent their number has increased by leaps and bounds. In England there is one Registrar-General's Office. On the Continent there are, apart from a Central Statistical Office for each country, numerous State and municipal statistical offices. Moreover, in the Registrar-General's Office priority is given to all matters dealing with registration; in the continental statistical offices population statistics are not at such a disadvan-

tage. A Registrar-General is rightly chosen according to his qualifications to function in this capacity, and he may never have dealt with population statistics before his appointment. On the Continent a civil servant becomes director of a statistical office after having worked for many years either in the same or another statistical office, and it is inconceivable that he be appointed without having had experience in population statistics. It may be argued that the Registrar-General has demographic assistants who are in charge of this section of his office's work. But experience shows that this offers no solution of the problem. The Registrar-General George Graham, in 1879, concluded his last annual report by saying: 'Lastly, I must express to Dr Farr, whom in 1842 I had the good fortune to find here presiding over the Statistical Branch, my grateful acknowledgment of the important services he has ever since continually rendered. He is acknowledged throughout Europe, the United States, East Indies and the Colonies as one of the first statist of the day. To his scientific researches and reports I attribute any reputation that may have accrued to the General Register Office of England and Wales from the time he accepted office in this Department.'⁶ All this was perfectly true. Dr Farr, in the Statistical Branch, made the best use of the basic data he got from the Registration Branch. But his efforts to improve those data failed. The birth registration form introduced in 1837, when the Registrar-General's Office was created, was quite defective. From a demographic viewpoint it was much less satisfactory than, for example, the birth registration form introduced in Sweden in 1775. Dr Farr became early aware of that. In his report for 1842 he wrote:

... no provision has yet been made for determining the simplest fundamental facts - the foundation of all reasoning on the subject - such as ... the ages of mothers of children ... Upon many of these points the greatest ignorance prevails, writers on population depending on rough approximations derived from scanty, imperfect, and often erroneous data, because the censuses and registers of Europe have not yet been taken and abstracted upon a comprehensive, well-considered plan.⁷

He made similar complaints over and over again. In his report for 1867 he said:

Two grave defects in the registers of the United Kingdom deprive them of much of their utility as pedigrees, and as records of facts for the solution of the great problems of population. Neither the age of mothers at the births of each of their children, nor the order of birth is recorded; so that the number of children borne by women at different ages, and in the course of their lives, cannot be ascertained. This defect was supplied in the first schedule of the Scotch Act, but the important parts of the schedule were unfortunately discontinued after 1855⁸.

It is pathetic to read in the Congress reports of the International Statistical Institute the appeals which Dr Farr made for the registration of the mother's age in all countries keeping birth records.⁹ His appeals were not altogether in vain. By 1921 England was practically the only civilized country where this question was not asked. The birth schedule of 1837 remained unchanged until 1938, when Parliament insisted on its reform. If Dr Farr had been Registrar-General - and it was the great disappointment of his life that he was never promoted to this position - the basic demographic data for many decades would have been as ample in England as in the British Dominions, the United States, France, Germany, Austria, Italy, the Scandinavian countries, etc., etc. As he was second in command his country derived only very limited benefits from the last prominent demographer it has produced. That he accomplished as much as he did was due to the fact that he was not only competent but also obsessed by an urge for research and a deep interest in demography. None of his successors combined these qualities.

Another cause of the deterioration of British demography is the lack of contact between administration and science. On the Continent many directors of statistical offices are at the same time university teachers; they sometimes hold seminars in their office and teach the students the whole process of statistical technique. It is quite usual for such students, after they have terminated their university studies, to be appointed assistants and later directors of municipal or other statistical offices, and to become in their turn

university teachers. Nothing of the kind exists in this country. The scholar interested in population problems usually knows very little of the technique of statistical administration; he cannot appraise the technical difficulties and the cost of carrying out the proposals which he may suggest to the administration; he cannot effectively answer the arguments of the civil servant. The administration, on the other hand, focuses its attention on its immediate needs and is, as a rule, reluctant to acknowledge specific legitimate needs of the demographic scholar. This, of course, does not affect the pure mathematician interested in methods but not in results. As a consequence thereof, the foreign reader of a review such as the *Journal of the Royal Statistical Society* must be struck by the high standard of the purely mathematical contributions dealing with a fictitious population, and by the scarcity of demographic papers. The barrier between administration and science deters the British scholar from demographic research; another deterrent is the low standard of the official vital statistics, compelling him to resort to foreign statistics which are difficult to understand without a knowledge of foreign conditions and foreign languages. The general level of demographic doctors' theses submitted to English universities is high, but a demographic study of the authors would probably reveal that comparatively few are born in this country.

3. *English census statistics.*

Censuses have been taken regularly every ten years from 1801 up to 1931. These censuses furnish valuable information to demographers, sociologists, economists, geographers and administrators. I shall deal here only with those data which are of special interest to the demographer. They refer to the sex and age composition of the population, to marital condition, and to fertility. The sex was ascertained from the first census onwards. The ages of the population were obtained in 1821 and again from 1841 onwards. The marital condition was asked for the first time in 1851, and thereafter

DEMOGRAPHY - SCIENCE AND ADMINISTRATION 47

at each subsequent census. Questions relating to fertility were put only in 1911.

At the early censuses the omission of persons was apparently frequent, but from 1841 on the censuses in England have probably been as complete as in any other country. This does not mean that the omissions are altogether negligible. There is, for example, a tendency everywhere to leave out young infants, and this tendency is strengthened in England by the habit of delaying the registration of births. Thus the General Report on the 1921 census states:

It may be observed that the population always comprises a large number of newly born infants whose births have not yet been registered, and that this number may easily have been as high as 50,000 or 60,000 at the date of the census. The attitude of mind which regards such children as not having been placed upon the official roll and not subject, therefore, to the census procedure is an intelligible one ...¹⁰

It is impossible, of course, to estimate even approximately the numerical importance of the omissions at various ages, but a guess that they aggregate something like one or two per 1,000 of the total population would probably not be far off the mark. They are offset in part by the erroneous inclusion of persons temporarily abroad and by double counting.

While, then, the English censuses may be considered fairly complete, the demographic details are not as accurate as might be expected. The mistakes in the returns on sex, to be sure, are undoubtedly few. But the age data are less satisfactory. There is in England, as everywhere, a tendency to report age in round numbers - that is, in numbers ending with 0. It may suffice to mention that, according to the 1931 census, there were 245,684 men aged 50, but only 215,999 aged 51; 280,182 women aged 50, but only 240,793 aged 51. Ages ending with the digit 7 are unpopular, while there is a marked predilection for ages ending in 8. According to the 1931 census, there were more men and more women aged 58 than aged 57, more men and women aged 48 than aged 47, more men and women aged 38 than 37, etc. That the age statistics of many other countries are not so

defective as those of England is due to the fact that in their census forms they do not ask for the age at the last birthday, but rather for the date of birth. It seems desirable that henceforth England should follow their lead in this respect.

Misstatements of marital condition are also quite frequent. The large excess of married women over married men which appears at every British census can be explained only in part by the fact that the number of married members of the Army, Royal Navy, and Merchant Service who are temporarily abroad exceeds the number of married foreigners who are temporarily in this country and have left their wives at home. It seems that tens of thousands of women report themselves as married though they are not. But a much graver defect is the inadequacy of the statistics of divorced people. Until 1921 all persons over 15 years were asked to state in the census forms whether they were single, married, or widowed. As to divorced persons, it was evidently left to their discretion which of these three alternatives they liked to choose, and if by any chance they were obstinate enough to reveal their actual marital condition they were probably counted as widowed. In 1921 a discreet attempt was made to identify divorced persons. The heading of the column relating to marital condition now read:

For persons aged 15 and over write 'Single,' 'Married,' 'Widowed,' or if marriage dissolved by divorce write 'D'.

The innovation was introduced with great reluctance and with the conviction that it would prove to be a failure. The report states:

At the 1921 census an attempt was made for the first time in this country to ascertain the number of divorced persons in the population ... The total number returned in this category amounted to 16,682 in all, of which 8,464 were males and 8,218 females ... It is greatly to be feared, however, that doubts as to the value of such returns, which were felt and expressed when it was first decided to include the inquiry in the general census questionnaire, have proved only too well founded, for from an examination of the records of the divorces which have been granted year by year, after making full allowance for reductions in the numbers by mortality and by a very high remarriage rate, the

expected numbers might well be put at a figure twice as large as the total recorded above, and it appears more than probable therefore that a large number of persons failed to return the desired information.¹¹

The number of men reported as divorced was then slightly greater than the number of women. Since the remarriage rate of divorced men is much higher than that of divorced women, misstatements of marital condition must have been particularly numerous among divorced women. According to the 1931 census the divorced males numbered 13,546 and the divorced females 19,169. It will be interesting to read the official comment on these figures in the General Report, which has not been published as yet.

From a demographic standpoint the 1911 census is still to-day the most valuable of all because it included questions concerning fertility. Each married woman had to state the number of years the present marriage had lasted, the number of children born alive to the present marriage, the number of children still living, and the number of children who had died. The question about the duration of marriage was answered incorrectly by a very large number of wives. The tendency to concentrate on round numbers - 10, 20, 30, 40 - was even more pronounced in the case of marriage duration than in the statement of ages. Moreover, misstatements of short durations were exceedingly numerous. The ratio of couples returned as married 0-1, 1-2, and 2-3 years to the number of marriages concluded in the three years preceding the census was 68, 85, and 89 per cent respectively, the understatement of the number of couples returned as married under one year being mainly due to a desire to conceal ante-nuptial conception. The instruction to return only the children of the present marriage was often disregarded. On the other hand, many mothers seem to have omitted in their statements children who died young.¹²

The first census embodying fertility questions had been taken in 1875 in Massachusetts, and many other countries followed suit. The decline in the birth rate made, of course,

the demand for such enquiries ever more urgent, and between 1920 and 1936 alone fertility censuses were taken in France (1921, 1926, 1931, and 1936), Holland (1920 and 1930), Norway (1920 and 1930), Hungary (1920 and 1930), Estonia (1922 and 1934), Spain (1920 and 1930), Italy (1931), Czechoslovakia (1930), Germany (1933), the Union of South Africa (1921 and 1926), Southern Rhodesia (1921 and 1926), Northern Rhodesia (1921 and 1931), Australia (1921), New Zealand (1921 and 1936). But in this country no such enquiry has been made since 1911.

Another grave defect which must be mentioned in this connection is the date of publication of the English census reports. The fertility enquiry of 1911 was made throughout the United Kingdom. The complete reports for Scotland and Ireland were both published in 1913. For England and Wales a first volume, containing a number of basic tables, was published in 1917; the second volume, containing all other tables and the text, was published in 1923. I have already mentioned that the text volume of the 1931 census is not yet available. The volumes published so far contain merely tables. This, it seems to me, is an intolerable situation which should be remedied and could be remedied easily. I happened to be connected with the United States Census of 1900. The publication of the 1890 census reports had been very slow. But Congress was not willing to run a similar risk again. The Census Act of March 3rd, 1899, provided therefore:

The only volumes that shall be prepared and published in connection with the Twelfth Census, except the Special Reports hereinafter provided for, shall relate to population, mortality and vital statistics, the products of agriculture, and of manufacturing and mechanical establishments ... and shall be designated as and constitute the Census Reports, which said reports shall be published not later than the first day of July, nineteen hundred and two.

We then had barely two years for the preparation and publication of the reports, not only of a most elaborate population census but also of a most intricate census of agriculture and of manufactures, in a country with a popula-

tion nearly twice that of England and Wales and with an area more than fifty times as large. But we knew that we had to do it, and all the volumes, tables and text, covering nearly 10,000 large, closely printed pages, were published within the prescribed time limit.

4. *English vital statistics.*

Civil registration of births, marriages, and deaths was started on July 1st, 1837. The records are probably as complete as in any other country. Birth registration, to be sure, was somewhat defective prior to the introduction of compulsory registration in 1875, but it is safe to say that nearly all children born thereafter have been registered, though often with great delay. Death registration was almost complete from the outset. Marriage statistics, of course, are absolutely complete, since, while people may be born or may die without being registered, there is no marriage without registration. As to accuracy the vital statistics suffer from defects similar to those of the census statistics. The age at death is too often reported in numbers ending with 0. There is also, at least among middle-aged persons, a predilection for ages ending in the digit 2. Here again the remedy lies in asking for the date of birth instead of the age in years.

Until the new Population (Statistics) Act came into force, the Registrar-General's *Statistical Review* showed merely the number of male and female, legitimate and illegitimate live births and still births, registered in each quarter. From July 1st, 1938, on the *Statistical Review* shows in addition the births by order of birth, age of mother, and duration of marriage. The *Decennial Supplement* gives in addition data about fertility by occupation in the period centring around the census. The *Supplement* dealing with the year 1921 was published in 1927; the volume which is to cover the years 1930 to 1932 is not available as yet. But the records of the father's occupation are not the only ones of which little use has been made. From 1837 on, the registration form recorded the date of birth and the date of registration. But in the

published birth statistics no account (until 1944) was ever taken of the date of birth. We know the number of births registered in a given year, but we do not know the number of births which occurred in that year. According to a statement published year-in year-out in the Registrar-General's *Statistical Review* 'the average time lag between occurrence and registration is usually about a month.' This seems to indicate that the average time lag is usually though not always about a month. But even if the average time lag were always exactly one month, the difference between the number of registered and of actual births might vary considerably from year to year. It may well be, therefore, that even in peacetime the changes in the official birth figures sometimes convey a distorted picture of the actual changes in natality. The number of registered births dropped from 613,972 in 1932, to 580,413 in 1933, and rose again to 597,642 in 1934. The number of registrations in the last quarter of 1933 (129,810) was over 10,000 lower than in any other quarter from 1850 to this day. It is possible, of course, that the number of births was actually as low in the last quarter of 1933 as indicated by the official statistics. But until there is proof to the contrary it seems safer to assume that the drop in the actual number of births in 1933 was smaller than the drop in the number of births registered in that year. The official birth rates* in 1933-7 were 14.4, 14.8, 14.7, 14.8, and 14.9. It may well be that the actual birth rate in 1933 was not lower than in 1935. And which year had the highest birth rate is anybody's guess.

In periods in which rationing of food induces parents to hasten birth registration or de-rationing of food removes this incentive the official birth figures, of course, are absolutely misleading. From 1918 to 1919 the number of registered births (which alone appeared in the official statistics) increased from 662,661 to 692,438 or by only 29,777. Actually, the number of births registered in 1918 was much

* For explanation of how these and other values are calculated, see *Glossary*.

larger than the number of births that occurred in that year while the reverse was true in 1919, and the increase in the number that occurred has been estimated at three times the increase in the number of births that were registered. In 1938-40, the numbers of births registered were 621,204, 619,352, and 607,029 respectively. How many births actually occurred in 1938 it is impossible to tell. In 1939 and 1940 - according to the *Statistical Review* for 1940 (published in 1944) - the numbers were 614,479 and 590,120. It is most welcome that the Registrar-General, who naturally is more interested in birth registration than in birth statistics, has at last yielded to the demographers' urgent demand for dealing with births according to the date of the event, and has published the numbers of male and female legitimate and illegitimate live and still births which occurred in 1939 and 1940. But the statistics for those years, showing order of birth and duration of marriage, unfortunately still refer to registered births and, therefore, do not convey a true picture of the decline in fertility during the first period of the war. These statistics include for 1939 54,144 live births which occurred in 1938 but were registered only in 1939, while they exclude 49,271 which occurred in 1939 and were registered in 1940. The statistics for 1940 include those 49,271 births, but exclude 32,362 which occurred in 1940 and were registered in 1941.

The deaths up to now have likewise been shown by date of registration. But here the consequences are less serious. According to a statement published each year in the Registrar-General's *Statistical Review* 'the time lag between occurrence and registration is usually only a day or two.' It would seem, however, that the exceptions to this rule are very numerous. Of the twenty-three tables on mortality embodied in the *Statistical Review* one shows the deaths by months of occurrence. A comparison with the returns by date of registration reveals that the differences are sometimes pretty large. From the third to the fourth quarter of 1935, the number of registered deaths increased by 22,677,

or 22.7 per cent, while the number of actual deaths increased by 24,785, or 25.1 per cent. It seems desirable, therefore, that henceforth not only the births but also the deaths be classified throughout by date of occurrence.

The Population (Statistics) Act introduced only some questions of minor importance to be asked on the registration of death. But the collection of details about deaths has been fairly ample for many years, and the mortality statistics of England compare favourably with those of other countries. I would even go so far as to plead for the curtailment of the published tables in one respect. I would advocate that the publication of standardized death rates be stopped. The standardized death rate, which is meant to convey a better picture of the trend of mortality than the crude death rate, has decreased from 20 in 1876-80 to 9 in 1935-9. The crude death rate has decreased only from 21 to 12. This crude rate is no adequate gauge of the trend of mortality, as it is calculated without regard to the changes in the age composition of the population. Since the age composition is now more favourable than sixty years ago, the crude death rate overstates the decline of mortality. The official standardized death rates are those which would have been recorded if the age composition of the population had been the same as in 1901. All depends, of course, on what year is taken as a standard, and it so happens that by choosing the year 1901 the decline of mortality in the last sixty years is overstated still more than by a comparison of the crude death rates. To compute what the death rates would have been sixty years ago and what they would be now if the composition of the population by age at both dates had been that of 1901 is perfectly futile. It is just as futile as to compute what would have been the death rates then and to-day if, both in the late seventies and now, the composition of the population by occupation had been the same as in 1901. But the computation of such standardized death rates is not only futile, it is a nuisance. Nothing, I am convinced, has contributed so much to conceal the actual trend of mortality in England as the

regular publication of these standardized death rates. The correct death rate derived from the life tables has decreased in the last sixty years from 23 to 16 or only by about 30 per cent. England, which in the 1880's had a lower mortality than, for example, Germany, Holland, and Switzerland, had in the 1930's a higher mortality than those countries. The time that is wasted in computing misleading standardized death rates could not be better used than in studying why the progress achieved in lowering mortality in England lags behind what has been accomplished in a good many other countries.

The position is particularly unsatisfactory as regards life tables. Thirty years ago such tables were available for 1838-54, 1871-80, 1881-90, 1891-1900, 1901-10, and 1910-12. Since then tables have been published only for 1920-2 and 1930-2. While the earlier tables comprised all years from 1871 to 1912, the recent ones cover only six of the thirty-two years elapsed since 1912. Formerly the bad years were taken with the good years. It is obvious that the results of the new procedure, which completely ignores not only the war years but also such a bad peace year as 1929, are much less conclusive. The existing life tables should be supplemented without delay by a life table for females for 1911-20, and by tables for each sex for 1921-30 and 1931-9. The necessary data are all available in the Registrar-General's Office, and by using modern short-cut methods which yield results sufficiently accurate for all practical and scientific purposes a life table can be computed in a few hours.

The marriage statistics in England are not as ample as the death statistics but they are not as meagre as the birth statistics were until 1938. As in the case of the death statistics it is not so much the tables showing the basic figures which need modification but rather the published rates and averages. The Registrar-General's *Statistical Review* gives the marriage rates of (a) bachelors and divorced men, (b) widowers, (c) spinsters and divorced women, and (d) widows. This lumping together of single and divorced persons was immaterial sixty or seventy years ago when the

marrying divorced persons numbered only 100 a year and constituted only 0.2 per cent of the remarrying persons, but it is most confusing to-day, since in 1939 no fewer than 10,698 or 20 per cent of the remarrying persons had been divorced. One average which is particularly misleading and which is published every year in the *Statistical Review*, is the average age at marriage. This figure was rising before the war and gave the impression that people nowadays marry¹² later than in former times. But it was rising only because the population was ageing. It is obvious that in countries where owing, for example, to a heavy decrease in fertility the proportion of younger spinsters decreases, the average age at marriage may increase, even if the nuptiality of the younger spinsters rises while that of the older spinsters declines. Since 1931 the frequency of marriages in England has increased enormously. Yet the average age at marriage of spinsters rose in every year up to 1937. No wonder that people who see those figures conclude that the increase in the number of marriages was largely due to the contraction of marriages that had been postponed during the depression. But a correct computation of the mean age at marriage of spinsters, that is of the mean expectation of single life at birth for those who eventually marry, shows that it has actually decreased in every year since 1931. The additional marriages concluded in that period were then early marriages, not delayed marriages.

The correct mean age at marriage can be derived only from a nuptiality table which is constructed according to the same principles as a life table, and which shows the probability of marrying. Many official nuptiality tables have been computed in the course of the last forty years in foreign countries, but none in England. Our knowledge of the frequency of marriages, in so far as it is based on official documents, is, therefore, quite defective. As in the case of life tables all necessary data are available in the Registrar-General's Office, and nothing could destroy as quickly the myth that our population problem may be

solved by an increase in marriages as the publication of official nuptiality tables.

5. Official statistics and population problems.

In February, 1944, the Government set up a Royal Commission on Population. Its first term of reference is 'to examine the facts relating to the present population trends in Great Britain.' Let us assume that, on being appointed, a member of the Commission would have tried to examine the facts by consulting all published official statistics. He would have found that the most recent general census available was the one of 1931 and that the covering report was not yet prepared; that no fertility inquiry had been made since 1911; that all birth statistics were based on the date of registration and were therefore somewhat uncertain; that the most recent statistics of the father's occupation were those of 1921; that no data collected under the Population (Statistics) Act of 1938 had been published so far. If he had looked for gross and net reproduction rates or for nuptiality tables, he would have seen that such rates or tables have never been published and never even been mentioned. If he had asked for any official publication on population trends he would have been given the Registrar General's White Paper *Current Trend of Population in Great Britain*, of which Earl De La Warr said in the House of Lords on June 8th, 1943: 'This Report, if I may say so, is completely misleading and therefore extremely dangerous. I cannot see the reason or excuse for a Report of this character unless it is just an attempt to allay public concern on a problem that is in fact dangerous to the extent that the public do not recognize it.' If he had pointed out that the Government in that same discussion had announced that the Minister of Health 'proposes to publish a document on the lines of the recent White Paper, in which the statistical outlook would be set out and discussed in a reasoned and balanced manner,' he would have been told that the plan to publish such a document had been abandoned.

Since the appointment of the Royal Commission the position has improved considerably.

1. The results of the emergency census taken for purposes of national registration on September 29th, 1939, were published in the spring of 1944. The report shows the civilian population by sex, age, and marital condition. The returns are less trustworthy than those obtained at the general censuses, and the data for men are particularly defective inasmuch as they exclude (for England and Wales) about 900,000 non-civilians; but our knowledge of the age composition and the marital conditions of females has become much more up to date.

2. The Government published in the summer of 1944 data for 1940 collected under the Population (Statistics) Act; similar figures for 1939 were published in the autumn of 1944, and shortly thereafter for the second half of 1938.¹³ The tables for 1938 are accompanied by a covering report which shows the gross and net reproduction rates for 1938. Unfortunately the text is written in the same vein as the White Paper on current trends of population. The Registrar-General evidently thought that the main object of such a report was to fight what his Medical Statistical Officer called "that lamentable sense of national inferiority from which a large section of the population now seems to be suffering."¹⁴ The essence of his argument is that the population problem is not serious since a population decline could be prevented by encouraging marriage and particularly early marriage. For 1939 and 1940 only tables have been published, but no rates and no text. It is to be feared, therefore, that in their present state these statistics will be of little use to the members of the Royal Commission. However, the basic data are now available, and it should not cause insurmountable difficulties to relate them to other data hidden in the Registrar-General's Office and to provide an unbiased interpretation of the figures.

3. The Registrar-General, in the spring of 1944, published for 1934-43 'approximate reproduction rates, corresponding

to the births which occurred in each year and making allowance for a continuing improvement in survivorship conditions.'¹⁵ In his *Statistical Review* for 1938 he had included a net reproduction rate computed according to the principles accepted all over the world (and adopted by the League of Nations and the International Statistical Institute), which imply that such rates should be computed on the basis of current fertility and mortality. His new rates take no account whatever of current mortality. Although in 1940 fertility was lower and mortality was much higher than in 1935, he obtains for 1940 a higher reproduction rate than for 1935 because, irrespective of the facts, he assumes year-in year-out a steady gradual 'improvement in survivorship conditions.' These rates are not comparable with those of any other country; they are speculative because all estimates of future survivorship conditions are uncertain; they tend to overstate reproduction because they disregard fluctuations in mortality. But if they were recomputed on the basis of current mortality they would be most useful.

A start, no doubt, has been made in the course of the last few years to increase our knowledge of the demographic position of England. But much has still to be done in order to enable the Royal Commission 'to examine the facts relating to the present population trends.' A few things have been suggested before, such as the computation of life tables and nuptiality tables, of gross reproduction rates and of genuine net reproduction rates, and a proper analysis of the statistics collected under the Population (Statistics) Act up to 1944. But the most important task of all is the immediate taking of a special fertility (or family) census which would fill many gaps caused by the numerous sins of omission committed in the course of the last decades. Such a fertility census would convey a comprehensive picture of essential demographic facts upon which (to use Farr's words of 1844) 'the greatest ignorance prevails,' for example, the incidence of childlessness, the spacing of births, fertility differences between occupational and social groups,

and the distribution of families of various sizes. The information to be derived from such a census is indispensable for a serious study of the population problem and for the framing of an adequate population policy.

6. *Future of Administration and Science.*

It may well be that in some very small countries population and vital statistics do not suffer if they are prepared in a Registrar-General's Office, but this combination is to be found in no large country except England, and it has yielded here quite unsatisfactory results. It seems, therefore, advisable to create an independent office dealing with population and vital statistics. Whether such an office should also be put in charge of some other related statistics is not of decisive importance. *The Economist* of October 7th, 1944, contained the following suggestion:

It is already contemplated to take the census of production each year after the war; formerly it was taken every five years. There is a strong case for taking a census of distribution at frequent intervals, perhaps every two years. Similarly, the interval between population censuses has been far too long in the past; it might well be reduced from ten to five years. But if the number of such statistical investigations is to be increased and if they are to be carried out at short intervals, the existing facilities at the disposal of the Government will have to be overhauled. It might be an advantage, for example, to follow American practice by creating an organisation analogous to the Bureau of the Census, that is by making a single organisation responsible for all censuses. The creation of a strong permanent organisation would not merely simplify the procedure, but it should make for speed in the assembly and publication of statistics.

The American Bureau of Census is in fact in charge not only of population censuses but also of vital statistics, and an organization analogous to this Bureau may prove as beneficial here as it has proved in the United States. The only reasonable objection which may be raised against the proposed change is that it would *not necessarily* 'make for speed in the assembly and publication of statistics.' But it is an incontrovertible fact that for several decades publication of population statistics in this country has been slower

than in any British Dominion or Colony or in any foreign country. An outstanding example of speed in publication is to be found in India. The area covered by the Indian census of 1931 was thirty times as large as that of England and Wales, and the population was nine times as large. Yet, the forty-nine volumes of that Indian census which included an admirable text of several hundred thousand words were published within thirty months after census date. Would anyone suggest that it is easier to take a census in India, that it is easier there to instruct the supervisors and the enumerators (2,000,000 as compared with 40,000 in England), to assemble the filled-up forms at headquarters, or to train the necessary clerical staff for coding, punching, and attending the electric tabulating machines? And would anyone suggest that with the present organization it would be possible here to achieve what has been achieved in India?

Difficulties, it is true, may arise in finding the adequate personnel for the new organization. The leading men should be competent, unbiased, eager, and preferably young. It is here that Science can play an important part. Our universities should establish chairs for demography, and the students of demography should be thoroughly trained in statistics. This training would enable them to find a position as statisticians if they fail to secure employment as demographers. But the demand for demographers will be considerable in the future. The increasing importance of the population problem will afford many opportunities of doing useful work in this country, and the great development schemes in the Colonies cannot be carried out effectively without the assistance of expert demographers.

In the field of demography, as in many other fields, the Administration needs the co-operation of Science. But there is hardly any other field in which Science needs so much the co-operation of the Administration. The demographic scholar can discover methods by which the basic demographic data may be correctly interpreted but he himself cannot collect those data. They can be collected only by

the Administration, and his chances of getting the data he needs will obviously be much greater if he contributes towards the education of the future personnel of the Administration.

NOTES

1. *Natural and Political Observations*, p. 72. London, 1662.
2. See *ibid.*, p. 12.
3. Thomas Sprat, *The History of the Royal Society of London, for the Improving of Natural Knowledge*, 3rd ed., p. 67, London, 1722.
4. See *The Economic Writings of Sir William Petty*, ed. by Charles Henry Hull. 2 vols. Cambridge, 1899.
5. 'An Estimate of the Degrees of the Mortality of Mankind, drawn from curious Tables of the Births and Funerals at the City of Breslaw,' *Philosophical Transactions*, vol. 17, No. 196, Jan. 1693, pp. 596-610; 'Some further Considerations on the Breslaw Bills of Mortality,' *ibid.*, No. 198, March 1693, pp. 654-6.
6. *40th Annual Report of the Registrar-General of Births, Deaths, and Marriages in England* (1877), p. xl.
7. *6th Report* (1842), p. xxvii.
8. *30th Report* (1867), p. 222. See also *14th Report* (1851), p. xiii; *16th Report* (1853), p. 4; *27th Report* (1864), pp. xix-xx; *Supplement to 35th Report* (1872), p. xi.
9. See, for example, *Congrès International de Statistique à la Haye, Compte-rendu des travaux de la septième session, seconde partie*, p. 533. The Hague, 1870.
10. *Census of England and Wales, 1921, General Report with Appendices*, p. 78. London, 1927.
11. *Ibid.*, p. 84.
12. See R. R. Kuczynski, *The Measurement of Population Growth*, pp. 83-90. London, 1935.
13. See *Statistical Review for the Year 1938, Tables, Part II, Civil*; same for 1939 and for 1940. London, 1944.
14. *British Medical Journal*, March 21st, 1942.
15. See *Births, Deaths and Marriages registered in the Quarter ended 31st December, 1943*, p. 1.

Reprinted from *The Eugenics Review*, 37, 1 (April, 1945).

What the Earth is made of

ERNEST TILLOTSON

Introduction.

THE earth is nearly a sphere having an equatorial diameter of 12,756·776 kilometres (7,927·06 miles) and a polar diameter of 12,713·7 kilometres (7,900·29 miles). It has a mass of $5\cdot98 \times 10^{24}$ kilograms (5,870 million million million tons) and it thus weighs 5·5168 times as much as it would if it were made solely of water. The corresponding ratio for average surface rocks is 2·6, and thus we see that the interior of the earth must be made of material much more dense ('heavy' when we take unit volume). Our knowledge of the interior of the earth is obtained by inferences from observations and experiments made within the uppermost 4 or 5 kilometres (3 miles) of thickness or depth. In our task of finding what the earth is made of we therefore proceed by making a mental model of the earth, the parts of the model being constructed from our observations and deductions from experiments. We can then amend this model if necessary as new facts come to light. That we do alter it from time to time and that we are still prepared to do so, and to discuss alternatives, indicates that our present model is not perfect. Its lack of perfection is caused by our observations being incomplete, and some of our deductions faulty. Let us examine the 1948 earth model, and the observations and deductions used in the making of it.

A glance at a cliff-face shows the rock to be arranged in layers and these layers to be concentric with the earth's centre. It is not, therefore, a wild flight of fancy to imagine the earth to be made up of concentric spherical shells, something like an onion, these shells increasing in density as we go towards the centre (see plate 20). The next obvious

step is to try to discover how thick these shells are individually, and of what kind of material they are likely to be made. Let us first look at the outer 4 or 5 kilometres thickness.

What the geologist has to say.

When the soil and sub-soil have been cleared away from part of the earth's surface by wind, streams, glaciers or other natural agencies, or by the hand of man as in quarries and mines, we can see the rocks. Rocks may be classified into three types, sedimentary, igneous and metamorphic. Sedimentary rocks are those which have been deposited by the agency of wind or water; igneous rocks have been formed by volcanic action, that is they have been molten, and metamorphic rocks are either of the above altered by heat or pressure or both. Sedimentary rocks may be recognised by examining the individual mineral grains of which the rock is composed, when these are seen to be water- or wind-worn, and by the presence in sedimentary rocks of the remains of the fishes, animals or plants which were living at the time when the rocks were deposited and which are now found as fossils. By examining these fossil remains and by noting carefully the relative positions of undisturbed horizontal beds as they occur one above the other we can distinguish a sequence, which, in the British Isles has been worked out, and the groups of rocks given names as follows: at the base of the sequence and therefore the oldest sedimentary rocks, pre-Cambrian, and above this in order as the rocks were deposited more and more recently, Cambrian, Ordovician, Silurian, Devonian, Carboniferous, Permian, Triassic, Jurassic, Cretaceous, Eocene and Oligocene, Pliocene and Recent. Igneous rocks may be recognised in that the individual mineral grains have not been rounded, the rocks in mass are not arranged in layers or strata, but they have been intruded into the surrounding sedimentary rocks and they do not contain fossils. The character of metamorphic rocks is indicated by their position amongst other rocks, i.e., near intensely contorted strata or igneous in-

trusions, and by their appearance in hand specimens or in thin slices under the microscope. If at one time they were sedimentary and contained fossils these will have been altered beyond recognition. The examination of the individual mineral grains singly or in thin sections under a microscope assists identification.

The various minerals needed for our civilisation are found concentrated in ore bodies, which makes commercial exploitation possible. The presence of these minerals in the crust as a whole is often in very small proportions. For example iron is present to the extent of about 5 per cent by weight of the whole crust, copper 0.01 per cent and lead 0.002 per cent. The concentration of minerals into ore bodies is often by volcanic action, but occasionally by water sorting or other natural process.

The thicknesses of the individual strata vary greatly from one to another, and even the same stratum is of different thickness from place to place. Sometimes one or more strata may be entirely absent, causing a non-sequence or unconformity at that point. At the thickest places as in the Alps, Andes, Himalayas and Rocky Mountains the total thickness of sedimentary rocks may be 4 kilometres (2½ miles) whilst under the Pacific Ocean it is supposed that they are entirely absent.

What the geophysical prospector tells us.

When a geologist examines the stratified rocks at the earth's surface and finds them dipping or their bedding planes sloping into the ground he can take the angle of dip and with a knowledge of the topography draw a section to show what happens to them underground. In this he is assisted by other observations of other strata in the district, and possibly by the reappearance of the original strata (recognised by fossil assemblage) further away. This is called extrapolating from known data to unknown regions. The only real proof that his extrapolations are justified is by drilling and examining the cores so obtained. This is very

expensive and nowadays there are cheaper ways of testing geological inferences obtained in the above manner. These less expensive methods are called geophysical prospecting, because we make use of the physical properties of the rocks and examine them with the apparatus which physicists normally use in their laboratories and observatories. The four chief methods are magnetic, electrical, gravitational and seismic. We will examine these methods in some detail, particularly the seismic method, as they are also useful for the larger work of probing the deep interior of the earth.

The magnetic method depends upon the magnetic susceptibilities of various minerals in the rocks. The presence of deposits of iron, nickel or cobalt ore can readily be detected by this method even if they are buried to a depth of over 1,500 metres (5,000 feet). Ores which occur in association with, say, magnetite or pyrrhotite can also easily be located by this means, though the percentage of iron present in a mineral is not the only criterion of its magnetic susceptibility. Occasionally certain igneous rocks such as basalt, diabase, diorite or serpentine are more strongly magnetic than, say, haematite, which contains a higher percentage of iron. In magnetic surveys the vertical component of the earth's magnetic field at a given point is usually found, as this in practice is more sensitive for the purpose of locating underground deposits or determining tectonic structures than finding the horizontal component. In most instruments of the type used in the field this vertical magnetic force is compared with an opposing force in the instrument due to, say, gravity, torsional or bifilar suspension or the magnetic force of an auxiliary magnet. The magnetic survey is the quickest and cheapest of all surveys. It does not need a trained observer and is independent of topography. Its limitations are due to the fact that only some rocks have magnetic properties. It has proved extremely useful in such countries as Sweden.

The gravitational method of obtaining evidence concerning underground structures depends on the variations of

density of the rocks as we go from place to place. The work is not carried out by means of a pendulum which would measure the absolute value of gravity in any place, but usually by means of a torsion balance which measures horizontal *variations* in gravity. The first effective torsion balance was constructed by Baron Roland von Eötvös, Professor of Physics at the University of Budapest, in 1888. One type of the Eötvös torsion balance consisted of a light horizontal beam with equal masses of platinum at each end – one actually at one end of the beam and the other suspended at a depth of about a metre from the other end of the beam. The system was suspended from the mid-point of the beam by means of a delicate suspension wire, the torsion in which acted in opposition to the tendency of the earth to rotate the beam into one definite position. The suspended beam carried a mirror by means of which, in association with a scale mounted on a separate stand, a telescope could be used to observe changes in the position of equilibrium of the beam. The greatest success of the instrument lies in delineating the extent of large bodies of very high or very low density when these occur fairly near the surface. The boundaries of salt domes, associated in some parts of the world with petroleum deposits, can be mapped efficiently, after the necessary corrections for the effects of surface features have been applied.

Electrical methods have been used to determine the depths and dips of hidden layers of rock and are based on the differences in electrical conductivity between adjacent rock bodies. These differences may be very large and hence there is a greater range in this method than in any of the other methods. The oxidation of certain sulphide ores produces natural earth currents, but in other cases it is necessary to set artificial electric currents in the earth and ascertain their effects.

The seismic method of prospecting is probably the one which is capable of the greatest accuracy to-day and the one which has the greatest possibilities of advancement in

the future. The cost is fairly high but it pays dividends by its efficiency and it is much cheaper than drilling. Relatively simple geological structures such as salt domes, anticlines, and horizontal strata have been dealt with successfully up to the present time. The earthquake waves are produced artificially at a known instant by firing electrically a quantity of explosive. In the neighbourhood of the Gulf of Mexico, when prospecting for salt domes, 70 to 90 kilograms (150 to 200 lb.) of T.N.T. are buried 4 to 6 metres (14-20 feet) underground and fired electrically. In that part of the world the hole is sometimes left unfilled. In England and more populated areas less explosive is employed, though the people of Leyland in Lancashire thought they had experienced a natural earthquake when a shot was fired in a 30 metre (100 feet) borehole in the neighbourhood when prospecting for oil on Sunday morning, 21 July, 1946. The action of the explosive is to generate at least two types of waves. In the first type, called primary (P), the particles move backwards and forwards in the direction of the wave, which is thus longitudinal. It is of the same nature as a sound wave though often the frequency of the wave is such that human beings cannot hear it, and an instrument is required for its detection and reception. The second type of wave is called secondary (S), and as the wave passes any point the particles in its path move to and fro at right angles to the path. It is thus a transverse wave, and is slower than the P wave. Moreover it must be transmitted through solids. Liquids and gases cannot transmit S waves. In seismic prospecting at present P waves only are used, though in future S waves may be employed. The instruments used to detect these waves, which are similar to two waves present in natural earthquakes, are called seismographs and the records obtained by them seismograms. In principle, the horizontal seismograph consists of a weight fixed at one end of a nearly horizontal rod, which is pivoted at the other end. When the earth quakes everything moves except the weight, which stands relatively still. The relative movement

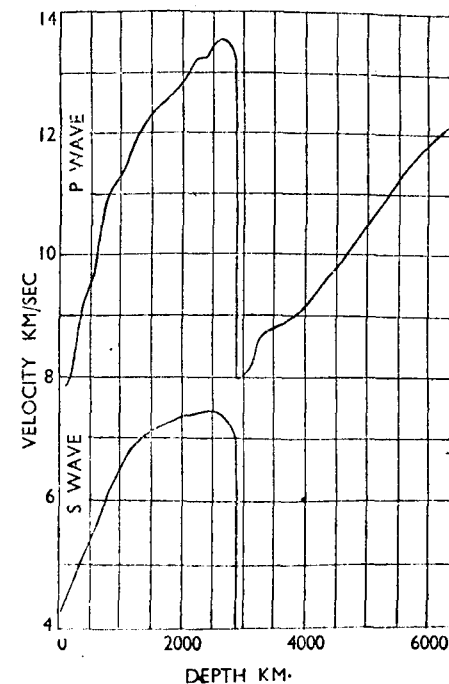


Fig. 12.—The speed of travel of both S and P waves increases as we go deeper into the Earth, until the core is reached.

of the weight and the surroundings, a few thousandths of an inch, is magnified and then recorded on paper fixed to a revolving drum. Time marks are also put on the paper.

There are various ways of arranging the work in seismic prospecting, but here we will consider only two ways — refraction shooting and reflection shooting. Suppose in the first instance that the area has been geologically surveyed and that a salt dome is to be expected. We first wish to find the horizontal extent of the salt which is surrounded by Tertiary strata. If we fire a charge of explosive at S (Fig. 13) and have seismographs at A, B and C, equal

distances from S, to receive the P waves, and if the ground were homogeneous (uniform throughout) the waves would all be received at the same time on the three seismographs.

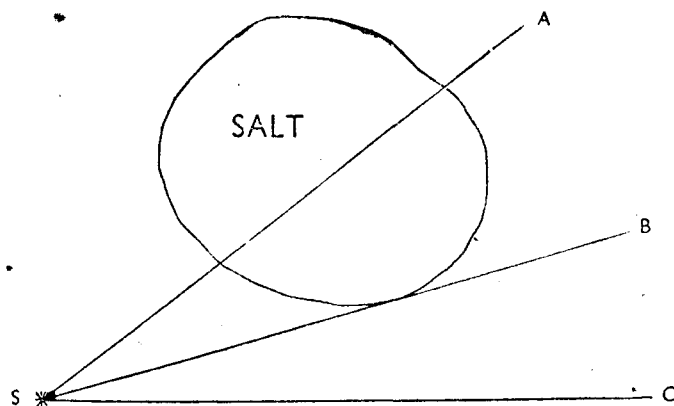


Fig. 13.—A plan view of seismic prospecting.

But as SA travels through the salt, which transmits P waves twice as fast as the surrounding rock, the waves will arrive at A before they arrive at C. By using a large number of seismographs it is relatively easy to distinguish the tangential ray SB and hence to find the limit of the dome in this direction. The prospector may then work round the deposit in a similar manner.

The next problem is to find the thickness of the overburden — the depth of the top of the salt. If an explosion at S (Fig. 14) causes an artificial earthquake to have S as its focus, P waves will travel outwards in all directions from S. Those vertically downwards will continue uninterrupted; those making a small angle with the normal to the surface will travel to the top of the salt and there be bent or refracted in a similar manner to that of a light wave going from air to water. A ray SP striking the top of the salt at the critical angle A in fig. 14.

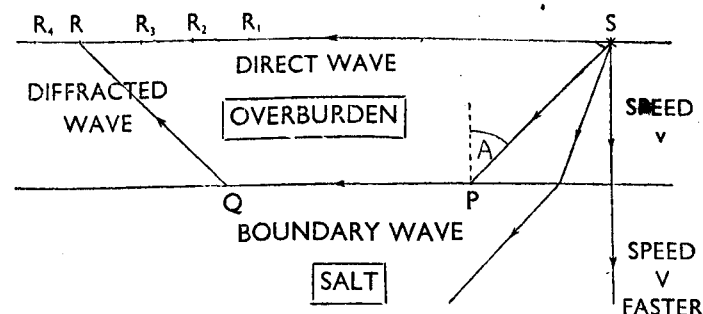


Fig. 14.—A section through the Earth's surface.

$$\text{(such that } \sin A = \frac{v}{V} \text{)}$$

will not go deep into the salt but will travel along the boundary with a speed equal to that in the salt (V). v is the speed of the P wave in the overburden. From every point along the boundary diffracted waves will come up through the overburden to the surface and these will be received by seismographs placed at R₁, R₂, R₃, R and R₄. The direct wave, with its whole course through the overburden, will also be received by these seismographs. On each seismogram the times of arrival of the direct wave and diffracted wave are noted and the point R found where the two waves arrive simultaneously. If the distance SR=2x and if the path SPQR is such that for SP,

$$\sin A = \frac{v}{V}$$

then it may be proved that the depth of overburden y is given by

$$y = x \sqrt{\frac{V-v}{V+v}}$$

Now v and V can be found from the slopes of the distance-time graphs for the direct and diffracted waves.

It would be impossible to use the above method if the speed of the P wave in the lower layer of rock were slower than the speed of the P wave in the upper layer.

Reflection shooting is more difficult. It depends upon PQ (Fig. 14) being a good reflecting horizon, and the shot in the borehole being sufficiently powerful for echoes to be transmitted through the overburden so that they can be recognised as such at R₁, R₂, R₃, etc. This must be so in spite of the much larger direct waves arriving at the seismographs, and moreover the timing of the echo must be carried out with an accuracy of a thousandth of a second. Only really good modern apparatus will accomplish all this. If the earth is multi-layered at that point, the prospector must find which horizon is giving the echoes but otherwise the field problem is merely one of finding the speed of the P waves in the upper layer and applying this to the timing of the echo. For this work a seismometer with short natural period and low sensitivity is fitted with an electromechanical transducer, usually of the moving coil type. The P wave is thus mechanically received and the movement converted into electrical impulses. These are then magnified several million times by means of an electronic amplifier which is also a band filter, letting through only frequencies of 30 to 60 per second. The amplifier is also controlled so that it only starts its work after the direct waves have passed and when the echoes are due to arrive. The electrical impulses are then passed through a small d'Arsonval type galvanometer and recorded photographically on a strip of paper moving 2 or 3 feet per second. Time marks are placed on the same strip by means of a mirror reflecting a spot of light from an oscillograph oscillating at 100 cycles. Clear marks are thus placed on the strip each five-thousandth of a second. A seismic survey verified and measured the shale-capped carboniferous anticline at Eakring in Nottinghamshire, England, where oil was later found by drilling.

Thus far we have considered the exploration of the outer

layer of the earth, which is the one handled by geologists. This portion of the earth is called by geophysicists the sedimentary layer. It is largely sedimentary in the geological sense as it is composed chiefly of such rocks as limestone, sandstone, shale and clay, which are the products of weathering and reconstitution of the original stony surface material. We will now probe deeper.

The interior

The basement complex of pre-Cambrian rocks, about 1,500,000,000 years old, is visible in Finland and consists there of 52.5 per cent by weight of granite, 21.8 per cent mixtures of granite, sediments (the earliest deposited on Earth) and old Volcanic rocks, 4.0 per cent granulites, 9.1 per cent schists (a metamorphic rock whose crystals are distinguishable to the naked eye, which is conspicuously thinly foliated and in which felspar is absent), 4.3 per cent sandstones and quartzites, 0.1 per cent limestones and dolomites and 8.2 per cent basic rocks unlike granite but once molten. It will be seen that it is truly granitic in composition, originally largely igneous though partly sedimentary, but now largely metamorphic in texture. There may be local variations in this metamorphosed mass of pre-Cambrian rocks, but according to seismological evidence it is fairly uniform in character and about 10 kilometres (6½ miles) thick. It is called the granitic layer and envelops the whole earth with the exception of the Pacific Ocean.

Under the granitic layer is another uniform layer some 20 kilometres (nearly 12½ miles) thick which has a composition intermediate between granite and basalt and which, in consequence, is called the intermediate layer. This may also be absent from the bed of the Pacific Ocean, but it is present everywhere else in the Earth.

The sedimentary, granitic and intermediate layers together form the Crust of the earth which took its name at the time when it really was considered to be the only solid

part of the earth. Natural earthquakes and very large explosions such as the atom bomb tests in New Mexico on 16 July, 1945, and Bikini on 24 July, 1946, the Burton-on-Trent explosion of 27 November, 1944, the Port Chicago explosion of 17 July, 1944, and the Heligoland explosion of 18 April, 1947, all sent P and S waves right through the crust and thus give us data concerning its constitution, structure and size.

Below the crust is the mantle, which continues to the edge of the core. The top of the mantle is probably the floor of the Pacific Ocean and it appears likely that large local reservoirs in the mantle, caused by a temporary relief of pressure, were the sources of the great basaltic lava flows of the Deccan of India and the Giant's Causeway in Northern Ireland. The mantle is thought to consist wholly of basalt (dark coloured, dense, basic, alkali rock poor in silica) crystalline for the top 70 kilometres (43½ miles) and glassy below. Towards the base of the mantle metallic ores are probably present in greater proportions. Elsewhere various small changes in composition or texture have been suggested from time to time, but these changes are not sufficiently obvious to have gained universal acceptance. The mantle is solid throughout and transmits S waves. It has a rigidity of 16×10^{11} dynes per square centimetre (10,352 tons wt. per sq. inch) which is about twice as rigid as steel.

The base of the mantle is at a depth of 2,780 kilometres (1,728 miles) from whence there is a gradual change to 3,000 kilometres (1,864 miles) where the core begins. The core ranges in density from 10.1 gm. per c.c. (630.5 lb. per cu. ft.) at the outside to 12.2 gm. per c.c. (761.6 lb. per cu. ft.) at the centre. It is thought to be liquid and may be composed of iron and nickel. It focuses earthquake waves which pass through it (as a magnifying glass focuses light) to a point in the antipodes and around this focal area is a shadow zone stretching roughly from 102° to 143° from the epicentre (point on the earth's surface directly above the

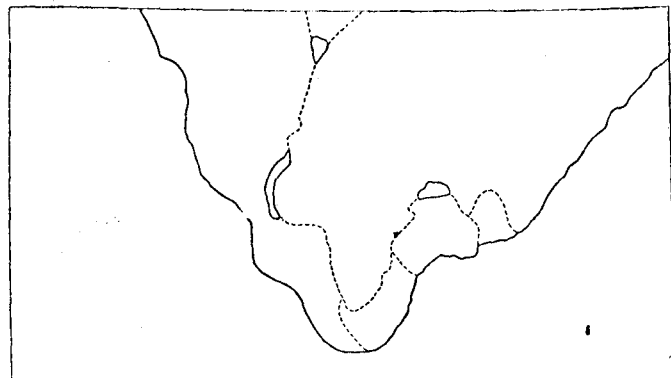
starting point of an earthquake). Furthermore, echoes of waves from natural earthquakes reflected from the core boundary can be distinguished on seismograms and these, as in reflection shooting (seismic prospecting) enable us to find the depth of the boundary. The high pressures existing and the physical properties of the material at these high pressures keep the core stable, though at these temperatures and pressures it may constitute a transition region where physical and chemical properties begin to change strongly. (These conditions are nothing like the conditions inside stars, however.)

Temperatures within the earth.

The temperature gradient in the crust of the earth is such that it gets about 1°C warmer in every 30 metres (1°F for every 18 yards deeper approximately) though there are large variations such as over salt domes and granite plugs, and over basement rocks. The heat current through the earth's surface is about 10^{-6} calories per second per square centimetre (3.69×10^{-6} British Thermal Units per second per square foot) and this appears to be largely accounted for by the estimated spontaneous production of heat by the known average radioactive mineral content of rocks in the uppermost 40 or 50 kilometres (25 to 31½ miles) of the earth. Furthermore, laboratory experiments tend to confirm that at high pressures rocks and minerals and metals are superconducting to heat and we are thus led to the conclusion that the interior of the earth no longer contributes to the heat flow through the surface. Except for the outermost part of the mantle the temperature in the interior of the earth appears to be not far from the temperatures existing at the time when the crust solidified.

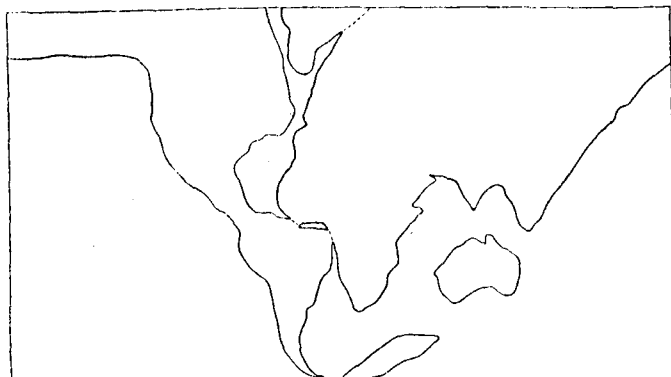
When the earth was originally gaseous it is thought to have had a temperature of $6,000^\circ$ absolute ($5,727^\circ\text{C} = 10,340^\circ\text{F}$); the crust became solid after about 15,000 years and soon afterwards the surface cooled sufficiently for oceans to condense. The temperature of the earth's centre

may now be just over $2,000^{\circ}\text{C}$ ($3,632^{\circ}\text{F}$), the temperature at about 80 kilometres ($49\frac{1}{2}$ miles) being about $1,200^{\circ}\text{C}$ ($2,192^{\circ}\text{F}$).



(a)

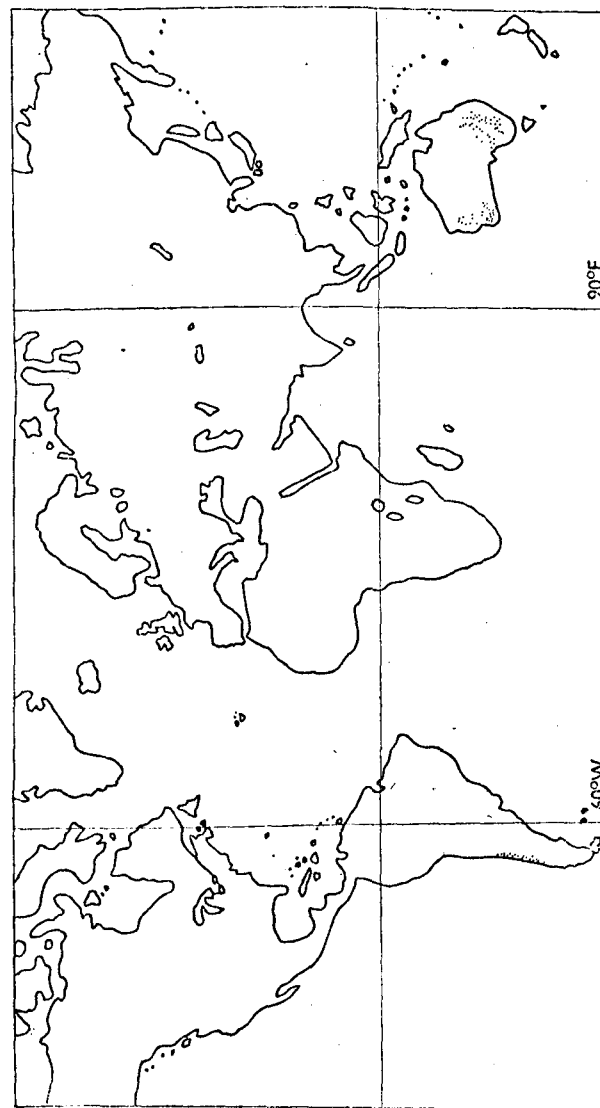
STAGE 1.



(b)

STAGE 2.

Fig. 15.—The Wegener theory of Continental drift supposes that at one time all the continents of the world were bound in one great land mass (see (a)). Under the influence of the Earth's rotation they began to tear apart, attaining an intermediate stage (b).



STAGE 3. Fig. 16.—The world to-day: the continents are separated. But signs of the old state of affairs remain in the structure of the continents, and in the distribution of certain animals and plants.

Forces in the earth's crust.

A consideration of the conditions under which the deposition of certain sedimentary rocks took place together with an examination of their fossil content, shows us that areas which are now dry land were at one time beneath the sea. It is one of the fundamental aims of geology to reconstitute past geographies and to determine, from the time when the oceans first formed on the earth, the changing positions of the shore lines. During the course of this time, points on the earth's surface have changed their relative vertical positions through several thousands of feet, and rocks which were deposited horizontally on the ocean floor have been raised to become dry land, and often compressed laterally until they have in some cases been left standing vertically, and in other cases as in the Alps and the Highlands of Scotland been turned right over. Ours is truly a mobile earth. The pressure appears to have come from different directions at different times though the whole process has been extremely gradual. For example, pressure from the North-east and/or South-west which folded the rocks of the Charnwood Forest area into axes which run North-west to South-east took place in pre-Cambrian times about one thousand million years ago. The NE - SW Caledonian axes typical of parts of Scotland were caused mostly in post-Silurian times about 400,000,000 years ago. Later on the direction changed again, for in post-Carboniferous and post-Triassic times (from 150 to 50 million years ago approximately) we find the fold axes were made to lie roughly north and south as in the Malvern Hills.

Sometimes, when the lateral pressure which causes mountains to be formed is too great for the strength of the rocks, the rocks crack or 'fault' and at the same time an earthquake takes place from the point where the first crack appears. Relative movement of the rocks on both sides of the fault also takes place. Earthquakes start or have their 'foci' at any depth in the earth down to 700 kilometres (435 miles), and earthquakes at this depth appear to have

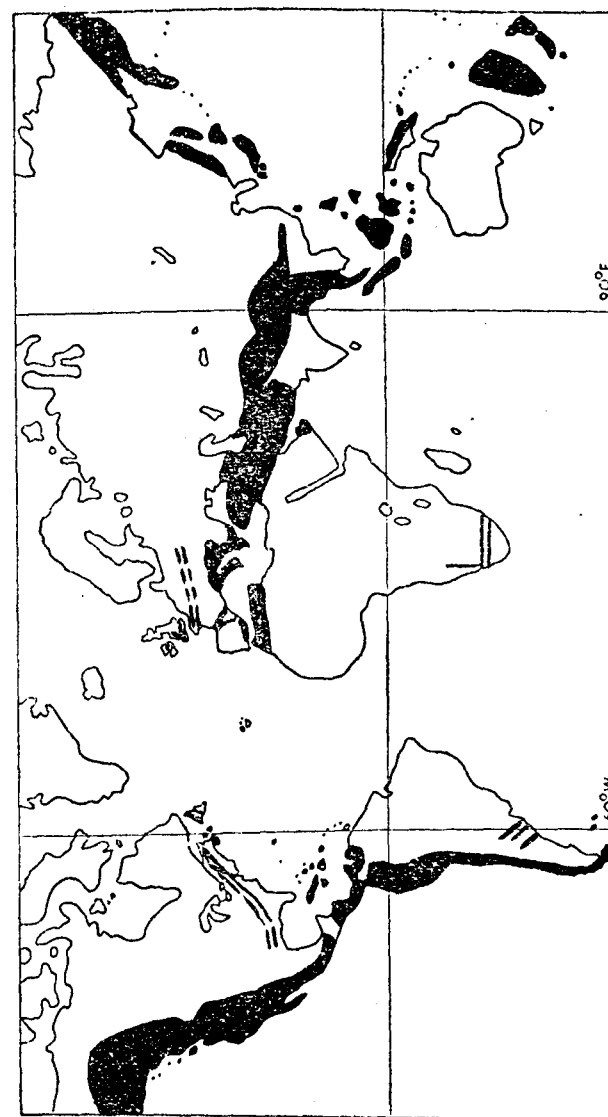


Fig. 17.—A world map of recent mountains.

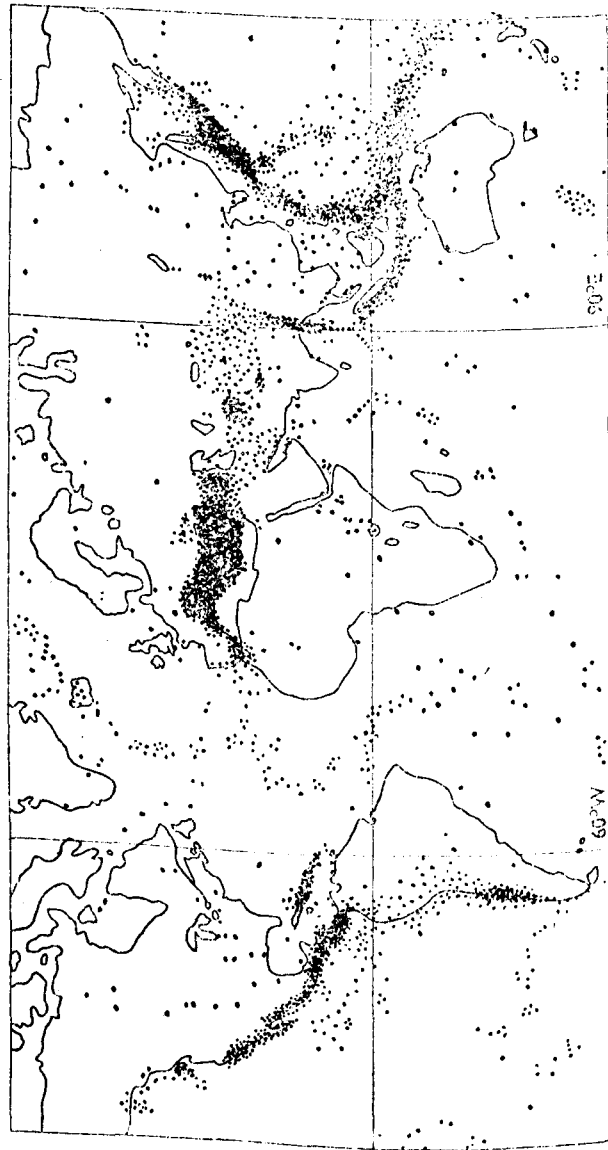


Fig. 18.—Map showing earthquake epicentres.

been caused in the same way as earthquakes which start at the earth's surface. This point is, however, still under discussion. If earthquake epicentres are plotted on a map according to their depth of origin interesting results are obtained. We find that earthquakes with shallow foci, i.e. 0 – 100 kilometres (0 to 62½ miles) and intermediate foci, i.e., 100 to 290 kilometres (62½ to 181 miles) are closely associated with Tertiary or more recent mountain building (within the last 60 million years), whilst deep focus shocks, i.e. 290 to 700 kilometres (181 to 435 miles) are associated with pre-Cambrian topography (1,000,000,000 years old). What was the origin of the lateral forces in the earth's crust which from time to time have been responsible for all this movement and mountain building? Some have been due to the contraction of the earth's interior consequent on cooling, and some may have been caused by variations in temperature within the crust and upper mantle owing to radioactivity, but the full story is not yet known. Mountain building appears to have gone on at a much more rapid rate during some periods of the earth's history than at others. Let us take a look at the 'foundations' of the crust, i.e. the upper part of the mantle.

The crust of the earth is often called SIAL on account of its estimated average composition by weight, which is SiO_2 , 59.12 per cent, Al_2O_3 , 15.34 per cent, other minerals making up the remainder. The material beneath this is often called the SIMA because of a preponderance in it of minerals containing silica and magnesium. Now the upper part of the SIMA, called the astheno-sphere, yields to the crust above it. During the last (Pleistocene) ice age Scandinavia was depressed about 700 metres (765½ yards) by a load 3 kilometres (1½ miles) thick of ice, and is even to-day gradually recovering from the load. The theory of isostasy concerns this hydrostatic support of the earth's crust by the mantle and states (in one of its forms) that each vertical column of the earth's crust, in regions that have not been recently disturbed, which have a radius of at least 10 kilometres (6½

form of carbon, are found. The aerolites are largely composed of igneous rock whose average chemical composition is very like that of terrestrial peridotitic rocks (a rock in which crystals of olivine may be readily distinguished and in which there is very little else. They are extremely poor in silica and are definitely ultra-basic). Meteorites appear to have come originally from some mass of cosmic matter which had a spherical form, increased in density towards the centre, and had cooled from a molten mass free from water before it disrupted. It is suggested that this cosmic mass was similar in constitution to the earth, with possibly the outer crust and atmosphere missing. Meteorites support the hypothesis that the mantle of the earth is composed of ultra basic rocks and that the earth's core is made of iron with the addition of some nickel.

Finally, the elastic constants of the earth's layers obtained from the rates of transmission of earthquake waves, are found to be the same as those obtained from a consideration of the tides set up in the earth by the gravitational forces of the moon and sun.

Thus we find that although the 1948 model of the earth is better than previous models, there are still large parts of our model to improve before we have a model which will work exactly like the earth itself. Observation is the key to the problems involved.

Glass

K. L. LOEWENSTEIN

THE art of glass manufacture depends on melting together a number of substances, which, on cooling rapidly, will set to give a clear, transparent solid. From a physical point of view this corresponds more to the liquid than to the solid state. For example, the change from liquid water to solid water (ice) takes place at a definite temperature, and as long as both water and ice are present the temperature of the mixture will be 0°C or 32°F ; with glasses this is not so on cooling, the viscosity increases continuously, and there is no definite temperature at which the glass can be said to solidify; similarly, on heating glass from room temperature it will soften more and more without any break in the temperature curve. This is best illustrated by means of a graph,

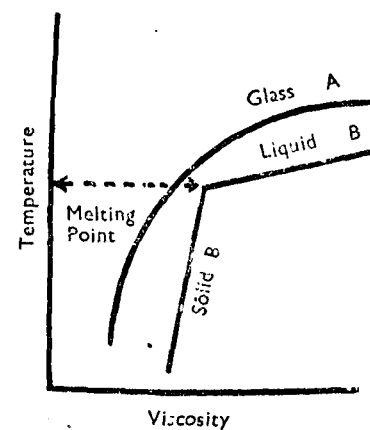


Fig. 2P.

plotting viscosity against temperature. Glasses will give a curve of the continuous type (A), whereas liquids and solids give the discontinuous type (B) (Fig. 21).

For this reason the glassy state has been termed 'vitreous,' and glasses are often described as 'super-cooled liquids.' The decomposition of glass into solid constituents can take place, especially when the glass is cooled very slowly; this is called 'devitrification' This often takes place years after a glass article has been made, and there is no known method by which it can be arrested.

The properties of glass must be in accord with the functions that it has to fulfil. For example, the wires of electric light bulb filaments have to be sealed into glass. To get a good seal, not only must the coefficients of expansion of the glass and the metal be as nearly identical as possible, but also the rate of change of these coefficients must be the same at any temperature. Any serious deviation from these conditions is liable to produce fracture.

The constituents of glasses:

(1) Glass-formers. These are pure substances which can exist as glasses by themselves, and without which no stable glass can be made. For example, silica (sand) can be made into a glass, but lime, present in most glasses, or lead oxide, present in many, cannot form a glass.

By far the most important glass-former is silica; the others are germanium dioxide, beryllium fluoride, and phosphorus pentoxide; the last, in particular, has now been extensively investigated for special electrical and optical purposes.

What makes a substance a glass-former is a question of relative atomic size of the atoms making up the substance. Firstly, the cation (silicon, phosphorus, etc.), must be large enough to space around itself a number of anions (oxygen, fluorine, etc.), to permit the formation of a three-dimensional network structure; and secondly, the anion must be small enough in relation to the cation to fit into such a

structure. For example, beryllium fluoride is a glass-former; on substituting a smaller atom, e.g. lithium, for beryllium, or a larger atom, e.g. bromine, for fluorine, we find that the substitutes are too small or too big respectively to fit into the network structure.

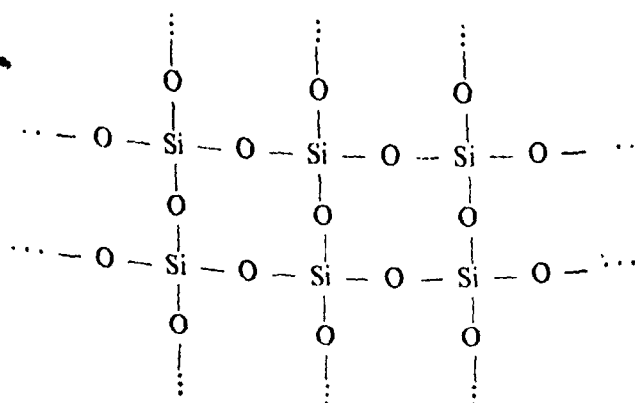
It has been found that the ratio of the atomic radii of the cation to the anion must be approximately 0.3.

The high stability of silica glasses compared with other glass-formers is ascribed to the fact that the ratio of the number of cations to anions is 1:2, so that each silicon atom can group around itself four oxygen atoms in the shape of a regular tetrahedron, and that each oxygen can hold two silicon atoms; furthermore, as no oxygen can hold more than two silicon atoms there is a certain flexibility in the joining together of SiO_4 tetrahedra. Hence the energy in the supercooled state is not very different from that in the crystalline state, and the tendency to devitrify is small.

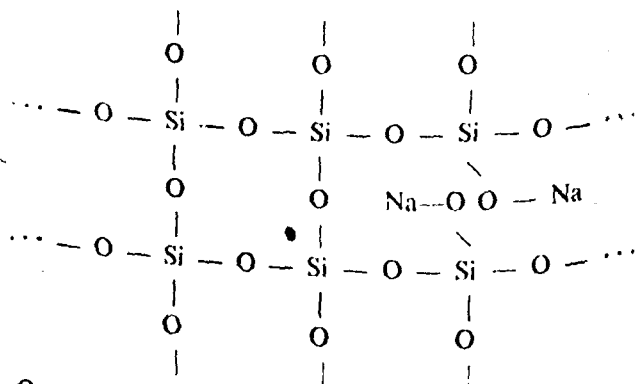
If, however, other groups besides the SiO_4 tetrahedra are incorporated in the glass structure, the tendency to devitrify is increased.

(2) Other constituents:—The breaking-up of the three-dimensional SiO_4 network by other constituents will lower the melting point of the mixture, and good glasses can be obtained at temperatures as low as $1,250^\circ\text{C}$, compared with $1,750^\circ\text{C}$ for the melting point of silica.

The network can be broken up by any substance that is soluble in the glassmelt, and which has no tendency to separate out on cooling. The oxides of the alkalis (lithium, sodium, potassium, rubidium, and caesium) are the most powerful substances to break up the silica network, as each molecule of alkali oxide breaks up one — Si — O — Si — link, thus:—



Silica in the vitreous state.



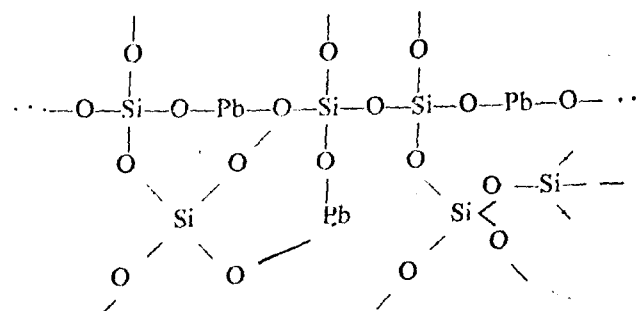
One molecule of sodium oxide, Na_2O , has entered the glass structure. (The diagrams are projections on to a plane, and do not represent relative distances, etc.)

It can be seen that there must be a limit to this process of breaking up the silica network; this limit is the formation of sodium silicate (waterglass). In glassmaking the limit is governed by the stability of the resultant glass, as a rough guide not more than 18 per cent of alkali oxide can be

incorporated into a glass without asking for trouble: the glass tends to devitrify even in the molten condition, and the finished article will be attacked rapidly by the carbon dioxide in the atmosphere.

The continuity of the silica network can also be broken up by replacing some of the oxygens by fluorine. Oxygen is divalent and can, therefore, act as link between the individual tetrahedra; fluorine is monovalent and therefore cannot. This method has its drawbacks as fluorides are not very soluble in the glassmelt, and the pots in which the glass is melted are badly corroded.

Calcium oxide (lime), barium, lead, zinc, and magnesium oxides do not break up the silica network, but thin out the tetrahedra, e.g.



The extent to which these divalent metals can be added in the form of their oxides depends on the properties of the resultant glass, the solubility of the oxide in the glass, etc. With lime, barium oxide, zinc and magnesium oxides, the glass becomes unstable when about 25 per cent is used; with lead oxide, however, glasses containing 90 per cent have been made and are quite stable. Lead crystal glasses contain 20 to 30 per cent of lead oxide.

The melting properties of most glasses can be improved by adding boric and/or aluminium oxides. These, when added in small quantities (up to 5 per cent), can form tetra-

hedra, which, being of a different size to those of silica, will distort the symmetry of the SiO_4 groups in relation to one another. This gives greater flexibility to the glass structure without breaking up the network.

The sum total of all these effects on the originally symmetrical network is that it is impossible to represent visually and accurately the structure of a glass as was attempted above. All that can be said on the subject is that the average position of the individual constituents is, say, one alumina group for every 50 silica groups, on the average every 24th $\text{Si} - \text{O} - \text{Si}$ link is broken by an alkali atom, or, for every three SiO_4 groups there is one lead atom thinning out the structure. But we must remember that these are average positions, and deviations may be considerable. It is impossible to picture the structure of glass; the lines and chains of atoms are twisted and turned in all directions; the space allotted to each atom varies, some are squeezed in, some have more 'breathing space,' but, on the average, they just about fit in. In technical language this has been termed a random network structure, common also to mixtures of liquid substances; this is another reason why the physicist calls glass a supercooled liquid.

Viscosity

DR A. E. BELL

VISCOSITY plays perhaps as large a part in the phenomena of everyday life as does friction – with which it is related. It is in fact an effect we commonly exchange for friction when the latter is undesirable. When we wish to reduce the drag between two metal surfaces, as in the moving parts of machinery, we introduce a thin layer of oil which acts as a lubricant. This replaces frictional loss of energy by a loss due to viscosity, but with a suitable lubricant the loss is much less, and the wear is eliminated. The viscosity, or resistance to flow offered by the lubricant in the bearing, remains an unavoidable nuisance, but fortunately the viscosity of a liquid in practically all cases shows a marked decrease as the temperature rises, and this means that considerably less energy is lost once an engine is warmed up. As Dr Bowden pointed out in his article on Friction the existence of such a drag between bodies is in some respects useful, in others a nuisance. It is the same here with viscosity. A reasonable amount of viscosity in a paint is an advantage since there is little flow during the time taken for the paint to harden. Too much will mean that the brush marks are retained too long and surface tension will not be able to flatten the surface before it is hard. Fortunately there are certain liquids which exhibit a reduced viscosity when they are stirred rapidly, the effect being called thixotropy. A small amount of thixotropy may be an advantage in a paint or varnish – while it accounts for some of the attractive qualities of whipped cream!

Much attention has naturally been given to the composition of lubricating oils and the properties which are desirable in them. They are all compounds of hydrogen and carbon with many carbon atoms arranged in fairly long chains.

These molecules slide over each other and one layer seems to cause a drag on the next through the entanglement of the chains. This sliding of one layer over another, like the sliding over each other of the cards in a pack, is called a shear. In Newton's great *Principia* of 1687 it was pointed out that the shearing force F is connected with the rate of shear, dv/dx , the viscosity η and the area of the layers A by the equation

$$F = \eta A \frac{dv}{dx}$$

and this is still the relation employed in measuring viscosity. Since liquids commonly adhere very strongly to the surfaces of solids the flow of a liquid through a uniform circular tube involves the sliding of concentric layers over each other with the fastest movement at the middle and a zero velocity next to the wall. This is what is called laminar or 'Newtonian' flow, and it means that the contour of the velocities across a section of the tube is of the form shown on the left.

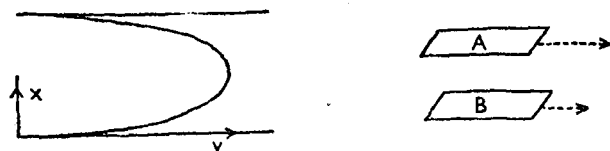


Fig. 22.

If we think of two flat layers A and B one centimetre apart, one moving relatively faster than the other by a velocity of one centimetre per second, then clearly A will exert a forward drag on B and B will exert a backward drag on A. The force exerted per square centimetre of either layer is the viscosity of the liquid.

In a general way it is clear that viscosity results from forces existing between the molecules as they pass over each other, and it is natural that physicists should be anxious to connect the viscosity of a liquid with the known properties of its molecules. Measurements of viscosity are relatively simple and involve study of the rate of flow of liquids through fine tubes or the rate of fall of a small metal sphere

through a column of liquid. What is wanted is a science of the subject, and straight away we see that to understand viscosity we need to know a good deal about the internal conditions, one might say the constitution, of liquids. These, we know, are composed of molecules in fairly close contact and the forces between the molecules are unquestionably of electrical character. Nevertheless we can get on quite well without concerning ourselves very much about the origin of the forces and simply occupy ourselves with 'billiard ball' molecules moving about in quite a Newtonian manner. I mean that we need to think only of the momentum of moving particles and the effects of forces on them, and not concern ourselves at all with electrons and protons, still less with those wave aspects of matter with which modern physics makes us familiar.

At first sight it certainly seems as if liquids are really only very highly condensed gases or vapours. You have only to cool and compress air sufficiently and, as we all know, it begins to liquefy. Thomas Andrews² studied the change in the case of carbon dioxide in the sixties of last century and he produced evidence to show that there is in fact a 'continuity of state' between the liquid and gaseous conditions of that compound. His method was to study the change of volume with pressure at any given temperature, and he did in fact plot the curve for the pressure/volume relation for a series of temperatures. The shapes of these curves or isothermals are shown in the figure. Andrews found that for any temperature below 31°C there is a particular pressure at which the gas begins to liquefy: this might be called the saturated vapour pressure for this temperature. So long as both gas and liquid were present in his apparatus the pressure remained constant. So soon as the gas was all liquefied it was found that great increases in pressure produced no appreciable decrease in volume, on account of the incompressibility of the liquid. The result is that the left-hand portions of the curves are all vertical.

The continuity of the two states can best be appreciated

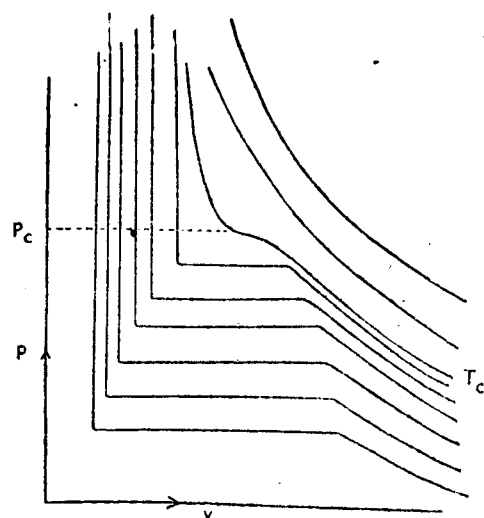


Fig. 23.—The behaviour of carbon dioxide at different temperatures (see text below for explanation); P, pressure; V, volume.

by running the eye from the bottom left-hand part of the figure up towards the top right-hand part. The curve marked T_c is the isothermal for the critical temperature, and at the pressure p_c the conditions are in fact exactly intermediate. On watching a tube half-full of liquid carbon dioxide, as this temperature is reached it is seen that the liquid surface vanishes without the liquid having boiled away. This is because the liquid and gas states have under these conditions the same density and they are as a consequence indistinguishable. If we represent the liquid at a temperature lower than T_c by the densely crowded condition of the molecules in (a) and the gas at that temperature by (b), then at T_c and p_c the two pictures have to be made identical through the expansion of the liquid on the one hand and the compression of the gas on the other.

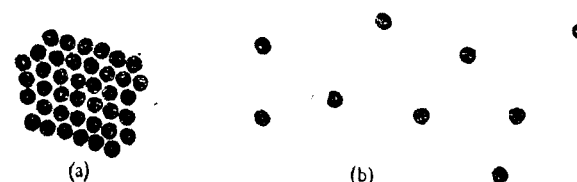


Fig. 24

In his first scientific publication the Dutch scientist Van der Waals produced (1873) an elegant explanation of the facts discovered by Andrews and others. This successfully treated the liquid state as originating from the gaseous state, and the combined work of Andrews and Van der Waals was for many years extremely influential. Jeans remarked, however, that 'there was a certain element of luck' in Van der Waal's equation having predicted the change from the gaseous to the liquid state. Today the old kinetic theory, according to which liquids closely resemble gases (in that their molecules are distributed in a purely random manner) has undergone considerable changes. The nineteenth century chemists, it must be admitted, made remarkably good progress with a theory of solution in which the molecules of the liquid were treated as if they did not exist! The dissolved substance was supposed to have its molecules distributed just as if they were the molecules of a gas and for many purposes the solvent liquid could be forgotten. There was a certain paradox in saying that this success provided support for the kinetic theory, but so it was generally supposed to be. The phenomenon of Brownian movement was probably the best visual demonstration of the general correctness of the kinetic theory. In this phenomenon, microscopic particles such as those found in Indian ink, or soil, or consisting of pollen grains or bacteria, when placed in water are to be seen under the microscope to be constantly vibrating or gyrating about fixed points so that the whole field of view is gently trembling in its appearance. Einstein gave a detailed

mathematical theory of the phenomenon, which is due to the unequal bombardments of relatively huge floating particles by the surrounding water molecules. All that Brownian movement shows therefore is the existence of molecules in a state of vibration.

It has not proved to be possible to give a mathematical account of viscosity on the basis of the old classical theory outlined in the preceding two paragraphs, and moreover there are other matters such as the existence of a more or less sharp melting point which are equally (or more) difficult. You have only to remove enough of the energy of the molecules of a liquid and they begin to take up a pattern and 'crystallise'. When the crystals are melted how do we know that the pattern completely disappears? The answer appears to be that it quite often does not. Professor G. W. Stewart³ was the first to draw attention to the fact that where it was possible to obtain X-ray diffraction patterns for both the liquid and the crystalline states of a substance there was a general resemblance between them. In the case of a crystal the pattern is formed by the scattering of a pencil of X-rays by the regularly spaced atoms, and it would be thought that the molecules (and thus atoms) of a liquid would be distributed in so random a manner that no pattern would be produced. This proves not to be so, although the pattern is far less easy to detect than for the crystal.

Following up this lead Professors Bernal and Fowler in 1933 published a long paper⁴ on the constitution of water. They considered a wide variety of evidence including the X-ray scattering effects and they were able to conclude quite definitely in favour of a particular kind of 'quasi-crystalline' structure for water near its freezing point. Their views fit in well with Stewart's theory that liquids have a 'cybotactic' structure, viz., that there are small aggregates of molecules arranged as they would be in a crystalline particle. Thus in water the same pattern which is found repeated in all directions in the ice crystal is present in water even though it extends over only a limited number of molecules at a time.

The pattern is constantly forming and vanishing at any given point so that on the average only a certain degree of order prevails.

A new theoretical approach has therefore been made to the subject and it is supposed that a molecule is 'all the time within a field of force of about as many other molecules as in the solid phase; it does not, as in a gas, spend most of its life in comparative freedom and interact strongly with other molecules only occasionally'.⁵ Although the picture is not yet very clear the results from the experimental and theoretical sides seem to be in general agreement. Useful evidence can for example be obtained from the study of glasses, which are simply liquids cooled down below their freezing points but without crystallisation having taken place. It is as if the particles of the liquid had been frozen in their positions so that they can be studied at leisure. (No doubt it is not really quite so simple as that.)

It must be emphasised that an article of this size and scope cannot do more than present a general picture. But enough has already happened to make it quite certain that the old kinetic theory of liquids is quite inadequate. For the specialist there is a large review of the subject by the Russian physicist Frenkel.⁶ Frenkel shows repeatedly that we must give up the belief that the physical and mechanical properties of liquids are radically different from those of solids, and indeed the present position is that liquids have as much or more in common with crystalline solids as they have with gases.

Quite inevitably the new view of the liquid state has meant a considerable change in the explanation of viscosity. In a flowing liquid it is now supposed that some of the momentum of one layer is transferred to the next (i.e. in laminar flow) through the operation of transient molecular forces of the sort found in crystals. On the older theory, molecules were supposed to be lost from one layer to the next, and it was this which was supposed to bring about the transfer of momentum. It is clear that the loss of faster-moving mole-

mathematical theory of the phenomenon, which is due to the unequal bombardments of relatively huge floating particles by the surrounding water molecules. All that Brownian movement shows therefore is the existence of molecules in a state of vibration.

It has not proved to be possible to give a mathematical account of viscosity on the basis of the old classical theory outlined in the preceding two paragraphs, and moreover there are other matters such as the existence of a more or less sharp melting point which are equally (or more) difficult. You have only to remove enough of the energy of the molecules of a liquid and they begin to take up a pattern and 'crystallise'. When the crystals are melted how do we know that the pattern completely disappears? The answer appears to be that it quite often does not. Professor G. W. Stewart³ was the first to draw attention to the fact that where it was possible to obtain X-ray diffraction patterns for both the liquid and the crystalline states of a substance there was a general resemblance between them. In the case of a crystal the pattern is formed by the scattering of a pencil of X-rays by the regularly spaced atoms, and it would be thought that the molecules (and thus atoms) of a liquid would be distributed in so random a manner that no pattern would be produced. This proves not to be so, although the pattern is far less easy to detect than for the crystal.

Following up this lead Professors Bernal and Fowler in 1933 published a long paper⁴ on the constitution of water. They considered a wide variety of evidence including the X-ray scattering effects and they were able to conclude quite definitely in favour of a particular kind of 'quasi-crystalline' structure for water near its freezing point. Their views fit in well with Stewart's theory that liquids have a 'cybotactic' structure, viz., that there are small aggregates of molecules arranged as they would be in a crystalline particle. Thus in water the same pattern which is found repeated in all directions in the ice crystal *is* present in water even though it extends over only a limited number of molecules at a time.

The pattern is constantly forming and vanishing at any given point so that on the average only a certain degree of order prevails.

A new theoretical approach has therefore been made to the subject and it is supposed that a molecule is 'all the time within a field of force of about as many other molecules as in the solid phase; it does not, as in a gas, spend most of its life in comparative freedom and interact strongly with other molecules only occasionally'.⁵ Although the picture is not yet very clear the results from the experimental and theoretical sides seem to be in general agreement. Useful evidence can for example be obtained from the study of glasses, which are simply liquids cooled down below their freezing points but without crystallisation having taken place. It is as if the particles of the liquid had been frozen in their positions so that they can be studied at leisure. (No doubt it is not really quite so simple as that.)

It must be emphasised that an article of this size and scope cannot do more than present a general picture. But enough has already happened to make it quite certain that the old kinetic theory of liquids is quite inadequate. For the specialist there is a large review of the subject by the Russian physicist Frenkel.⁶ Frenkel shows repeatedly that we must give up the belief that the physical and mechanical properties of liquids are radically different from those of solids, and indeed the present position is that liquids have as much or more in common with crystalline solids as they have with gases.

Quite inevitably the new view of the liquid state has meant a considerable change in the explanation of viscosity. In a flowing liquid it is now supposed that some of the momentum of one layer is transferred to the next (i.e. in laminar flow) through the operation of transient molecular forces of the sort found in crystals. On the older theory, molecules were supposed to be lost from one layer to the next, and it was this which was supposed to bring about the transfer of momentum. It is clear that the loss of faster-moving mole-

cules to a slower-moving layer would have this effect. Frenkel, and in this country Professor Andrade, have worked out theories of viscosity which are more in line with the new ideas. As Professor Andrade expressed it, on the supposition of a 'transitory and fluctuating crystallisation,' or a temporary holding of hands between individuals of two parallel rows moving past one another, a theoretical equation can be derived which holds good for the viscosities of quite a number of elements in the liquid state. This equation is

$$\eta = K \sqrt{\frac{A T_m}{V_A}}$$

where A is the atomic weight of the element, T_m the melting point on the absolute scale of temperature, V_A the volume of the atomic weight expressed in grams and η the viscosity. The constant K has the value 5.1×10^{-4} . A modified equation can be applied to those liquid *compounds*, the molecules of which have sufficiently uniform fields of force around them.

Unfortunately water remains too difficult a problem. The water molecule possesses a field of force which is by no means uniform, and as Bernal and Fowler have shown the amount of 'order' in the liquid is relatively high. It would seem that the existence of cybotactic units must interfere with laminar flow as conceived in the simplest way. The consequence is that there is as yet no complete physical theory for the viscosity of water. Much work has been done, however, on the effect of dissolved substances such as salts.

The initial effect of all salts in very weak solutions seems to be to raise the viscosity. It is known on other grounds that the ions into which salts are dissociated have considerable restraining influences on each other's movements, and viscosity measurement of salt solutions does provide in fact a useful means of examining inter-ionic forces. The ions of salts are surrounded by large numbers of water molecules

which orientate themselves in the electrical field of the ion. There are, besides, ions of opposite charge always in the vicinity. The result is an ionic atmosphere which is constantly shed and rebuilt as the ion moves through the liquid. The effect on the viscosity of the solution has been calculated, notably by Falkenhagen, and some agreement between theory and experiment can be recorded.

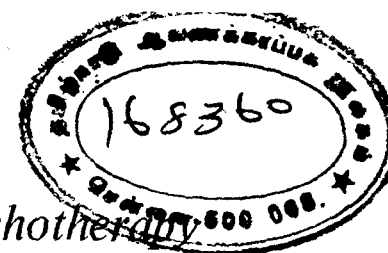
It is something of a relief to turn to the solutions of non-electrolytes in solvents other than water, for here the conditions are a good deal simpler. Here we almost certainly approach the state of affairs envisaged in Andrade's theory. The molecules of the dissolved substance affect the viscosity in a way which depends in a relatively simple manner on their size and shape. Staudinger, for example, uses the viscosity of a solution (e.g. in tetralin) in order to compare the lengths of molecules. Thus he has shown that the relative alteration in viscosity

$$\frac{\eta_{\text{solution}} - \eta_{\text{solvent}}}{\eta_{\text{solvent}}}$$

depends directly upon the concentration of the dissolved substance and the size of its molecules. Provided the substances compared are of the same chemical type the method gives a quick means of comparison of the sizes of molecules produced in the important process of polymerisation. In polymerisation the chemist is causing small molecules to join up, usually in a head to tail fashion, to produce very large molecules, and in the production of lubricating oils and in the formation of substances for use in plastics this is of great economic importance.

REFERENCES

1. *Science News*, vol. 4, p. 139.
2. Bakerian Lecture, Royal Society (*Proceedings*) 1869.
3. *Chemical Reviews* 6 (1929) 483.
4. *Reviews of Modern Physics* 2 (1930) 116.
5. *Journal of Chemical Physics* 1 (1933) 515.
6. N. F. Mott and R. W. Gurney, *Reports on Progress in Physics* (1938) 48.
7. *Kinetic Theory of Liquids* (Oxford University Press, 1946).



Group Psychotherapy

STEPHEN LESTRANGE

DURING the last generation the treatment of the mentally sick has made great advances. No longer does the onset of serious mental illness mean the incarceration and the custodial care of the patient pending a possible spontaneous recovery. Instead, there is an almost universal counsel of hope and active therapeutic measures are available. The dramatic empirical success of the new physical treatments – electrical treatment, insulin therapy and ‘Mind Surgery’ – must not be allowed to obscure the more solid advances which have been made in the understanding and treatment of the milder forms of mental illness. Mental illness – ‘Insanity’ – still has a great deal of stigma attached to it, but with the increase in knowledge of the function and structure of the normal mind, and the realisation that even the most polished intellect is flecked with emotional bias and miniature pathological reactions, people are more willing to go to their doctor with what they used euphemistically to call their moods, turns and eccentricities. Doctors in their turn realise more often that nervous symptoms can and should be treated.

Knowledge of the how and why of mental disorder first began to take scientific shape with the work of Freud at the end of the last century. Even now, fifty years later, so subjective is the nature of the material that argument still rages round many of the early findings, and the body of knowledge which is widely accepted or open to direct proof remains extremely small. Freud recognised that neurosis was largely, if not entirely, a price that was paid for civilisation and the many advantages gained from a complex social organisation. Most of his investigations led him to the study

Am 1, 146
24.8.5

of mental development in the very circumscribed life of the infant and its conflicts with a hostile or indifferent environment. The infant is intensely individualistic; he has no knowledge of, nor does he care about, the needs of others. At his first breath he is a pulsating but powerless bundle of life incapable of fending for himself or even of expressing his needs save by indiscriminate signs of distress. Fortunately his wishes are satisfied almost before they arise. Surely, then, here is a being completely dependent on others; he is unaware of this, however, and feels no obligations in return. It is the age of omnipotence.

The Freudians unfortunately interpret the consciousness of the newborn in terms applicable to our adult minds but which can bear little direct relationship to what the infant may be thinking – even if his mental processes bear any resemblance to those of adults. Whatever the conscious concomitants, however, it remains true that as the infant develops he has to give more and receive less: the world no longer revolves around his whims and there are other human beings whose needs and desires involve the postponement or loss of his own satisfactions. Meals must await the arrival of others and a kind word, if he is lucky, replaces a prolonged caress. Fears, beliefs, modes of reaction which are developed at this early stage lie deep in the formation of his character and are hard to change. They will, however, influence him throughout his life. These patterns of behaviour are forgotten – ‘repressed’ – and they are synonymous with the unconscious forces of the Freudians.

As the child grows up there are continually problems of social adjustment facing him at a conscious level: Play-mates – School – Parties – College – Work – Leaving Home – Marriage – these are all milestones in the development of his ability to handle interpersonal relationships; throughout life there is a continual adjustment to society. Living in a civilised community is not all electricity and drainage, and the price to be paid is not merely the rates. Civilisation involves the continual suppression of antisocial trends, the

inhibition and redirection (sublimation) of many primitive drives, especially those of aggression and sex. There must be a continual modification of the individual's wishes and behaviour in order that he may avoid conflict with his fellows: a kind of thermostatic control, with the regulator set at a conventional level of permitted self-seeking. This control mechanism may break down in several ways. The individual may refuse, for instance, to accept the sacrifices imposed by the community and become a criminal or a revolutionary. More often open conflict is unsought, and arises either because of a constitutional weakness of the personality or because the problem to be solved becomes too great. This failure to control the situation leads either to mental illness, or the stress may cause apparently physical disease such as peptic ulcer or colitis. In the causation of mental illness external strain generally interacts with a poor personality structure. The balance between these two factors is well illustrated by the experience of war-time. Here the external stress is increased and may become so great that the most stable personalities will react with marked neurotic symptoms.

Illness begets its own treatment. Psychotherapy is the combined attempt of the patient and his physician to unravel the causes of the conflict and attain an understanding of the underlying basis of the symptoms. Relief through psychotherapy can be obtained by three main methods.

Removing the cause: many of the aetiological factors lie in childhood or even further back – in the inherited make-up of the patient – and it is of course impossible to remove these. The burden of real and present troubles is, however, often a large one, for the neurotic tends to accumulate failure. Adjustment of the environment is therefore an important part of the treatment of the patient and it is the plaint of the social worker not that there is so little to do, but that it is so difficult. A patient who suffers from claustrophobia may find the long journey to work in a crowded train a terrible ordeal. His impulses to jump from the

compartment may become so strong that he no longer dares to make the journey and he stops off work. Financial worries now add to his misery. The treatment of claustrophobia may unfortunately be long drawn out and the social worker must step in to fill this gap. She is not content with helping the patient to obtain financial aid from the appropriate authority or charitable institution. Bread and butter may be essential for the patient's existence, but self-respect is just as fundamental to mental health and the social worker should find the patient similar work near his home. Alternatively, she can help him to obtain accommodation within walking distance of his original job.

Altering the patient: Many conflicts are due to reactions more relevant to the child than the adult; or they may be due to false beliefs which the patient has acquired early in life and which colour his picture of reality. To take a simple example: boys are often brought up to believe that women are different – angels in thought, noble and unsullied by the unpleasantness of life's struggle. This is known to psychiatrists as the error of the pedestal. Alternatively, the contrary delusion is equally common – that women are scheming, false and a certain ruin to the man who meddles with them. This is known, perhaps a little more cynically, as the error of reality. Both these errors are typical of the misconceptions a patient may have about reality. They are a type of fallacy inherent in human thought – generalisations from the particular and exaggeration of the importance of what we happen to have experienced ourselves. These fallacies do not easily suffer spontaneous correction, and for two main reasons. First, they are often beliefs stamped early into the mind of a child by a parent. To the child, the words of a parent or parent-surrogate have all the authority of the *ex cathedra* statement of a being who is at one and the same time the most feared and the best-loved person in the world, and whose omniscience has not yet been questioned. To deny these beliefs is to deny the parent himself, and however emancipated we may feel this arouses immense

unconscious resistance. Secondly, these fallacies are almost always half-truths, and there is consequently a lot of evidence to bolster them up: women are good and they are evil.

One further example of a type of misconception of major importance is that children are almost always taught absolute standards of right and wrong. Their criteria of ethical behaviour are often excellent, but the ideals are unfortunately painted in strong blacks and whites and much unnecessary conflict is engendered by this distortion. The man who believes that women are pure is unable to reconcile this with the reality of a human help-meet who, for his sake, cheats the milkman and jeopardises the rationing scheme. If, on the other hand, he regards women as the embodiment of evil, desire continually wars with fear. If we believe that virtue is absolute, it is disheartening to find ourselves always among the goats. Man has a need of virtue and a sense of right-doing, and if he is to keep going happily he must feel that some of his arrows are falling near their target, or at least travelling in the right direction.

This, succinctly, is our second point: the patient is struggling to fit comfortably into a niche in the external world. If his picture of reality is blurred and false he will find this more difficult – perhaps impossible. Here the psychiatrist can help him. Any intelligent onlooker can point out the foolishness of these misconceptions (they are not called delusions because so many normal people subscribe to them), but the neurotic is shy of rational argument and the only effective approach is by demonstrating to the patient how these ideas originally arose and how they gained the power they undoubtedly possess. Better still is the analytical method of leading the patient back over his life and letting him discover these early influences for himself.

I mentioned three main methods of treatment. We have already created a dichotomy – our patient and his environment. Our third approach therefore must embody the other two. We call it adjustment; fitting the patient into the real

world. Having made our assault on the environment (social therapy) and having altered our patient (individual therapy) we do not, unfortunately, find an easy fit. Bluntly, he must lump it, and we must help him. This is supportive therapy. We can help him to develop his assets, seize his opportunities and learn acceptable methods of gaining self-esteem and satisfaction. Like a patient with heart disease, he has to be taught to live within his limitations. Briefly: no moon and no tears. Critics of psycho-analytical technique say that what is not nonsense is common sense. Of course this is true and one is tempted to reply with the adage that common sense is an uncommon virtue. Sense, that is to say, rational thought, is a treasure that is available now and again to everyone, but never continually to even the most critical. Most problems in living – those between persons as well as those between nations – have a sensible solution, but it is a matter of common sorrow that this is so rarely applied. No psychiatrist is himself perfect, but he is a specialist in interpersonal relations. In a world where perfection is only a rumour the lame can sometimes teach the paralysed to limp.

One of the limitations of psychiatry has always been the amount of time which the therapist should devote to each patient. Orthodox psychiatrists may devote several hours to the investigation of one case, and treatment may entail many more. An analysis may require two or three hours a week for a year or longer. A second limitation is that the patient's problems are often ones of social adaptation. He sees his doctor in private, however, takes his advice to heart and goes away to try and put it into practice. How much better it would be if his doctor could see him in his trouble and try to help him adapt on the spot. Both these limitations are removed to some extent by group psychotherapy. This implies that a group of patients is treated together rather than individually. It is not however merely several individuals receiving treatment at the same time. For the patient the group is an entirely different situation

from an individual treatment session. His mental orientation is altered by the need to consider the other patients, and he forms emotional ties with them as well as with the therapist. This group situation brings to the forefront social urges which many psychologists would attribute to a primary instinct – the herd instinct; it is in fact an attempt to reconstruct the social pattern in miniature.

Originally, group treatment was used to expedite the after-care of patients with physical complaints. The first recorded use of group therapy was the experiment of Dr J. H. Pratt in Boston. In 1905 Dr Pratt arranged a series of meetings for a group of tubercular patients and gave them formal instruction in hygiene and careful living. The original impetus for the experiment was a desire to save time, but Pratt soon noticed that the patients enjoyed meeting together and gave each other mutual encouragement. He was not slow to develop his discovery and soon started groups for the treatment of other chronic diseases, such as raised blood pressure and diabetes. So successful were his results that his methods were quickly adopted by workers in other parts of America. In the field of mental illness the benefit of group activities had been noticed earlier. In 1904 Dr Camus and Dr Pogniez working in the Salpêtrière (Paris) published a report on their experience with the prolonged rest treatment devised by Dr Weir Mitchell. They noted in passing that the patients in the large public wards were considerably more cheerful than the patients treated in private rooms. These workers however did not appreciate the importance of this observation, and the credit for the first conscious use of group psychotherapy must go elsewhere. By 1909 Dr Marsh was experimenting in America with the group treatment of the psychoneuroses. In Vienna in 1911 Dr Moreno began his famous experiments with problem children. He treated these patients, too young for formal analysis, by methods analogous to play therapy. He encouraged the children to act out their fantasies in miniature plays, and thus laid the foundations of psycho-

drama. These methods soon caught on and it became clear that patients who received group therapy showed greater improvement than those treated individually. They were getting something extra. What is the cause of this extra benefit?

Psychotherapy has two main weapons – a direct appeal to the intellect and an indirect approach through the emotions. The emotional situation which arises between doctor and patient is far older than the study of psychological medicine and has in fact always played a part in medical treatment. In psychotherapy this emotional bond assumes a very important rôle and has been dignified by the special name of 'transference'. The formation and subsequent control of a 'positive' (helpful) transference between the therapist and his patient is one of the most difficult aspects of psychotherapy. A positive transference is easily acquired to start with, but as treatment progresses and the patient is faced with unwelcome home truths and the loss of some of the protection he obtained from his illness he evinces marked hostility. This is the stage of 'negative' transference and unless it is handled carefully he may break off treatment. Even a favourable emotional situation can sometimes interrupt treatment as effectively as it can aid it. A crucial point in psychotherapy is the careful dispersal of the therapeutic transference when the treatment is drawing to a close. Mistakes here will lead either to a pathological dependence on the analyst or to a sudden loss of all the ground so exhaustingly gained. It is in its different approach to the transference situation that group psychotherapy makes an important advance.

The basis of the emotional life is very imperfectly understood, but it is clear that normal people have a great need to love and be loved. Part of this emotional pool is used up on individuals – parents, siblings, spouse, friend, lover – some of it is directed towards groups of people – the village, the office, home and country; or even towards symbols or abstract conceptions such as the flag or democracy. It

seems obvious, and practice appears to bear this out, that group treatment offers a more natural environment for the adjustment of some emotional disturbances than does the intense focal point of individual analysis. Group therapy also enables the patient to form more diffuse and less intense emotional ties while under treatment, and he is spared the necessity of switching them on and off suddenly. Friendships formed in the group may be continued into everyday life. Group therapy is not a comprehensive panacea and should be part of a wider therapeutic plan. First the patient should be given a lengthy diagnostic interview at which a comprehensive history of his illness and the background of his life is recorded. This should be followed by a thorough physical examination to exclude or evaluate any organic disease. If any doubt remains, special pathological tests will almost always settle the point. Only when the patient has been thus diagnostically assessed is it possible to lay down a rational plan for treatment. Group therapy may be the main treatment suitable for a particular patient, but even so it must be supplemented by individual therapy. Such individual treatment should if possible be given by the doctor who runs the group. On the other hand, group therapy may be an adjuvant to the physical methods of treatment.

The conditions which benefit most from group psychotherapy are the psychoneuroses – particularly anxiety states and reactive depressions – and the psychosomatic illnesses such as peptic ulcer, high blood pressure and asthma. In these latter conditions worry and tension aggravate the organic lesion if they do not actually cause the physical changes. As every member of the group will influence the progress of the others, the selection of patients suitable for a group requires considerable clinical acumen. The inclusion of one unsuitable patient may completely disrupt the group and block further progress. On the one hand, the acutely disturbed or deeply depressed patient will distract and upset those less ill. On the other, the supercilious, argumentative

intellectual will spread a highly contagious atmosphere of non-cooperation. It is not essential that all the patients start at the same time, and in practice it is found helpful if a new patient joins a pre-existing group: old members make him welcome and explain the mysteries to him. Those who have never had any psychotherapy before resent the idea that their troubles are primarily mental and are inclined to doubt the efficacy of mere talk. The personal story of recovery from symptoms very similar to the new patient's own, told by one of the old members, carries more conviction than all the blandishments of the most enthusiastic psychotherapist. The latter also appreciates the piecemeal arrival of his clientèle because it gives him an excuse to reiterate his fundamental concepts.

The number of patients admitted to a group depends on the type of treatment. The 'deeper' the therapy (i.e., the more analytical) the smaller the number which can be satisfactorily treated. Two or three patients treated together will automatically form a group, while eight to twelve is probably the ideal number. With more than twenty members the session is likely to degenerate into a lecture and while this didactic form of treatment can still be very helpful, it lacks the plasticity of a small integrated group. At the first session, then, the therapist will find himself surrounded by headaches, backaches, giddiness, weak legs, indigestion, palpitations, depression, tremors, paralyses, pains before the eyes, pains behind the ears, pains all over and just plain nerves. With examples and explanations in simple terms the therapist shows how symptoms like these can be caused by mental conflict and worry. He draws everyday analogies to illustrate these mechanisms at work. Fear, he reminds the group, can cause sweating, trembling, nausea and diarrhoea - physical symptoms; a sudden shock - even the mere sight of blood - may cause loss of consciousness; tension and awkward postures cause muscular discomfort and pain. He emphasises that rigorous medical examination has excluded any physical cause for their symptoms and

leaves them to draw their own conclusions. The therapist proceeds to give a short series of lectures on simple psychological mechanisms. He will explain how we repress unpleasant ideas and how the emotions attached to them may still influence our thoughts and actions. When the patients have become used to each other and have acquired some knowledge of psychological mechanisms they are required to write their own life stories in the light of this new knowledge. Selected passages are then read out - anonymously - and everybody is invited to comment on them. It is illuminating, but not altogether surprising, that patients will often recognise more easily in others than in themselves the psychological causes of symptoms. Making this step, of course, helps considerably towards the attainment of insight into their own troubles. Throughout group treatment the patient is seen individually at frequent intervals when the more personal points and intricacies of his own illness are discussed.

In many clinics this fundamental educative side of the treatment plays a minor rôle and most of the benefit is derived from the emotional re-adjustment obtained. Many patients have never really attempted to integrate their lives with their neighbours: social conventions have dammed back the primitive expression of their drives and they have never acquired the ability to enjoy the compensating outlets that civilisation offers. Other patients have become so pre-occupied with their symptoms that they have withdrawn from social activities. The psychotherapist therefore attempts to establish a communal organisation; he encourages the patients to take up hobbies and to develop any interests they have in common. He lays emphasis on the duties of the individual to the community, encourages co-operation and friendship between patients, and by pointing out how everyone can contribute in some way to the general good, bolsters up their self-respect. He teaches them muscular relaxation, which besides having a beneficial physiological effect renders them more suggestible and paves the way for

the acceptance of encouragement and reassurance. Many therapists, especially in America, conclude each treatment with exhortative passages from the poets and minor philosophers. The more inspirational a therapist's technique the more closely does a group session resemble a revivalist meeting. Indeed, such movements as the Oxford Group use similar methods. 'Alcoholics Anonymous,' a group formed by alcoholics for the treatment of alcoholics, emphasises that a fundamental step towards the success of their treatment of the chronic drunk is a change of face equivalent to a spiritual conversion. The mass cures, which in the past accompanied outbreaks of religious fervour, were spontaneous examples of the mechanisms underlying group psychotherapy. Special methods such as psychodrama are extensions of the same principles. The patients themselves act out their conflicts and seek solutions in imaginary situations.

In assessing the results of group psychotherapy it must be remembered that it is the chronic neurotics who respond best and these patients are notoriously unresponsive to other methods of treatment, usually continuing in a state of interminable disability. The modest achievements which can truthfully be attributed to group treatment are therefore a decided therapeutic advance. In one follow-up questionnaire 90 per cent of patients replying said that they had improved, and most of these to a considerable extent. When allowance was made for those who did not reply, and for those who had defaulted from treatment, the proportion showing improvement fell to just under 50 per cent. Group therapy takes deliberate advantages of man's gregarious nature, and is thus planted on a firm theoretical basis. Further, it has stood the test of time. It can therefore safely be added to the armamentarium of the psychiatrist, but it must be remembered that group therapy is not a specific and when further light has been thrown on the aetiology and pathology of the neuroses it may well give way to a more direct method of attack.

Research Report

A. W. HASLETT

Muscle measurements.

FOR sheer delicacy of measurement and finesse of experiment it would be difficult to beat some recent observations by Prof. A. V. Hill of University College, London, on the heat changes which take place during a single twitch of a muscle. The rise in temperature which accompanies such a twitch is only about 0.0025 degrees Centigrade. This in itself would present no particular difficulty, for although it represents a small enough temperature change by ordinary standards, very much smaller differences in temperature can be measured. The real difficulty was that, in order to follow a single muscular twitch through all its stages, and particularly to record accurately its beginning, Prof. Hill needed to measure both the contraction of the muscle and its temperature at such speed that the state of affairs at intervals of time only two thousandths of a second apart could be distinguished. It was the combination of these two requirements of speed and sensitivity which made the measurement problem difficult.

The first stage was to translate the temperature changes to be measured into changes of electric potential. This was done by a small electric thermometer or thermopile about as thick as a red blood cell is long. Its maximum output, under the conditions of the experiment, was about four millionths of a volt. This in turn was translated into a small movement of a mirror attached to a fine wire coil suspended between the two poles of a magnet. In outline this is a standard arrangement. The idea is that a beam of light is directed on to the mirror, and the deflection of the light beam when the mirror moves is observed on a distant scale.

The details of the magnification system which Prof. Hill employed cannot be given. But the effect in the final display was as if the light beam from the mirror had been observed from a distance of some 285 yards down the road. And the main point was that to stand such enormous magnification and at the same time give the required high speed of response, and steadiness of movement, the mirror system had to be almost perfectly balanced and provided also with a form of mounting which was almost completely immune to outside vibration. The key part of the instrument was constructed in Prof. Hill's laboratory by Mr A. C. Downing.

The measurements, about which all this trouble were taken, were made on the leg muscle of a frog at a temperature of zero Centigrade. Under these conditions, the contraction period of a single twitch was spread out over about 0.4 seconds, and the following period of relaxation lasted about 0.6 seconds. And to recover to its original state, the muscle required some 45 minutes.

No new or sensational theory of muscular action has been put forward on the strength of these measurements. Merely, a few concrete facts have been brought together which any future theory will have to explain. One is that, although the muscle takes from 7 to 10 thousandths of a second to 'get going', it is impossible to separate in time the beginning of the release of heat and the beginning of the physical movement of contraction in the muscle. The two effects are simultaneous within the two-thousandths of a second accuracy of the experiment. Second, the release of heat appears to take place straight away at its maximum rate. The effect is as if the muscle was already 'set up', ready for action, and needed only to be released or 'triggered' by a nerve stimulus. Third, there is the peculiar fact that in a single muscular twitch, the initial release of heat in contraction is independent of the amount of physical work done by the muscle. 'It would be a very odd kind of cartridge,' as Prof. Hill has said, 'in which the energy released in the explosion had no apparent connection with the muzzle

velocity of the bullet.' And finally, there is no sign of the absorption of heat, as opposed to its release, at any stage of the complete cycle of contraction, relaxation and recovery. Negatively, this seems to disprove fairly decisively one theory of muscular action which has been much discussed. But neither is there any other complete or consistent theory which can be put forward. Prof. Hill, who should know, thinks that the time is not yet ripe.

Hill. *Proceedings of the Royal Society* (in the press).

Coral island.

One of the advantages of research conducted for the United States services is a certain lavishness of effort, with a tendency to follow up sidelines which are of more interest for their own sake than for their ostensible purpose. An example was the explosion of the Bikini atom bomb which, among other activities, led to the deepest boring yet undertaken on any coral island. A report which has lately been issued shows that this was carried to a depth of 2,556 feet, and it is seriously suggested that an 8-10,000 feet boring should now be made. Also, a large number of deeper soundings were made by the method of setting off explosive charges near the surface and observing the times at which earth waves were recorded at different points on the atoll.* By this second method it was established that the depth to basement rock was about 8,000 feet in the centre of the lagoon, but there were variations of many thousands of feet in the corresponding figures for different parts of the complete atoll.

The interest of what was done at Bikini, and still more of the proposed deeper boring, is in relation to theories of the growth of coral islands. The picture which is still, in the main, generally accepted is that which was first put forward by Charles Darwin on the strength of observations which he made during the same voyage of the *Beagle* which led him to his more famous theory of evolution. This was the first attempt to cover in a single theory the three chief forms

*See *What the Earth is Made of* (page 63).

of coral growth. These are the fringing reef, the barrier reef, and the atoll or coral island. By the first of these is meant the type of reef which grows directly out from the coast and at low tide can be seen to be connected with it. The barrier reef, on the other hand, looks to be a separate structure from the continent or island which it separates from the ocean beyond. The Great Barrier Reef of north-eastern Australia stretches for about a thousand miles, acting as a more or less continuous breakwater separating Queensland from the Coral Sea. On the other hand, in the typical barrier reef island, the barrier reef is an irregular circle, with the island in the middle, and usually a break in the reef on the leeward side, and safe anchorage in the lagoon between reef and island. Finally, in the atoll or coral island (in the literal sense) there is no central island, but only a broadening of the coral reef on which vegetation has become established.

Darwin's suggestion was that, in the typical case, these three types of formation could be regarded as stages in a single process. He supposed that the original island was volcanic, as many Pacific islands obviously are. In the first stage, he thought, a fringing growth of coral would become established round the edge of the island - although, in fact, there are no known examples. Then, in the second stage, the island would begin to subside, and the reef would grow upwards to keep pace with the subsidence, but would now be reared above a part of the island which was submerged, giving the effect of a barrier reef. Finally, in the third stage, the original island would have subsided so far as to have totally disappeared, and only the barrier reef to which it had given rise would be still visible above the surface.

The theory is attractive in its simplicity, but fails completely to explain one further fact - that the depth of nearly all atoll lagoons lies between about 120 and 240 feet, with a relatively flat sea bed at that level. There is no obvious reason why every subsiding volcano should consistently stop subsiding when its summit is at the same depth below the

surface. Nor can the theory, by itself, account for the very large amount of debris needed to level off the gap between the sloping sides of the former volcano and the barrier reef. It was therefore pointed out by the American geologist H. A. Daly that during the period when the continents of Europe and North America were covered in ice sheets, the level of the oceans would have been lowered by some 300 feet. During this period, any pre-existing coral would have been unable to survive, and he thought that the effect of waves and storms would have been to wear away the top of the island to a roughly level platform. Then, about 25,000 years ago, when the level of the oceans would have begun to rise again, from the melting ice of the continental sheets, coral would have become established round the edge of these eroded platforms, and the final result would be a barrier reef, with a central lagoon of roughly the required depth. It is against the background of these two theories, which many geologists believe can be usefully combined, with the possibility of more violent movements at intermediate stages, that the results of test borings through deep coral structures are most naturally considered.

Fifty years ago, the first such boring was undertaken by the Royal Society on Funafuti atoll 500 miles north of Fiji, and since then there have been further borings on the Great Barrier Reef of Australia, in the East Indies and in the island chain south of Japan. But by a substantial margin there has been none so deep as that lately carried out on Bikini atoll, nor was there the same information about the depth of the basement rock. Only the preliminary results of the Bikini boring are yet available. These bring out two main points. The first is that the layers of the atoll foundation which are now 300-575 feet below sea level appear to have been at one time exposed above sea level. This suggests that, as well as the two main processes, there must at one stage have been a quite considerable upheaval. The second is that the layers from 1,790 feet down to the lower limit of 2,556 feet are lower miocene and must have been formed

therefore between 20 and 30 million years ago. This would imply an average rate of growth not more than one-thousandth of the maximum rate of a foot in ten years which is possible for coral structures. In other words there would be an ample margin for interruptions and reversals of movement. Consequently, if money, time and equipment, can be made available it would be a project of real interest to explore a single deep structure down to and into the basement rock beneath. The suggestion is that the boring should be in the centre of the lagoon, so as to give the earliest chance of reaching the basement rock; and that, in conjunction with a sunken landing barge, the existence of a number of flat-topped platforms reaching within twelve feet of mean sea level, would make such a central boring possible.

Ladd, Tracey and Lill. *Science*, 1948, 107, 51.

Radar and colonial survey.

Methods of mapping from the air, with radar control to fix position, which were developed during the war for use in the Far East, are now being applied on a large scale in Africa by the Colonial Survey Organisation. Since April of last year upwards of 150,000 square miles have been photographed in this way. This includes the whole of Swaziland, and parts of Bechuanaland, Tanganyika and the development areas of Uganda, as well as some further work in Kenya. Priority has been given to those areas scheduled for ground-nut planting, and the whole programme has been tackled with a speed and urgency which would have been impossible with conventional methods of surveying. There is also the advantage that, pending the translation of air photographs into maps, the photographs themselves are immediately available for any purpose for which large numbers of copies are not required. These are on a scale of 1 in 30,000 - or not quite two inches to the mile. Mapping from the air photographs is limited at present to a rate of 80,000 square miles a year, so that the bottleneck is on the

mapping side rather than in the taking of the photographs on which the maps are based.

The system of radar control is based on measurements of the range of the aircraft from radar beacons installed in convenient positions on the ground. The main transmitter, with a system of this kind, is in the aircraft, and the beacons on the ground can be thought of as so many reflectors from which radar echoes are returned to the aircraft. This is the reverse of the original air-warning use of radar, when the transmitter was on the ground, and the echoes which were measured came from the aircraft. The only difference otherwise is that instead of relying on simple reflection, the radar beacons are provided with auxiliary transmitters which are triggered or 'set off' by the aircraft transmissions, thus giving stronger reply signals. The measurement which is obtained is that of the range of the aircraft from the beacon, and it is usual in survey work to fly on a series of circular courses about one of the beacons. In this way, if photographs are taken at regular intervals of time, and successive courses are accurately spaced, it is possible to make certain that there are neither gaps nor overlapping. With the addition of a second beacon, or ground station, the position of the aircraft is fixed. The ranges of the aircraft from both are directly displayed, and arrangements are made for a photograph to be taken of the instrument panel showing these two ranges, the height of the aircraft and the tilt of the aircraft at the same instant that each survey photograph is taken. The earliest experiments with this method were carried out in Anglesey, and later the Colonial Survey Organisation carried out further trial surveys in West Africa. The present surveys in central and east Africa are believed, however, to be the first large-scale application to practical mapping. The limit of accuracy is stated to be about 75 yards.

Prof. C. A. Hart of University College, London, who was lately appointed to the first university chair of surveying and photogrammetry, was directly concerned with the

earlier wartime development which led to the present application in Africa and is continuing research on the still more accurate development of radar control methods. An appreciably higher standard of accuracy is, in fact, already attainable – but the equipment is neither so simple nor so mobile as that used by the Colonial Survey Organisation. Also as might be expected, the amount of work required in translating air photographs and radar control measurements into accurate maps increases considerably as the standard of accuracy is pushed up. There is therefore a substantial gap technically between the type of survey which is now being carried out in Africa, and the revision of home Ordnance Survey plans, on which there are heavy arrears waiting to be caught up. At the largest scale used by the Ordnance Survey, an error of 75 yards would represent a distance of about two-thirds of an inch on the final plan. There seems little doubt that Ordnance Survey standards could be reproduced by a combination of radar control and air photography. But although this may be possible technically, it would clearly be unwise to assume that, because the radar method is rapidly proving itself in Africa, the ground surveyor and his traditional 'chain' have been rendered universally obsolete.

Radio and the sun.

Radio methods of observation look like giving new information about sunspots. It has been known for some time that abnormal bursts of radio radiation appear to come from the direction of the sun at times when the sun's disc is unusually disturbed – that is, when there are many or large visible spots – and it has been calculated by Sir Edward Appleton that, under these conditions, the sun is the most powerful radio transmitter of which we have knowledge. It was left, however, for M. Ryle and D. D. Vonberg at the Cavendish Laboratory at Cambridge to make the first measurements of direction sufficiently accurately to enable an estimate to be made of the width of

the active area on the sun's surface. They found that, at the most, the width of this area could not be more than about ten minutes of arc, or some 300,000 miles on the sun's surface. Any error in their measurements would have made this an over- rather than an under-estimate – and as a spot measuring not quite 150,000 miles across was visible at the time, it seemed a reasonable guess that the visible sunspot and the observed 'radio spot' covered roughly the same area. Since then, the sun has been kept under continuous radio observation, in quiet times as well as disturbed, over a period of some ten months. Records have been made on two different frequencies, 80 and 175 megacycles respectively, corresponding to wavelengths of 3.75 and 1.6 metres. And it has been found that one useful indication of conditions on the sun is the proportion of energy which is put into these two frequencies. Under disturbed conditions, the amount of power radiated on the higher frequency is several times greater in proportion. Finally, taking this change in energy distribution as the criterion, it was noticed that there was a tendency for these abnormal periods to recur at 27 days intervals, corresponding with the sun's known period of rotation. This, in itself, is not particularly surprising. Sunspots themselves are liable to recur at similar intervals and for the same reason, although it is unusual for even a large sunspot to make more than one or possibly two repeat appearances. What is interesting, and rather surprising, about these Cambridge measurements is that, during the whole of this ten months' period, there seem to have been only three main areas of activity on the sun. One of these has survived for as many as 11 rotations, another for 7 rotations and the third for 5. It is early to draw definite conclusions, and Mr Ryle has himself pointed out that these results cannot yet be regarded as providing sufficient proof. But at their face value, they suggest that to the radio observer disturbed regions on the sun's disc have a very much longer life than they present to the astronomer; and that visible sunspots

should be regarded as only one stage in a longer history of disturbance, of which the earlier or later stages cannot be seen.

Ryle. *Nature*, 1946, 161, 136.

Film of the sun -

Considerable and well deserved attention has been attracted among astronomers to a film of the sun made by Prof. H. D. Menzel at the Climax, Colorado, high altitude observatory of the Universities of Harvard and Colorado. A copy of the film has been presented to the Royal Astronomical Society and this has been shown lately in London. It was taken through a special type of filter and shows the behaviour of prominences seen at the edge of the sun's disc. The general effect can only be described as that of a fireworks display on an immense scale, originating in a surface of almost continuous disturbance and movement, and with a variety of forms which any manufacturer of fireworks could only envy. These included narrow luminous jets, shooting rapidly outwards and looking (probably deceptively) to return along the same line on which they had started out; a great eruptive surge, developing into a luminous mushroom spreading outwards from the sun's disc; a cloud suspended, as if held in its place by electrical forces; and an inverted cone, in much the same shape as an umbrella blown inside out by the wind. It could be seen also that many luminous 'shoots', for example between one cloud and another, followed curved courses as would electrically charged particles if moving under the influence of strong electrical or magnetic forces; and, visually, at least, one got the impression that inward moving streamers were being sucked into particular areas on the sun's surface. Films of this kind, taken in broad daylight, of events at the outer edge of the sun which would normally only be visible during total eclipse, were first made by M. Bernard Lyot, the French astronomer. Prof. Menzel's set a standard which will take some beating - although, naturally, the ultimate value of such a film is not in the visual effect which

it produces on the spectator but rather in the mass of information which it provides for the formulation and checking of theories of solar activity. From that point of view, the dominant impression must be that conditions on the sun are not so simple as the contributions of theoretical astrophysicists would sometimes suggest.

- and of Penguins

Another scientific film which was shown lately in London was made as part of the work of the Falkland Islands Dependencies Survey. It includes a series of motion studies of Adélie Penguins, which, in the hands of competent zoologists, should be a useful supplement to the extensive observation made earlier by Levick on the same species and recorded in his book *Antarctic Penguins*. Many shots are included of penguins jumping from sea on to land ice, with an angle of take-off of about sixty degrees, with a slide in toboggan position on the ice as the usual ending. Points which could be noted were the quickness of pick-up to an upright position, and the balance shown afterwards in glissading, still upright, down a steep snow bank. Except immediately after a jump, they would use their tobogganing method for such a slope. But having jumped, slid and got up, they prefer to remain so. When the tobogganing method was used for straightforward travel on snow, with both feet and flippers used for propulsion, it compared very favourably with the similar efforts of a Weddell seal, humping its back like a caterpillar at each forward movement. So few zoologists have had the opportunity of studying the movements of penguins at first hand that it would seem well worth while making a slow-motion version of this and any further film studies of the kind which may become available.

Blood and its evolution

The red cells of the blood play so necessary a part in the conveyance of oxygen to all organs of the body that one

would have expected the manner of evolution of the blood pigments to have received more attention from zoologists than has so far been the case. Prof. H. Munro Fox, of Bedford College, London, is among the very few who have worked seriously on the problem, and he has lately obtained quite convincing evidence that the blood pigments should be regarded as biological accidents – that is, that having originally been used for some other function, the incidental advantage given by their oxygen-carrying capacity was then ‘seized upon’ by evolution and developed into one of the key attributes of all higher forms of life. The most obvious indication comes from the multiplicity of these pigments. Four different substances are known, which carry out this same job of oxygen transport. Three of them are compounds formed by the union of a protein with three comparatively simple iron-containing substances. The fourth, instead of iron, contains copper. A more precise indication comes from the water-flea *Daphnia* which, like a human being or mammals generally, may possess haemoglobin – but, at least in adult life appears to get along equally well with or without it. It has been shown in Prof. Munro Fox’s laboratory that individual *Daphnia* can gain or lose haemoglobin within a few days; that they swim as vigorously when apparently ‘anaemic’ from its absence as they do when red-blooded; and that water from ponds containing red-blooded *Daphnia* can induce the development of haemoglobin in initially pale individuals from other ponds. Whatever reason, therefore, may have led to the appearance in *Daphnia* of haemoglobin, it cannot now be used for oxygen transport. Prof. Munro Fox’s second piece of evidence was obtained during a visit to the Marine Biological Station at Naples. While working there he found, first, that there were families of marine worms of which some species contained haemoglobin, while others, closely related and living under similar conditions, contained a different blood pigment, chlorocruorin. Finally, he found a particular family, *Serpula*, in the members of which both pigments

were to be found in the same individual and at the same time. If both had been evolved in response to other and possibly different needs, and then found useful as oxygen-carriers, it would be an odd coincidence, but not in itself particularly unlikely, that both substances should be present in the same species. Haemoglobin could have originated through one process, and chlorocruorin through another, and both independently been pressed into service as carriers of oxygen. But if the need which was first filled had been that of oxygen transport it would be unlikely that two separate mechanisms would have become established, when either by itself was sufficient to do the job.

Fox, *Nature*, 1947, 160, 431, 825.

Meson particles.

Twenty years ago, it was supposed that nature made use of only two kinds of fundamental particle as ‘building bricks’ for the construction of atoms, and so ultimately of the material universe. These were the negatively charged electron – the atom, if you like, of electricity – and the positively charged proton, otherwise the nucleus of the hydrogen atom, with a mass some 1760 times as great. To these, in 1932, Sir James Chadwick added the neutron, which was without electrical charge and of roughly the same mass as the proton. It was also discovered that positive electrons could exist as well as negative, and to match the positive electron it was thought possible that negative protons might also exist. But for almost a generation it was taken as axiomatic that no particles were likely to exist of masses intermediate between those of the electron and proton. There were electrons, the lightest particles known, and there were protons and neutrons – but nothing in between.

Then, in 1935, on purely theoretical grounds, the Japanese physicist Yukawa postulated the existence of an intermediate kind of particle with a mass about one hundred times that of the electron. No one at the time paid much attention because the existing picture appeared reasonably

cut-and-dried and, in the absence of experimental evidence, few contemporary physicists felt inclined to upset it. In 1937, however, Dr Carl Anderson of the California Institute of Technology produced clear-cut evidence for the existence in cosmic radiation – radiation, that is, reaching the earth from outer space – of a positively charged particle with a mass estimated at 200 times that of the electron. The real existence of meson particles had therefore to be accepted, and an enormous amount of experimental work has since been done on the observation of cosmic radiation mesons. Again, the tendency at first was to be conservative. Measurements suggesting that mesons might have more than one mass were critically examined, and reasons were found for distrusting the evidence, and for sticking to the simplest interpretation – namely that meson particles did exist, but that they could all be taken to be of the same mass. The pendulum has now swung for the second time in the opposite direction. At the University of Bristol, Lattes, Occhialini and Powell have shown fairly conclusively that there must be at least two meson masses, one twice as great as the other. And at Manchester University, Rochester and Butler have produced evidence pointing to a third possible meson mass, heavier than the other two – as indeed had been suggested by the French physicist Leprince-Ringuet a year earlier. Finally, there is a third suggestion, on much less direct evidence, of a considerably lighter kind of meson. All this is clearly going to take time to sort out. And the purpose of the present note is merely to place it on record that physics is once again becoming more complicated; and that not only do meson particles exist of mass intermediate between that of the electron and proton, but also that it is practically certain that mesons exist of more than one possible mass.

Lattes, Occhialini and Powell. *Nature*, 1947, 160, pp. 543 and 486.
Anderson, Adams, Lloyd and Ross. *Physical Review*, 1947, 72, 724.
Rochester and Butler. *Nature*, 1947, 160, 855.
Auger, Daudin, Fréon and Maze. *Comptes Rendus*, 1948, 226, 169.

Operational research.

So much has been said and written about the possible application of the methods of operational research to promote industrial production that a concrete example which has lately been instanced may be worth repeating. It is due to the Shirley Institute, the cotton industry's co-operative research establishment, and consisted in the quantitative investigation of the reasons for varying output in different mills. The method was to take each process separately, collecting figures for the amounts of cotton processed, numbers of workers, hours worked, types of plant, equipment and so on. The figures showed that for the same process and the same count of yarn, the number of man hours required per pound might vary by as much as three to one. By studying some of the apparently best and some of the apparently worst mills for any particular process, a good idea was then obtained of the reasons for these differences, practice in labour deployment, package size, lay-out or whatever it might be.

Up to a point, this could be described as 'merely common sense'. But it was of the essence of the inquiry that individual stages should be separately reviewed, and differences thus brought to light which, if overall efficiency had been made the yardstick, might have largely balanced out. The results, in any case, suggested that by bringing the practices of all mills up to the level of the best individual rooms, increases in output ranging possibly up to forty per cent might be obtained, with the same labour force and machinery. Whether this is called 'science' or 'common sense' applied systematically and through figures, the fact remains that it needed an investigation by trained scientists familiar with the industry to show how much could be done. This example was quoted by Mr Herbert Morrison who, as Lord President of the Council, is responsible for the Department of Scientific and Industrial Research.

Sir Edward Appleton, as Secretary of the Department, has also given lately his own abbreviated definition of opera-

tional research. 'It is the arithmetic,' he said, 'of "what happens when we do what",' adding that 'recent studies have shown that in some cases industries can take immediate steps by which output could be increased by as much as 20 per cent'. Other industrial research organisations quoted as having taken a prominent part in investigations of the same type were the Iron and Steel Research Association and the Refractories Research Association which deals with the manufacture of heat-resistant materials. For comparison, it may be recalled that one of the more spectacular wartime successes of Operational Research – in doubling the effective anti-submarine patrol time of a Coastal Command Group – was achieved by no more abstruse methods than a rational and critical overhaul of maintenance procedures, based on statistics which were no more difficult of collection.

A.W.H.

*

Nullarbor, Australia.

A recent holiday expedition from Port Lincoln in South Australia across the Nullarbor plains to Perth by a party of Melbourne naturalists has whetted the appetites of Australian scientists for further expeditions into this scientifically interesting but little-known region.

Travelling in a capacious tourist coach, the party followed the route of one of the grimmest and most courageous journeys in the history of Australian exploration, that made in 1841 by Edward John Eyre, the first to cross the 1,000 miles of waterless desert country (plate 16) and reach the west coast by land. Water tanks at intervals of about 100 miles along the road through the desert to-day provide water, potable though mineralised, for travellers, but the tanks are not to be relied on in dry seasons.

There is no accommodation for travellers – only a few scattered homesteads, each the centre of a million-acre sheep run, where courageous hard working and resourceful settlers carry on Australia's pioneer traditions. With their

nearest neighbour some 70 miles away, these settlers must rely on the 'Flying Doctor,' summoned by pedal wireless, in cases of sickness. Their children are pupils of the 'School in the mail box,' Australia's celebrated system of education by correspondence for children who live in remote areas where there are no schools.

Beneath the surface of the Nullarbor plains lies a far-stretching underworld of limestone caverns (plate 17), deep lakes and waterways. The 500 feet deep limestone deposit is a survival from the miocene, when the plain was part of the bed of the sea. From these caves, the scientific party brought back two rare carved stalactites composed of mineral gypsum instead of the usual calcite: these gypsum stalactites have never before been found in Australian caves, though they were previously described in the Mammoth Caves of Kentucky and a small cave in Southern France. (See plate 18).

The holiday exploring party was organised and led by Mr Russell Grimwade, President of the Melbourne National Museum trustees and a leading Australian drug manufacturer. It comprised a forester, a zoologist, an anthropologist, an ornithologist, and seed-collector and a naturalist-journalist, as well as the entomologist director of the Melbourne Museum and a botanist of the staff of the Melbourne Herbarium. All brought back many interesting specimens, aboriginal chopping axes (plate 19), parasitic plants, metallic blue spiders which hunt metallic blue ants, and the strange plant *kingia* which grows a foot in a century and lives for several thousand years.

Whalemeat

The whale is not a fish. Like the cow, it brings forth its young as calves and nourishes them on milk. Like the cow, it is a warm-blooded animal, with the vestiges of four legs inside its flippers. And like the cow we eat it. A blue whale, 100 feet long, provides about 12 tons of meat – the muscle which enables it to swim so powerfully, perhaps at 20 m.p.h.

or so, for a couple of hours at a stretch. This muscle is very fatty, and may contain anything from one to ten per cent of extractable oil. In the past it found a use, after the extraction of this oil, as a high-grade cattle food or a rich guano-type soil fertiliser; now the world food shortage has brought it on to the human dinner-table as a very valuable protein food. The blue whale's liver, which weighs about a ton, is also nutritionally important, for 3 per cent of it is oil rich in vitamin A.

But whalemeat, in its living state, is interesting from quite another point of view. The performance of the animal is so striking, both in speed and endurance of swimming, and in sheer strength, that one naturally wonders whether its muscles can possibly be the same in structure and function as those of the human body. Surely they must be much more efficient machines than those in our arms and legs. Otherwise, how could a whale tow a small steamer forward at seven miles an hour when its engines are racing full astern?

One way to compare the efficiency of machines is to compare the horsepower generated for a given weight of engine. Now human horsepower has already been worked out, and it seems that an athlete, like the dog, generates about a hundredth of a horsepower for each pound of muscle he has in his body. To get this result was relatively easy, because a man can be made to work away, for instance at pedalling a stationary bicycle, under laboratory conditions which permit of precise measurements. The whale is a more difficult case, and one or two assumptions have to be made, for instance that when towing the steamer it is doing as much work as would a submarine of the same shape and displacement. In other words, it is like a rigid steel body forging through the water, with a similar water skin friction resisting its passage. It may be, however, that the whale, being flexible and not rigid, and oiled all over, has less water-resistance to overcome, and wastes less energy in producing wake-eddies than the submarine. On the other

hand, experiments towing anaesthetised fish suggest that in fact they have much more friction, and not less, to overcome than a vessel of the same size and shape – but this may be due to their scaly rough surface.

Anyway, the assumptions probably do not completely spoil the calculation. They do not give the whale's true horsepower, but they do show whether this horsepower is a hundred times greater (or smaller) than the human. They are accurate to that extent, and after all, that is what accuracy is – always a relative matter, not an absolute thing. According to the experts, the value for the whale comes out at about 9 horsepower per ton of muscle, which is a figure lower than that for man, after all (four thousandths against one hundredth). So the answer seems to be that whalemeat is just like human meat, mechanically speaking, and the whale's fantastic power is the result only of its fantastic size. It has really no more kick to it than a cow.

Folic acid again

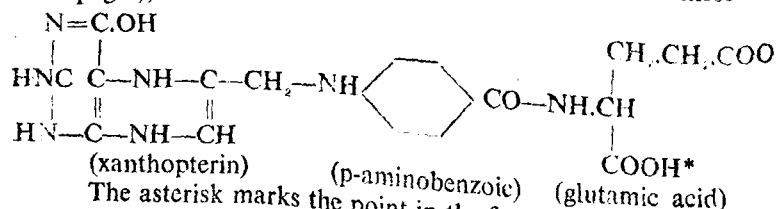
Since the report in an earlier issue of *Science News* of the discovery of this new water-soluble vitamin which is present in the leaves of plants as well as in liver and in yeast, further results have appeared to modify the original conclusions, and to justify a fresh account of the whole matter.

As in a great deal of modern vitamin research, workers got on the track of an unknown when they attempted to keep young chicks on a completely synthetic diet; one, that is to say, which contained only known proteins, fats, carbohydrates and minerals, and all the pure known vitamins. Something extra was needed in the food before the chicks would grow, something that prevented their getting anaemic. Experiments about the same time with various kinds of bacteria on simple casein media containing only known constituents, showed the same kind of result. Before the 'bugs' would grow a supplement had to be added. In the case of *Lactobacillus casei*, this was obtained from liver and yeast extracts; with *streptococcus faecalis* it was found in

spinach. In the early stages there was nothing to show that L. casei factor and 'vitamin B_c' (for chicks) were the same – in fact, the different research groups thought they were after different substances. It was only as they purified the substances they were chasing that they found themselves coming to a similar conclusion. The new vitamin could be separated from protein, adsorbed on a special charcoal and eluted again, and so on, until there was obtained a product with a rather characteristic 'colour' (visible only in the ultraviolet, where only a camera can 'see' it). So gradually three supposed vitamins turned out to be forms of one.

There was also, as it happened, a fourth approach. Since 1931 Dr Lucy Wills has been interested in a special kind of anæmia which comes on during pregnancy in India, and is curable with liver, or extracts of yeast such as Marmite. It was found that a similar anæmia could be produced in monkeys by feeding them the sort of food eaten by the poor in the slums of Bombay, and this too was curable with liver, which therefore contained a new food factor, Vitamin M (M for monkey). Gradually it became apparent that there was little to choose between vitamin M and vitamin B_c, and in all probability they are identical.

The new vitamin occurs in at least three different variations, however, of which folic acid or pteroyl glutamic acid is the simplest. It is made up of a molecule each of xanthopterin (butterfly wing pigment), p-aminobenzoic acid (already known as a member of the vitamin B group), and glutamic acid (an important amino acid, already mentioned in these pages), thus:



The asterisk marks the point in the formula where variation creeps in. Instead of one glutamic acid molecule, there

may be three, i.e., two more attached at the asterisk; or even seven, as though the mechanism which makes the vitamin in the cell had stuttered or stuck in the groove and gone on repeating 'glutamic acid ... glutamic acid ...' These variations are not all physiologically interchangeable: the form with seven glutamic residues, for instance, is effective in the diet for chicks, but useless in the bacterial test. There exists an enzyme in animal tissues which breaks down the more complex forms to simple folic acid.

It was natural, in view of the effect of folic acid deficiency on the blood of chickens and monkeys, to try the effect of it on various human blood diseases. In this way it was discovered accidentally that folic acid would 'cure' pernicious anæmia. That is to say the anæmia would clear up and the blood remain normal as long as the correct dose of folic acid tablets was being taken. The quantity required was quite large – 5 or 10 milligrammes per day. At first it was thought that this would become the modern treatment, replacing the nuisance and expense of crude liver injections. But one aspect of the disease had been overlooked.

Pernicious anæmia itself is an illness in which there is a failure to produce the red cells of the blood; only a few are made, and they are abnormally large. Associated irregularly with this (the causal connection is obscure) there are nervous symptoms, grouped under the heading of subacute combined degeneration of the spinal cord. The nervous symptoms may be marked and the anæmia slight; or the anæmia alone may be present, followed only after some time by the nervous degeneration, which has nothing to do with the severity of the anæmia. It was early found that folic acid, unlike liver, would not affect the nervous side of the illness, but it took about two years of experience to show that people with anæmia only, and this anæmia successfully treated with folic acid, might develop subacute combined degeneration nevertheless. In other words folic acid was dangerous, because it lulled doctor and patient into a false sense of security. It affected the blood, but not

the nerves, and was only a half-cure. The truly curative substance present in liver extracts is still (March, 1948) a mystery: it is not folic acid, but it may be a chemical relative.*

In another illness, however, folic acid appears to be having a marked success. Tropical sprue is a disease with great diarrhoea, and an anaemia somewhat like the pernicious type. In a matter of days from the start of treatment, folic acid often arrests the diarrhoea, the health of the intestines improves, and the anaemia lessens. So far no fly in the ointment has been announced.

* Very recently Dr E. Lester Smith of Glaxo Laboratories, Middlesex, has announced the isolation from liver of a red pigment which is powerful in treating pernicious anaemia, though it does not contain folic acid. Simultaneously, workers at Merck Laboratories, New York, have isolated a crystalline "vitamin B₁₂" which is also powerful in treatment. The relation between the discoveries of the British and American workers, and further news of them, will appear in a later issue of *Science News*.

Minor Additions in Metals and Alloys

DR F. A. FOX

It is to-day becoming increasingly understood that our present level of engineering technology is in large part due to our skill in manipulating alloys. However, although it is well known that, where mechanical strength is needed, pure metals are too weak to do what modern engineering requires, it is not so widely understood that the amount of the critical alloying addition is often remarkably small. Most non-specialists have the haziest idea of how alloys are constituted, and tend to think of additions of tens per cent being necessary before metal B will have much effect on metal A. Some of the best-known and oldest alloys do contain large amounts of the added metal, as for example zinc in copper to make brass, and copper in gold to make the usual jewellery alloys. As, however, the science of metallurgy has developed it has become possible to do two things – first to begin to isolate the causes of differences in behaviour of apparently similar alloys, many of which sometimes refused, with a subtlety which seemed almost perversity, to do what was expected of them, and secondly, by inductive reasoning from observations of the first kind, to formulate fresh alloys which would have some new combination of desired properties. Two other activities, one old and one new, also go on in the metals research laboratories of the country. The first is the much-practised pursuit of trial-and-error; in the evolution of new alloys this method is widely used, is expensive but occasionally very rewarding: it is widely used because metal alloy problems are so complex, and because some of the fundamental guiding prin-

ciples are only just beginning to be discerned. The new thing is the study of metal physics in which physicists and metallurgists and physical chemists undertake fundamental researches into the behaviour of metals and alloys from the point of view of 'pure knowledge'.

The studies in metal physics and also in the more empirical spheres of industrial research have intensified in the last ten years, and the natural result has been to throw up many new ideas and to lead to the revision of many old ones. As is also to be expected, our ignorance becomes correspondingly more obtrusive.

One of the things that modern metallurgical research has emphasised is the importance, often decisive, of the part played by minor additions, constituents and impurities in ruling the properties of metals and alloys. In this context 'minor addition' means some element, phase or compound which is present to an amount less than 1 per cent of the whole; sometimes decisive influences are exerted by an addition or impurity present in far smaller amounts, of the order for example of 0.002 per cent.

The alloying additions which are made to the metals of commerce and industry are usually themselves metallic, such as chromium to nickel to make the elements for electric fires, or copper to aluminium to make duralumin. Some important alloying additions are, however, non-metallic and some are even gaseous. A technically important example is that of steel itself, in which the addition to the metal iron of, say, half of one per cent of the non-metal carbon produces an alloy enormously stronger and more versatile than the parent metal. Another non-metallic material present in steel has in recent years been found to have a profound effect on its heat-treatment behaviour and on its toughness and other properties, and that is alumina. The alumina occurs as a by-product of the deoxidation of the steel by aluminium additions during the steel-making process; the aluminium is itself oxidised, leaving behind minute amounts of alumina highly dispersed through the steel.

The effective particles of alumina have not been positively identified even at the highest magnifications under the microscope, yet they affect the behaviour of the steel in which they occur in a radical way. The effect of gases on the behaviour of metals and alloys is often complex; generally they are harmful, but the addition of oxygen to copper is an example in the reverse direction; the presence of 0.010 per cent to 0.015 per cent oxygen in ordinary high conductivity copper (it occurs within the metal as copper oxide), is necessary to ensure that it will behave properly in manufacturing processes. Another example is provided by the use of carbon monoxide to improve the hardness of mild steel on the surface ('case-hardening'); the gas reacts with the iron, and iron carbide, the hardening constituent, is formed in the outer layers of the part which is to be hardened.

From the scientific point of view the interest is naturally focused on the picture which can be formed of the operative changes occurring within the structure of the metal.

Structure of metallic crystals.

It was long ago shown that metals are crystalline and that the ordinary alloys of industry are composed of a mass of crystals or 'grains,' which are clearly to be seen under the microscope if the piece is polished and suitably etched. It is also well known that the grains of a piece of metal, even though they do not advertise their crystalline symmetry by showing geometrically formed faces, have a beautiful internal regularity which can be revealed by X-ray crystallography or other means. Each crystal of a metal is built up by a symmetrical arrangement of atoms in space; these are packed together in the best way possible to accommodate the forces between them. The structure repeats indefinitely in all directions, and is bounded only by the limits of the piece itself, or, more usually, by the boundaries of the individual grain. The neighbouring grain has, of course, the same structure, but it is differently

oriented in space, and the grain boundary is a zone in which the atomic arrangement is locally not perfectly regular since it is one of transition from one orientation to another.

The geometry of the crystal structure or 'space lattice' is characteristic for each metal, although three common types

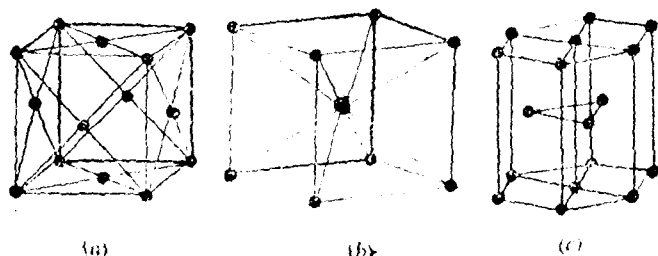


Fig. 25.—(a) Face-centred cubic structure: unit cell. (b) Body-centred cubic structure: unit cell. (c) Close-packed hexagonal structure: unit cell

of symmetry are recognised. Two of these are included in the cubic system, in which the unit of atomic 'brickwork' of which the structure is built is cubic in form. The two common types of cubic 'unit cell' are the face-centred cube and the body-centred cube. In the former the unit cell is a cube, with an atom at each corner and one in the centre of each face; in the latter, a cube with one other atom right in the centre of the cube itself (at the intersection of diagonals joining opposite corners). The third common type of unit cell is that termed 'close-packed hexagonal,' in which the unit is a hexagonal prism, with an atom in the centre of each end hexagonal face, and three others in symmetrical positions on the central plane parallel to the end faces (Fig. 25). Most engineering materials conform to one of these three types of structure, but the actual dimensions of the unit cell are characteristic of the particular metal (or alloy); minor deviations from these types also occur (for example some hexagonal metals are not close packed in structure, but are elongated or compressed axially from the geometrical ideal).

A great amount of research has been done in recent years on the detail of metallic lattices and the forces which hold them together; the dimensions are small, of the order of 10^{-7} to 10^{-8} cm. and the lattice forces large. The periodicity of properties shown in the table of the elements was an early pointer that atoms could not properly be considered as minute billiard balls, and that structures within the atom itself were critical. Metals are characteristically electropositive. This fact involves the implication that their atoms may be readily separated from one or more of the electrons which these contain. From the study of the forces in the space lattices of metals, the general picture has fairly recently emerged that the lattice points are occupied not by atoms, but by positive ions, from which one or more electrons have been separated; the electrons are more or less free to move in the interstices of this structure, forming what may be described as an "electron gas"; the electron gas or cloud is the decisive agency in holding the structure together, the ions themselves having a mutually repulsive effect. This picture is one for the most electropositive metals; other types of lattice bonding apply for those which are less characteristically "metallic" — for example the structures of arsenic, antimony, bismuth, selenium and tellurium are bound together by forces which are more like chemical linkages.

Mechanisms of alloying.

When an alloying addition is made to a metallic element, the question at once arises — how can the atoms of the alloying element be accommodated in the structure of the parent metal? Great numbers of investigations have been made over many years to follow the changes that occur when alloys are made; years ago the microscope and the cooling curve were the principal research tools and a great body of knowledge was built up with their aid. More recently X-ray and other techniques of modern physics have been adding to the data and are being also used to help formulate theories as to the

mechanisms of the observed effects. For many combinations of metals and over large ranges of temperature it has been found that the alloy structure scarcely differs from that of the parent metal, and that under the microscope only one type of crystal or grain is visible. Here the alloy is said to be a "solid solution", since the alloying addition must have somehow gone into solution in the parent metal. In other combinations of two metals, one may see under the microscope a mechanical mixture of two species of crystal or grain; these substances are only very seldom the pure constituent metals, they are themselves usually solid solutions of the two metals, often being of the kind - solid solution of B in A, mixed with solid solution of A in B. These various species of grains are called "phases" and they are usually assigned Greek letters to identify them; for example if one talks of an alpha-beta brass, one means a copper-zinc alloy of such a composition (for example, 60 per cent copper, 40 per cent zinc) which, it is found, has a structure consisting of the alpha phase together with the beta phase of that particular system. Phases usually exist over a range of composition; some may partake of the nature of intermetallic compounds; if they do, then their range of composition is, as would be expected, very limited.

If an alloying addition enters into solid solution its atoms must find places for themselves within the framework of the lattice of the parent metal. There are two possibilities - that the atoms of the alloying element should occupy lattice sites of the parent metal, or that the atoms of the alloying element should occupy spaces in the interstices of the parent lattice. These two types are called 'substitutional' and 'interstitial' solutions. The former is by far the more common; interstitial solutions are generally formed only by small atoms which can most readily be accommodated in the small spaces available. Carbon, nitrogen and hydrogen are the most common interstitial-solution forming elements, and that technically most important substance, the solid solution of carbon in iron, is interstitial: the carbon, which when

MINOR ADDITIONS IN METALS AND ALLOYS 141

it does so, comes out of the solution as iron carbide, Fe_3C , reposes within the iron lattice, when dissolved, at points equidistant from three neighbouring iron atoms.

The effect of foreign atoms in the parent lattice is naturally to cause a greater or lesser distortion of its regularity in the neighbourhood of each intruder. The distortion varies with the nature of the added atoms, and with their numbers; when sufficient foreign atoms have been added to make the solid solution just unstable, then a new phase is formed if further additions are made, and this phase is of such a structure that it can accommodate the interatomic forces more 'economically' than the one which has just become unstable. The point at which a new phase appears is called a solubility limit; one may say for example that at room temperature aluminium can hold up to 0.2 per cent of copper in solution, and that above this solubility limit a new phase appears if more copper is added.

Important though the effects of added elements which go into solid solution may be, the influence of the presence of separate phases is equally great. Separate phases can occur in a variety of forms ranging from needle-like to globular shapes; they may occur within the grains of the parent metal or in the grain boundary, they may be in massive or in film-like forms, and their mechanical state of aggregation may have an influence as great as or greater than their 'inherent' effect associated with composition. In no field of metallic structures is this better exemplified than in 'age-hardening' in which the whole beneficial effect of an alloying addition is contingent upon its being present in the appropriate form.

Age-hardening.

Age-hardening is the name given to the spontaneous hardening of an alloy on standing after heat-treatment; the metal hardens with the passage of time, like its makers' arteries. This effect, so well-known in the duralumin type of alloy, is basically due to the fact that in the alloys showing it there is a solid solubility limit which increases as the temperature

is raised. At 548° C for example, metal A may be able to take 5.7 per cent of metal B into solid solution; but at room temperature it may hold only 0.2 per cent in solution, and above this limit some separate phase – say the β phase – tends to come out of solution. Suppose now an alloy of 5.7 per cent B in A is heated to 548° C and held at that temperature for long enough for all the β phase to dissolve and to produce a homogeneous solid solution; if the alloy is then quenched in water, the piece may cool so rapidly that there is no time for the β phase to come out of solution as it would otherwise do. The condition is then that of an alloy containing 5.7 per cent, and not 0.2 per cent, of B in solution at room temperature; the extra 5.5 per cent of B is said to be held in 'forced' solid solution. As a result, the alloy is unstable and spontaneous changes begin to take place in the direction of precipitation; these changes cannot occur very readily, since at room temperatures what may be called atomic mobility is not great, and not only is the time-scale of precipitation slowed down, but the state of aggregation of the embryonic precipitate is different from that associated with stability.

There has been much research, which is still continuing, on the nature of the incipient and rudimentary precipitates which form under these conditions in different alloy systems; for example in the aluminium-copper system, for which the above figures apply (this is the basis of the duralumin type of alloy) the changes taking place are complex and are on a scale well beyond the ordinary light microscope to resolve, although the normal precipitate in the un-heat-treated cast alloy is readily visible under the microscope. The effect of the elementary type of precipitation occurring under age-hardening conditions is observed in many alloying systems; it is to cause a distinct, and sometimes a technically decisive, hardening and strengthening of the alloy, and electrical and other physical properties are also affected. These changes are wholly due to the special state of aggregation of the alloying addition; the alloy before

MINOR ADDITIONS IN METALS AND ALLOYS 143

the solution and ageing treatment is of the same chemical composition, but the second phase, when in massive form (massive at least on a microscopic scale) has a relatively small strengthening effect on the structure. The highly dispersed incipient precipitate exercises its great hardening effect by some mechanism of distortion of the parent lattice, which makes the metal more resistant to deformation (that is, harder) by making less easy that slipping movement on atomic planes by which a metal deforms under load. The details of the lattice distortion mechanisms vary from alloy to alloy according to the nature of the component metals.

Effects of minor additions.

The technical interest of minor additions to metals and alloys lies in four main fields; first, the addition may have some peculiarity which makes it so potent, that although minor in quantity it constitutes the main alloying effect of the system; second, it may be added to neutralise the harmful effect of some impurity which it is not practicable or economic to eliminate; third, it may be added to counteract some undesirable behaviour of an otherwise useful alloying addition already present, or finally, it may be added to modify some specific physical or chemical property of the metal to a maximum (or a minimum) extent. These effects the minor addition must achieve by going into solid solution in phases already present, or by causing a separate phase to appear.

Light alloy examples.

A good example in the first field is provided by the recently investigated alloys of magnesium and zirconium, some of which are now being developed for industrial application. Pure magnesium is not used for engineering purposes, since, unalloyed, it is too weak. One of the reasons for this weakness appears to be connected with the fact that it solidifies with a large grain size; it is not only a general fact of observation that metals and alloys tend to be weak if their grain structure is coarse, but it is found that magnesium alloys as

a class are especially badly affected if their structures are coarsened by any means. Research during the last few years has shown that zirconium has a large grain-refining influence on the structure of cast metal and if about 0.7 per cent of the alloying addition is present in an effective state, the grain size is reduced by a factor of ten or more. This is illustrated by plates 21 and 22; plate 21 shows the grain structure of pure cast magnesium as seen under the microscope at a magnification of 50; plate 22 shows the grain structure of a cast magnesium alloy containing 0.7 per cent of zirconium, at a similar magnification. The latter has mechanical properties which are very much improved, and its ability to deform before breaking is at least doubled. It is probable that the alloy owes its improved properties to some 'inherent' effect of the zirconium in solid solution as well as to the more mechanical effect of the reduction in the size of the grains. Not more than about 0.75 per cent will go into solid solution in magnesium; this is remarkable since from many points of view theory would suggest that a wide composition range should exist for the solid solution. It almost seems as if the zirconium, although not dissolving to such an extent as might be expected, crammed all its probable effect into the restricted solution range of about 0.75 per cent. A similar amount of zirconium enters into solid solution in alloys of magnesium containing 2 per cent to 6 per cent zinc, and also has a striking effect in refining the grain, strengthening the alloys and making them much more capable of being deformed before breaking; the latter is important technically as it means that forms such as sheet and forgings can much more readily be made from the ingot. Its useful effects increase more or less uniformly up to the solubility limit of zirconium in the parent material; further additions give no increased benefit, and in fact they are probably harmful, as the extra alloying element causes a massive new phase to appear which is likely to have a bad effect on the corrosion behaviour by favouring the creation of local galvanic cells within the metal.

The alloys of aluminium have been the subject of intensive research for very many years, and strong grain refinement has been observed to follow from minor additions of titanium, columbium and boron; but in this field the minor addition which might be classed as having the effect of a major alloying element is sodium, when added to the aluminium-silicon alloys. The normal alloy containing 11 per cent to 13 per cent silicon is distinctly weak and brittle under impact if it is cast in a sand mould to make a piece of any size. If however, 0.05 per cent of sodium is added to the molten alloy just before pouring, the metallographic structure is greatly refined and the mechanical properties of the casting are so much improved that the treated (or 'modified') alloy may safely be used for engineering parts such as crank-cases and cylinder blocks. Some of the sodium is lost in the addition, and only about 0.015 per cent appears in the finished casting. There are many interesting points about 'modification', chief among which is the fact that the alloy shows a modified structure even in the absence of sodium if it is quickly cooled; in other words, a thin sectioned piece cast into a metal mould would not require the addition of sodium to produce a modified micro-structure and good mechanical properties. The sodium, therefore, appears to be acting in some way like a super-cooling agent; it also delays the crystallisation of the silicon, since it is found that modified alloys solidify at lower temperatures than do the normal ones. If an alloy supercools it is usually for the same reason that the super-saturated hypo solution in the schoolboy experiment remains liquid - because of absence of inoculants. In the solidification of metals only specific inoculants are effective and it is probable that sodium has its remarkable effect by forming an 'insulating' coating on minute particles of impurities which, in the unmodified alloy, can act as inoculants and cause earlier and coarser solidification.

Counteractant additions.

An example of researches in the second field is provided

by the harmful effect of small amounts of bismuth as an impurity in copper. If copper contains more than about 0.005 per cent of bismuth its ability to be hot-rolled falls very seriously, and as hot rolling is the standard way of making copper sheet, this influence is economically most undesirable. Copper refined by an electrolytic process contains negligible quantities of bismuth, but it is always present in copper which has not been treated in this way, and no other satisfactory way of removing small quantities of bismuth is available. The amount of bismuth entering solid solution in copper is less than 0.002 per cent; any bismuth occurring above this limit in pure copper does so as metallic bismuth itself. Its strikingly harmful effect is exerted because its presence gives rise to small quantities of liquid metal when the copper is heated for hot-rolling; the liquid spoils the continuity of the metal and causes cracking; it also adversely affects cold-rolling because the bismuth, which is weak, tends to take up film-like forms between the crystals of copper, destroying their cohesion. If the copper contains oxygen, as it often does, then the bismuth occurs as the less harmful oxide, Bi_2O_3 ; above 800°C , however, metallic bismuth is formed. The approach to the problem has had to be along the lines of finding some addition agent which will neutralise the harmful effect of the bismuth, by combining with it to convert it into some relatively harmless form or compound.

Australian research showed that one agent which could be used to eliminate, at least at room temperature, the brittleness of bismuth-contaminated copper, was phosphorus. The use of this element is an easy matter for the copper refineries, since copper is often deoxidised with phosphorus, and all that is required is to ensure that there is enough phosphorus left in the metal to neutralise the bismuth; it seemed probable that the phosphorus worked without combining with the bismuth, but by causing it to assume a globular, non-film-like form, perhaps by some surface-tension mechanism. This answer was not perfect, however, since the amount of

phosphorus required (0.1 per cent to neutralise 0.011 per cent of bismuth) would itself interfere with the hot-working behaviour. Another line of research has been pursued in this country and a solution has been forthcoming which is effective for high as well as room temperatures. The attack was to test the effect of alloying additions which form stable compounds with bismuth; among these was lithium, which forms the compound BiLi , with a melting point of 1145°C . It was found that an addition of 0.036 per cent lithium entirely eliminated the harmful effects of bismuth up to 0.01 per cent. The formation of the high melting-point compound would naturally lead to the elimination of the liquid phase so harmful to the cohesion at high temperatures. The addition of lithium is also effective in preventing the bismuth embrittlement of copper-tin and copper-zinc alloys.

Another example of the 'set a thief to catch a thief' technique occurred in zinc alloy research work. Some years back a serious trouble with zinc alloy die-castings was that they were found to weaken spontaneously with time; they became brittle, and even sometimes disintegrated after a few months service. It was found that the trouble was due to the presence of minute amounts of tin, lead and cadmium in the alloy, which caused it to become susceptible at the grain boundaries to the effects of moisture and slight heat, and thus prone to atmospheric corrosion spreading along them, which caused weakening and disruption of the metal. The amount of tin which could have this unwelcome effect was particularly small – of the order of 0.002 per cent. As lead and cadmium were present in the crude metal, the problem of dealing with them was in the last resort a matter for experts in extraction of the zinc; in the meantime a temporary solution was found, however, by the addition of magnesium to the casting alloys. It was found that about 0.1 per cent of magnesium would make the alloys fairly stable to tropical atmospheric conditions provided the concentration of lead, tin and cadmium were not too high (0.005 per cent for cadmium and tin, and 0.010 per cent for

lead). At the same time the general corrosion resistance of the zinc alloys for casting was raised by eliminating the 3 per cent copper which they used to carry, and thus making them mainly a zinc-aluminium alloy. The exact mechanism of the improvement caused by the magnesium addition is not clear, but it is probable that it operated by the formation of harmless magnesium-containing compounds in which were 'mopped up' the dangerous impurities. The whole problem has now passed into limbo by the development of a new process for the extraction of zinc, which gives metal of higher purity than 99.99 per cent; although the woodpile should no longer contain the niggers, 0.03 per cent of magnesium is still added to the alloys in case of accidental contamination.

Preventing "weld decay".

Although it is possible to make a stainless steel without chromium there are many excellent reasons why our standard stainless steels should contain that element, and one of the best-known compositions for stainless steel is 18 per cent chromium, 8 per cent nickel, together with a small amount of carbon - say 0.1 per cent. In steel of this composition the chromium confers the bulk of the 'stainlessness', and if anything happened to withdraw the chromium, even locally, the steel would not be able to resist corrosion in the way expected of it. It turns out that this very thing can occur under some conditions. The structure of stainless steel is such that under most conditions the carbon present remains in solid solution; if however, the steel is heated in the range of 550° C - 750° C most of the carbon will come out of solution in the form of free carbide which appears at the grain boundaries. The carbide which forms in this steel is mainly chromium carbide (Cr_3C_2), although some iron and nickel carbides also form, and every 0.1 per cent of carbon in it is combined with 15 to 18 times that amount of chromium. When the steel is heated in the critical temperature range the carbon, having a small interstitial atom, diffuses quickly,

and comes from the interior of the crystals to reach the grain boundaries, at which precipitation occurs; the chromium atom, at these temperatures, diffuses to the 'rendezvous' far more slowly, and consequently the chromium making up the precipitate is drawn mainly from the outer portions of the crystals, near the grain boundaries. Since these outer layers must contribute enough chromium to supply 15 to 18 times the amount of carbon coming from the whole of the bulk of the crystals, they suffer a severe loss of chromium, and as a result the corrosion resistance of the metal is locally destroyed. Stainless steel which is in this condition, if subjected to corrosive conditions, suffers rapid attack at the grain boundaries, and in time may disintegrate. This effect was originally noticed in stainless steels which had been welded, and it was termed 'weld decay' in the shops, 'intergranular corrosion' in the laboratories.

This bad effect of an otherwise useful alloying element could be countered by three methods: by reducing the carbon content so that precipitation does not occur, by introducing some other element which has a greater affinity for carbon than has chromium, or by suitable heat-treatments. The first is not practicable from a manufacturing viewpoint, and the last is often not economic or convenient, especially for welded pieces. The search for the antidote element was, however, successful, and three answers were found. The most commonly used alloying addition to prevent intergranular corrosion of stainless steel is titanium, which has the required high affinity for carbon and forms a carbide TiC . It is found that if 0.5 per cent of titanium is present, essentially all the carbon which precipitates does so as TiC , and the chromium is left behind in solid solution to protect the steel against corrosive attack, which thus becomes immune from weld decay. In welding, however, the presence of titanium in the welding rod does not necessarily produce weld metal which is immune from intergranular corrosion, since a large part of the titanium may be oxidised during the welding, and the loss will be critical if several

crossing or overlapping welds have to be made. In these circumstances the alternative solutions may provide better results – 0.75 per cent of columbium or 2 per cent silicon – since these elements do not suffer such loss in the welding operation, and consequently can more readily ensure that the chromium is kept in its proper 'defensive' place in solid solution.

Restriction of grain growth.

Examples of the use of minor additions to modify specific physical or chemical properties of metals and alloys are numerous. Interesting illustrations lie in a field allied to that of grain refinement touched upon earlier; this is the restriction of grain growth. It is a general rule that metals which have been mechanically worked (rolled, forged, etc.) may tend to undergo grain growth when they are heated, i.e., certain grains grow by swallowing their neighbours. It often turns out that severe grain growth is undesirable since it is accompanied by loss of strength and hardness. If it is an unavoidable part of some operation that a metal or alloy has to be submitted to a high temperature, then it becomes desirable to try to restrict the resultant grain growth in some way, to avoid its weakening effects. A familiar example of metal exposed to very high temperature is that of the tungsten filament wire in ordinary electric lamps which runs at temperatures well over 2,000°C. If tungsten without any additions is used for filament wire, the grains become enormous as the lamp's life continues; as a result the metal becomes brittle and the grain growth causes distortion or sagging under the filament's own weight. An addition is therefore made to the tungsten during the manufacture of the wire, of an agent capable of obstructing the growth of the grains; 0.75 per cent of thorium oxide (thoria) is very effective in this way, being insoluble in the tungsten, and when finely distributed throughout the structure the particles present a durable mechanical obstruction to grain growth. Tungsten containing 0.75 per cent thoria thus

permits the production of filament wire with a much increased service life.

Before a worked metal undergoes grain growth when it is heated, it usually recrystallises – that is, among the crystals which were distorted by the working process, appear tiny new ones, which are not distorted; these begin to grow as the temperature is raised further, and they finally obliterate the old structure. The metal is then said to be 'completely recrystallised'. Most metals are much strengthened by being worked, and recrystallisation causes the increased strength to be lost with the obliteration of the old crystals. Pure copper is rather weak, but is valuable for its high electrical and heat conductivity; for some uses in which strength is needed as well as good conductivity, it is used in the worked state, such for example as hard cold-rolled sheet; suppose, however, that this copper has to be soldered; the heat of the soldering operation will perhaps be enough to cause recrystallisation and loss of the desired extra strength put in by the working. Here the problem is to find an alloying addition which will raise the recrystallisation temperature, thereby avoiding the softening, but which will not interfere with the good conductivity of the metal. Copper is most sensitive to minute traces of impurities, many of which cause a great drop in conductivity; the selection of permissible alloying additions is, therefore, very restricted. The answer to the problem is provided by the addition of about 0.1 per cent of silver. This element itself has a very high electrical and thermal conductivity, and has a negligible adverse effect on these properties in copper; it also has no bad effect on the strength of the worked copper, but 'postpones' to much higher temperatures the softening effects of heating for operations like soldering. This small addition of silver thus permits the use of copper when strength and high conductivity are required, and when a high temperature is to be met during manufacture.

This short and rather simplified review will perhaps help

to bring home to non-specialists the idea that the metal laboratories are to-day very much concerned with 'minor-addition-metallurgy'.

Beginners in this field may find it helpful to read first the article *Metals* by Sir Lawrence Bragg, which appeared in *Science News* 1. —Ed.]

GLOSSARY

BIRTH RATE: This is calculated as

$$\frac{\text{the number of babies born during one year} \times 1000}{\text{the number of individuals alive, half-way through}}$$

Thus, if the year is from January 1st to December 31st, the total number of people of all ages alive on June 30th, is the figure required for the denominator (see also under *Fertility*).

CASEIN: The chief protein present in milk.

COEFFICIENT OF EXPANSION: Different materials expand and contract to differing extents on heating and cooling. These differences between substances can be described quantitatively by comparing the actual expansions of unit lengths (1 cm) of each material when the temperature is raised by 1°C. The numbers so obtained are called coefficients.

DEATH RATE: The crude death rate for a population is

$$\frac{\text{the number of deaths registered in the year} \times 1000}{\text{the total number of people alive, half-way through}}$$

Here the word 'population' denotes a group living within definite geographical boundaries. Consequently, to get the correct number of deaths, allowance must be made for residents who die while away from home, and the deaths of visitors from abroad must be discounted.

Specific death rates in general are calculated in the same way, but refer only to one group of the population, e.g., teen-agers, bus-conductors, women, comparing the number of their deaths with the number of them all alive.

FERTILITY: An estimate of the extent to which people can and do multiply. Thus the Birth Rate (see above) is an attempt to measure it, but one liable to mislead since no account is taken of the number of women in the population of the right age to have children.

Therefore the Fertility Rate is sometimes calculated:

$$\frac{\text{number of legitimate births} \times 1000}{\text{number of married women between 15-45 years old}}$$

in which case an illegitimate birth rate must also be separately calculated on the same lines, to get a true final picture.

LIFE TABLES: How long can you expect to live? Quite a number of people besides yourself are interested in this question, Insurance companies for instance, and Government planners trying to arrange the future. If you are already ninety when you read this,

you know that the time is not long – much shorter than if you are only ten. A life table shows the probable expectation of life of a person at any age, calculated continually from the observed ages at which people die. Owing to the progress of public health measures and of medicine and surgery, and perhaps for other unknown causes, the ages at which most people die continually alter (people on the average live longer now than they used to a hundred years ago), so that the life table must be always brought up to date.

MAGNETIC SUSCEPTIBILITY: A coefficient which indicates how easy it is to magnetise a given substance. Every schoolboy knows that iron is easily made into a magnet (i.e., has a high susceptibility). Whereas aluminium, for instance, cannot be so treated, and therefore has a very low value.

About Our Contributors

A. E. Bell, Head of the Science Department, Clifton College, is a chemist who has become equally interested in the history of Science. His book on *Christian Huygens* appeared earlier in the year, and his chief interest is in the educational problem arising through the need for scientists of wide culture in modern civilisation.

I. W. Cornwall was born in India in 1909. He is a graduate of Cambridge and holds the Academic Diploma in Prehistoric Archaeology of the University of London, where he occupies the post of Secretary at the Institute of Archaeology. His chief interests are in human evolution, the skeletal remains of early men and sub-men, and the study of their environment and equipment as shown by Geology, Palaeontology and Archaeology.

J. L. Crammer has been editor of *Science News* from the first issue, originally under a pseudonym. He has had experience of Service life and of journalism, is a Science graduate of Cambridge, where he has done research, is medically qualified, and has practised medicine. He is now twenty-seven.

Frank Dickinson is a contact lens practitioner who has written several technical and general articles on the work which he has been doing for the past ten years. He has lectured extensively abroad and at training centres and technical colleges here, and is a member of the Council of the Contact Lens Society.

Francis Fox graduated in Metallurgy at Birmingham, and took his first job on a gold mine in West Africa. He has been engaged in metallurgical research ever since he returned to this country, and is best known for his publications on magnesium alloys, for which he was awarded his D.Sc. He is at present Deputy Director of the British Welding Research Association.

A. W. Haslett is a Cambridge graduate, who has made scientific journalism his career. He is Editor of *Science Today*, and for some months past has been a regular speaker in the B.B.C. Forces Educational Programme.

Robert René Kuczynski, who died at the end of last year, was one of the greatest demographers of our time. The University of London appointed him in 1939 to the first Readership in Demography established in this country, and in 1944 he was appointed demographic adviser to the Colonial Office. His publications include *Balance of Births and Deaths* (1928 and 1931), *Fertility and Reproduction* (1932),

- Magnetic method of prospecting, 66
 Mantle of basalt, the earth's, 74
 Marriage statistics, 55, 56
 Mechanism of alloying, 139
 Mesolithic man, 13
 Mesons, 125
 Meteorites, 83
 Mortality statistics, 53, 54
 Muscle measurements, 113
 National registration census, 58
 Nullarbor Plain, 128
 Omissions in censuses, 47
 Operational research, 127
 Palæolithic man, 13
 Penguin antics, 123
 Pernicious anæmia, 133
 Petrology, 22
 Phosphate-determinations in Archaeology, 14
 Pollen analysis, 22, 27
 Polymerisation, 99
 Popularity of different numbers, 47, 49, 51
 Post-operative care in grafting, 36
 Preparation for corneal grafting, 33
 Preservation of prehistoric material, 12
 Pressures in earth's crust, 78
 Psychotherapy, 101
 Radar-controlled air survey, 118
 Radio and sunspots, 120
 Recent mountains, 79
 Recording of field work in archaeology, 16
 Reflection shooting, 72
 Refraction shooting, 49, 71
 Reproduction rates, 58, 59
 Resistivity surveying in archaeology, 14
 Rock classification, 64
 Royal Commission on population, 57
 Royal Society foundation, 42
 Sedimentary layer of earth's crust, 73
 Seismic method of prospecting, 67
 SIMA, 81
 SIAL, 81
 Soil science, 22
 Standardized death rates, 54
 Structure of metallic crystals, 137
 Study of types in archaeology, 17
 Temperatures within the earth, 75
 Thixotropy, 91
 Torsion balance, 67
 Trading ramifications, 23
 Transference in psychotherapy, 108
 Varved clays, 27
 Viscosity, 91
 Wegener theory of continental drift, 76
 Weld decay prevention, 143
 Whale muscle, 130

GENETICS

BY H. KALMUS

Though genetics is the youngest of the biological sciences, it is exciting more attention among both scientists and layfolk at the present time than any other branch of biological inquiry.

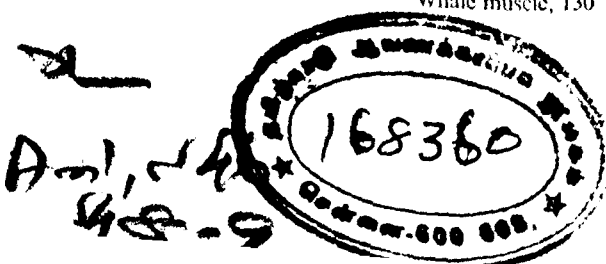
It has grown so fast, and in so spectacular a manner, that the majority of scientists, and even of biologists, cannot keep abreast of it; in consequence it has won an undeserved reputation for being difficult to understand. How undeserved, will be apparent to every reader of this book; for Dr Kalmus and his collaborator, Miss L. Crump, in comparatively small space, here set out in simple language the main principles of the science so clearly and concisely that every reader, however previously ignorant of its axioms and achievements, will be able to follow.

Part of the supposed difficulty, perhaps, resides in the special terminology which genetics has created to meet its special needs. But no one need be deterred from making a closer acquaintance with genetical principles on this account; the authors give the fullest explanations of the special terminology, and clones, alleles, polyploids will hold no mysteries for the readers of their pages.

What factors of biological make-up are inherited, in plants, animals, and man; how they are inherited; how mutations and variations arise and are transmitted; the importance of genetical knowledge to the gardener, farmer, stock-breeder, and human parent; all these matters are discussed and explained, and the text is elucidated by a number of simple diagrams.

A PELICAN BOOK - A179

one shilling and sixpence



BY H. STAFFORD HATFIELD

Dr Stafford Hatfield surveys the whole field of invention, which, as everyone now realises, we must cultivate intensely or perish. We cannot earn our living, as we conceive living, by mass production of standard articles. We must earn the wages of high skill, turning a shilling's worth of raw material into a pound's worth of goods unobtainable elsewhere. All experience shows that such goods are created by individual inventors, who have found this country very unresponsive in recent times. This book is addressed to such men, to explain themselves and to the public at large, more particularly the manufacturers who find them difficult, and the Government who have failed to create a patent law which ensures them a due reward under modern conditions. Nationalisation of industry creates a new and hopeful situation for the inventor, since pioneer invention, so often a dead loss for one firm after another till technical success is attained, is now a business which, if properly pursued will benefit the industry as a whole.

A PELICAN BOOK - A178

one shilling and sixpence

1

[illegible]

I4

Synthetic Fibres

Recent Progress in Television -

Molecular Architecture in Plants

The Electrical Activity of the Human Brain

Viruses and the Veterinary Surgeon

The Origin of the Solar System

Correspondence

The Calendar

The Discovery of the Bacteriophage - F. W. Twort

The Bacteriophage - Felix d'Herelle

The Nature of Bacteriophage - R. W. G. Wyckoff



SCIENCE NEWS

EDITED BY J. L. CRAMMER.

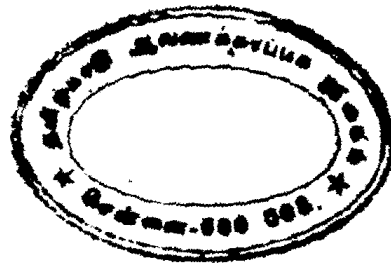
14

PENGUIN BOOKS

HARMONDSWORTH • MIDDLESEX

First published December 1949

Am 1, 246
249-14



Made and printed in Great Britain
for Penguin Books Limited by C. Nicholls & Co. Ltd,
Manchester

CONTENTS

Editorial	7
Recent Progress in Television by R. Howard Cricks	9
The Discovery of the Bacteriophage by F. W. Twort, F.R.S.	33
The Bacteriophage by Dr Felix d'Herelle	44
The Nature of Bacteriophage by Ralph W. G. Wyckoff	60
Viruses and the Veterinary Surgeon by G. P. West	69
The Origin of the Solar System by Sir Harold Spencer Jones, F.R.S.	74
Molecular Architecture in Plants by R. D. Preston	94
Correspondence	116
The Electrical Activity of the Human Brain by Dr William A. Cobb	119
Synthetic Fibres by R. W. Moncrieff	141
The Calendar by W. A. Price	153
Books Received	156
About our Contributors	157
Glossary	159
Index	160

How An Aeroplane Flies, Plates 1-9

Editorial

THE nucleus of this issue is a short symposium on the Bacteriophage, that parasite of bacteria which so well illustrates the old rhyme about little fleas having lesser fleas. The discovery, first by Professor Twort and then by Dr d'Herelle, of a submicroscopic germ which causes a disease amongst some of the bacteria responsible for human illnesses, raised great hopes that bacteriophages would prove of medical value in controlling these bacteria and curing these illnesses. These hopes have not been fulfilled, at least in the opinion of British and American doctors, and interest in the bacteriophage has waned.

With the coming of the electron microscope, however, there has recently been aroused a fresh curiosity, largely in the mind of Dr Wyckoff, who describes here the new work on the subject, illustrated by some fine plates (*Nos. 10 to 15*). Although this is at present only of general scientific interest, it may yet prove ultimately to lead to knowledge of practical healing value, and Dr d'Herelle's hopes and claims thereby gain a firmer foundation than they at present possess. In any case, our symposium is a contribution to history, since two eminent men describe how they came to make certain discoveries, and an indication of recent developments which may stimulate others to become interested in the subject.

It is evidence of the way that science evolves by the breakdown of artificial frontiers that Dr Wyckoff, although now working on a matter of biological and medical importance, is primarily a crystallographer, as is Dr Preston who works and writes now within what are usually regarded as the confines of botany. Dr Cobb, on the other hand, began as a practising medical man, and in his study of

'brain waves' has had to become more and more of a radio engineer. Even Mr Cricks, a cinematographer, is taking a close interest in television which technically is a totally different sort of moving picture.

The Astronomer-Royal needs no introduction. His account of the possible origin of the earth and its fellow planets is an interesting corrective to earlier popular accounts, showing that the subject is both less settled and more complicated than other writers have suggested.

Mr Moncrieff and Mr West have both written for *Science News* before; the veterinary notes link up with the bacteriophage symposium in a further reminder of the importance of viruses, while the article on *Synthetic Fibres* follows out a thought in *The Molecular Architecture of Plants*.

In the next issue of Science News there will be as a supplement a complete index to all the issues from 1 to 15.

Recent Progress in Television

R. HOWARD CRICKS

IT is now seventeen years since the B.B.C. incurred a certain amount of scorn by broadcasting television on John Logie Baird's original 30-line system. The signals were broadcast over the normal sound channels like ordinary radio, and showed in the receiver a small red-coloured image, which with some imagination one was able to recognise as a dancer or an acrobat.

Four years later in 1936, after competitive tests between high-definition systems sponsored by Baird and by Electrical & Musical Industries (E.M.I.) the latter's system was adopted. The B.B.C. could for some years claim to be the only source of high-definition television; much of the original equipment is still in use.

Some of us can still recall the amazement with which we realised that, in contrast to the 10 Kc/s* which represented the limiting factor of Baird's early transmissions, we were witnessing the results of a 5 Mc/s transmission – five million separate impulses every second which went to make up the high-definition picture.

High-frequency Transmission

Let us consider a little more closely what this statement means. Whether we are transmitting sound or picture, the method of transmission universally employed in Europe is known as *amplitude modulation* (a different system known

* c/s = cycles or impulses per second; Kc/s = kilocycles (thousands of cycles) per second; Mc/s = megacycles (millions of cycles) per second. Frequency f is related to wave-length λ (in metres) by the expression:

$$f \times \lambda = 3 \times 10^8$$

as *frequency modulation* is being experimentally employed in America). This means that a (carrier) wave of a fixed high frequency is generated, and its strength is

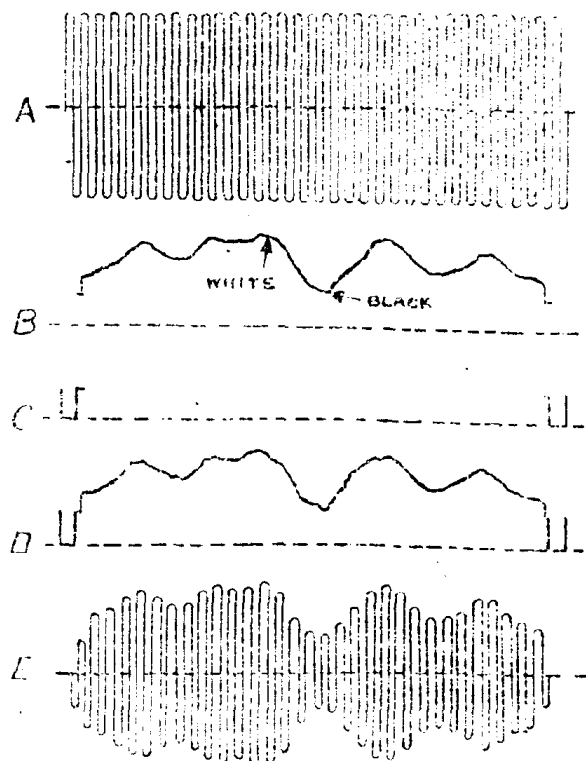


Fig. 1. - The principle of modulation of a Radio Wave. *A* the unmodulated carrier wave. *B* the signal received from the camera. *C* synchronising pulses. *D* combined picture and synchronising signal. *E* carrier wave modulated by the signal of *D*. (The single scanning line, i.e. what happens between the two synchronising pulses indicated, occupies less than $1/10,000$ th of a second, and in this period there are actually over 4,000 cycles of the carrier wave.) (Reproduced by courtesy of the *Ideal Kinema*.)

varied or modulated from instant to instant by the impressed impulses to correspond with the signal being transmitted.

The application of this principle to television transmission is illustrated in Fig. 1. *A* is the high frequency carrier wave, *B* is the actual wave-form of the picture, varying from light to dark, and *C* the synchronising pulses (the purpose of which will be referred to later); *D* is the combined signal wave-form. For transmission, the carrier wave is made to vary in amplitude or strength in accordance with the signal, as shown in *E*, which represents the waves actually emitted by the aerial.

Now, it is obvious that the higher the signal frequency (or speed of change in *B*) that it is required to transmit, or as it is expressed, the greater the band-width, the higher must the carrier frequency be, if the signal frequencies are not to undergo unacceptable distortion. While sound frequencies, ranging up to 10,000 c/s, can be satisfactorily transmitted on comparatively low frequencies (the B.B.C. Light Programme is radiated on a wave-length of 1500 metres or a frequency of 200 Kc/s), the fact that television calls for the transmission of 5,000,000 picture impulses per second requires that it shall be transmitted on correspondingly higher carrier frequencies.

This figure of 5,000,000 picture impulses corresponds to 2.5 Mc/s. The programmes from Alexandra Palace are broadcast on a carrier frequency of 45 Mc/s. Since the effect of the signal frequencies is to give rise to sum and difference frequencies*, the actual wave-band transmitted extends from 42.5 to 47.5 Mc/s. Television sound is radiated on a frequency of 41.5 Mc/s; since the band-width is

* When a wave such as a radio wave is modulated, new frequencies are produced equal to the sum and difference of the carrier frequency and the modulation frequency. Thus if a carrier frequency of F is modulated by a single frequency f , two new frequencies of $F+f$ and $F-f$ will be produced. The waveband necessary is equal to twice the maximum frequency to be transmitted.

negligible, there is practically 1 Mc/s separation between the two transmissions.

Something even more important which necessitates the use of these high frequencies, is the fact that at the lower frequencies used for sound broadcasting the requisite wave-bands just do not exist. Thus the whole of the B.B.C.'s medium wave-band, from 600 to 1500 Kc/s, is less than 1 Mc/s in width.

Unfortunately the necessity of transmitting on the high-frequency short wave-lengths means that the radius of reception is limited, since such waves have what is termed a quasi-optical range – they travel substantially in straight lines. Experience has proved that the frequency adopted at the inception of the Alexandra Palace services represents the most satisfactory compromise between the necessary frequency band and the range of reception.

Repeater Stations

This limited range of reception is the reason for the erection of a number of secondary transmitting stations, the first of which is at Birmingham. In order to provide inter-communication between London and Birmingham, four unattended repeater stations provide radio links between masts erected on the roof of the Museum telephone exchange in West-Central London and similar masts in Birmingham. The equipment for these stations I had an opportunity of examining before it was installed, at the G.E.C. Research Laboratories at Wembley.

At the top of 200-ft. masts are two pairs of paraboloid reflectors of tubular construction, one pair for reception of picture and sound respectively, and one pair for the re-transmission; at the foci of the paraboloids are small horizontal dipole aerials. The amplifiers contain complete standby circuits, and also alarm systems whereby in the event of any trouble a signal is automatically flashed to one of the terminal stations. Should the power supply (taken from the grid) fail, a diesel generator will start up automatically,

and within two minutes the service will be on the air again.

The equipment for the ultra-high frequency used – 900 Mc/s – bears no resemblance to orthodox amplifier circuits. At these high frequencies electricity travels along tubes or 'wave-guides', and an amplifier looks rather like a very neat job of plumbing. A resonance circuit consists not of chokes and condensers, but of tubes of carefully computed length. (See also *What Are Wave Guides?* in *Science News* 9, 143.)

Area of Reception

Over what area can viewers expect to receive Birmingham? One hears stories of freak reception of Alexandra Palace – the 1949 Boat Race, for instance, was clearly picked up in Capetown. But it may safely be said that the quality of reception beyond a 50-mile radius makes programmes of little interest except to the radio enthusiast.

Such an enthusiast has described¹ how over a period of two years he regularly picked up the Alexandra Palace programmes in Leicester, a hundred miles away. He found it necessary to convert a standard receiver by fitting a pre-amplifier; noise (the sound-radio term is still used to apply to the 'rice-pudding' effect sometimes seen on the screen) and interference were reduced by restricting the wave-band to 1 Mc/s. He also experimented with various types of aerials. In spite of these steps, he was only able to obtain unreliable results.

Even in the 30-mile range within which reception is normally satisfactory, there are certain snags. As the signal strength weakens, interference becomes more pronounced; flashes of light across the picture mark the passing of every motor-car. Much can be done to improve matters by careful siting of the aerial.

Less easy to overcome is the shadowing effect of hills. Television signals fortunately do not travel in perfectly straight lines from the transmitting aerial, and will jump over projections. A shadow area is, however, commonly found

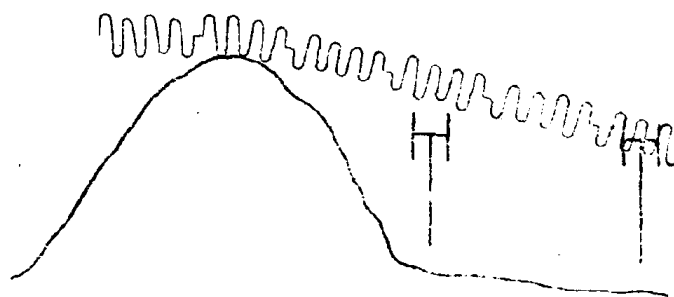


Fig. 2. - Shadowing effect of hills.

immediately beyond hills or prominences (Fig. 2); even tall buildings may have a shadowing effect. It has been stated that at considerable distances from the transmitter freedom from local screening is the most important factor in maintaining reception.

Obviously the range of transmissions can be increased by placing the transmitting aerial higher. A suggestion that must not be hastily dismissed is that an aeroplane should carry the transmitting aerial at sufficient height to assure a wide area of reception. It has been claimed that the use of 'stratovision' would be enormously less costly than building transmitters to cover an equivalent area.

American Transmissions

While we in Britain boast of our forthcoming second transmitter, in the United States the number of transmitters, already 50 or 60, may probably reach three figures. No fewer than 400 wave-lengths have been allocated for television transmission, 315 of them by more than one station. Any town is now limited to seven transmitters.

As against these facts, an experienced observer has stated² that the quality of transmission falls far short of that of the B.B.C., both on technical grounds and in programme material. This fact is not surprising, if one considers that

American television is financed by advertisers, who are naturally chary of spending large sums on programmes for the benefit of only a few thousand viewers.

On the other hand, American receivers can apparently pick up programmes with less interference from motor-cars. This is due to the 'reversed transmission' system employed; in the B.B.C. transmissions (see Fig. 1) white corresponds to full modulation (i.e., maximum strength of signal) and black to 30% modulation (zero being reserved for synchronising pulses), and consequently interference - which is in effect an additional irregular signal - appears as white flashes; in the American system, black is represented by full modulation, so that interference appears as black streaks which are almost unnoticeable.

The last point raises a matter of some concern to television manufacturers. While the adaptability of sound radio sets to use in any part of the world is limited only by their range of reception, there seems little likelihood that television sets can be made for use on the wide variety of standards which already exist.

True, in America sets have been made capable of receiving on as many as 13 different wave-lengths; but in all cases the picture and line frequencies are alike. Television transmissions in different countries vary as regards the number of picture traversals per second (50 in Europe as against 60 in America), the number of lines, and the sense of the transmission; yet another problem has been mentioned - the possibility of the introduction of frequency modulation. It appears that, regardless of wave-length, the possibility of making sets universally interchangeable is remote.

Principles of Television

London viewers have for the last year noticed a distinct improvement in the quality of O.B.'s or outside broadcasts - an improvement which it is to be feared is overdue in studio transmissions. The reason is that a new camera was brought into use, one of the first of its outings being the Olympic

Games in the summer of 1948. In the studio, on the contrary, the cameras of 1936 are at the time of writing still in use, although they are due to be replaced.

Before discussing the principles of the new cameras, let me clear up a widely-held misconception on the principles of television transmission. One frequently hears it said that the television picture is like a half-tone block – it is made up of small dots of varying brightness. As owners of television sets have no doubt already discovered, the picture is made up not of dots, but of lines – lines of light, produced admittedly by a moving point whose brightness varies with the light and shade of the picture that is being transmitted. The principle will be made clear by comparing the drawing (Fig. 3) with a half-tone block.



Fig. 3. – Construction of Television Picture. (Reproduced by courtesy of *Kinematograph Weekly*.)

The scanning lines – the complete framework of which is known as the *raster* – originate in the time-base circuits associated with the transmitting camera; the circuits produce, in addition to the so-called saw-tooth voltages which move the scanning point in the camera from line to line,

synchronising pulses which form part of the transmitted vision signals, and which in the receiver serve to trigger the time-base circuits in order to assure absolute synchronism between transmitter and receiver.

In B.B.C. transmissions there are 405 horizontal lines, with a repetition rate of 50 traversals or 25 complete interlaced frames per second (the term 'interlacing' means of course that instead of the lines being scanned consecutively the odd lines 1, 3, 5 ... are scanned in the first traversal, and the even lines 2, 4, 6 ... in the second traversal).

Let us consider a round-by-round commentary on what happens to the scanning spot during the production of one complete frame. On the sound of the bell the spot enters the screen top left; impelled by the rising saw-tooth current through the line deflector coils of the cathode ray tube controls, it moves towards the right. At the same time, it is being slowly moved downwards owing to the saw-tooth current in the frame deflector coils. As the spot moves across the picture its intensity varies according to the modulations of the radio waves, to form the light and shade of the picture; but, even though the spot may be totally extinguished for a deep shadow, the radio waves, as shown in Fig. 1B, never drop below 30% modulation.

Suddenly the synchronising pulse of Fig. 1C drops the modulation to zero. The current in the line deflector coils drops to zero, and the spot flies back to the left-hand side of the picture; during the time it is moving backwards it is of course blacked out, so that the fly-back is not visible.

While the spot has been covering one line and a fly-back, the frame deflector coils have brought it down the space of exactly two lines, so that after the fly-back it starts again on line 3. Thus it covers the whole picture area until it reaches the last odd line of the picture. At this instant, the frame pulse is received – a multiple repetition of the line pulse. The current in the frame deflector coils drops not to zero, but to such a value that the spot re-starts its second picture traversal on line 2. When it reaches the last even line of the picture,

a complete frame has been traversed. Again the frame pulse effects a fly-back, this time to line 1, when the whole process is repeated.

The Television Camera

In the receiver the screen consists of the fluorescent front surface of the cathode-ray tube. In the original camera – employing the so-called Iconoscope tube – the picture is focused upon an almost microscopic mosaic of photo-cells. Each photo-cell consists of a tiny speck of light-sensitive caesium, deposited electrolytically on a thin dielectric or insulating substance, backed up with a sheet of conducting material. Each photo-sensitive element constitutes with this conductor a minute condenser; the conductor carries a negative charge. The mosaic is scanned by a cathode ray, in exactly the same way as is the fluorescent screen.

As the luminous image falls upon the photo-electric mosaic, each element builds up a positive potential, whose intensity corresponds to the value of light falling upon it. Now, the electrons which constitute the cathode ray are actually particles of negative electricity. Consequently as the beam strikes an individual element of the mosaic, it neutralises its positive charge, and in doing so produces by capacitative action a minute impulse in the rear conductor. As the scanning beam neutralises the charges in the rows of photo-elements, so a variation occurs in the potential of this conductor which corresponds to the light and shade along each scanning line, and which is amplified and fed to the transmitter. The beauty of the system is that each tiny element is building up its charge until the beam strikes it, so that there is a storage effect.

This principle was originated by Dr Zworykin, of the Radio Corporation of America (RCA) and was developed by the technicians of Electrical & Musical Industries in this country. It was embodied in the Emitron camera of 1936. However, it suffers from certain inherent defects.

The worst of these faults is known as *shading*. When the

electrons constituting the cathode ray strike the photo-electric elements of the mosaic, the impact causes secondary emission of electrons. These stray electrons give rise to a spurious image, which is visible as a shading on the edges of the picture. This effect becomes more prominent as the velocity of the electron beam is increased to augment the sensitivity of the tube, and constitutes a limit to its sensitivity.

The Image Orthicon

Through the Image Iconoscope and the Orthicon[®], the present type of tube was evolved, known as the Image Orthicon. This is a really amazing piece of electronic

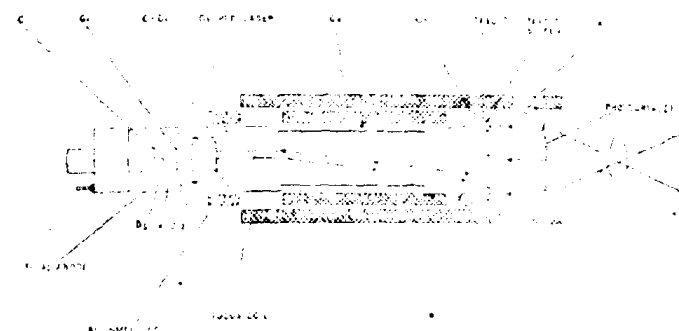


Fig. 4. – Section of Image Orthicon. (Reproduced by courtesy of Marconi's Wireless Telegraph Co.)

engineering, and is illustrated in section in Fig. 4, while Plate 16a shows its appearance.

The optical image is focused upon the photo-cathode, which is no longer a mosaic, but a light-sensitive surface not unlike the photo-cells used in exposure meters, except that it is coated upon exceedingly thin glass. The photo-cathode is maintained at a strong negative potential – 300 to 500 volts – and consequently the target behind it is in comparison positive, and will attract the electrons given off by the photo-cathode.

Owing to the magnetic field of the focusing coil, aided by the electrostatic field created by the accelerator ring G_0 , these electrons travel in straight lines, with the result that on the target is formed an electron image equivalent to the optical image on the cathode; actually the image is doubled by secondary emission*, although degradation of the picture ('shading') by the secondary electrons is prevented by the target screen. The polarity of this electron image is such that a positive charge represents the whites of the scene.

Turning to the other end of the tube, C is a cathode, G_1 a grid, and G_2 an accelerating anode (the latter carrying a 300v. charge) which between them form an electron gun, as in the cathode-ray tube. The resultant beam of electrons is focused by the main focusing coil, and also by the tube G_0 , which carries a positive potential of 160 to 240 volts. By magnetic coils (not shown), again similar to those of a cathode-ray tube, the beam is caused to scan the rear face of the target.

In the highlights of the scene (that is to say, where there is a relatively positive charge on the target) electrons are absorbed, but where the surface is relatively negative they will be repelled. Thus a modulated beam of electrons is reflected off the target back to the face of the dual-purpose element G_2D_1 ; here again secondary electrons are emitted, and these electrons, aided by the delightfully named 'persuader' G_3 , pass to the so-called 'dynodes' D_2, D_3, D_4, D_5 , which are at successively higher potentials, up to 1,200v., and constitute an electron multiplier, amplifying the electron beam by means of repeated secondary emission. The signal is finally received on the anode, which has a potential of 1,250v.

The sensitivity of the tube is such that for the optimum highlight illumination on the photo-cathode of 0.3 foot-candle, the output of the tube is 40 micro-amperes. Pictures have been obtained with the light of a single candle. The signal has a modulation of 25%.

* The emission of more electrons due to the impact of rapidly moving free electrons upon the molecules of a substance.

Camera Design

A further advantage of the new camera tube is that, owing to its smaller photo-cathode (only 1.6 in. across the diagonal) it is possible to use lenses of shorter focal length, from 2 ins. upwards, as compared with the 6 in. and 12 in. lenses of the original Emitron camera. This in its turn means not only that a picture of greater depth of focus is obtainable, but also that – as shown in Plate 16b – the camera can be equipped with a turret of lenses, like a cinematograph camera. In practice, two turrets, each carrying four lenses, are interchangeable upon the camera, one carrying lenses of short focal lengths giving wide-angle long-shot views, and the other lenses of greater focal lengths, simulating close-up views.

While change of lenses can produce the illusion of changing from long-shot to close-up, a better way of doing so is, of course, to track the camera towards the subject. Frequently however – notably in sporting events – this is not possible. An optical means of producing a similar effect employs a variable-focus or 'Zoomar' lens†.

An inherent characteristic of the television process is that it is capable of transmitting only *differences* in illumination – in electrical terms of transmitting A.C., and not the D.C. which would represent a variation in the mean brightness of the picture; although circuits now used tend to introduce a D.C. component, which effects a slight control over the mean brightness. This characteristic has two effects. First, within the range of sensitivity of the camera, the viewer will be comparatively unconscious of the level of the scene being transmitted, i.e. whether it is bright or gloomy; secondly, that ideally the television scene should contain areas of contrasted light and shade – this latter is in fact a point borne in mind by the scene designers at Alexandra Palace‡.

Film Transmissions

Cinematograph film forms an important part of television programmes, whether as news items (the B.B.C. news-reel

has attained a high reputation among Wardour-Street news-reel technicians), the transmitting of complete films as programme items, or the inclusion of short lengths of film to bridge awkward gaps in a studio transmission, to permit a performer to change his costume, or for outdoor 'establishing' shots. Until recently film transmission has never reached the same standard of quality as other studio transmissions.

There is a difficulty in the transmission of film by the ordinary intermittent mechanism of the studio camera or cinema projector. While in the television transmission nearly the whole of the transmitting time is occupied by the picture (only a small proportion of the transmission time being occupied by the synchronising pulses) the intermittent movement of the film occupies a quarter or more of the transmitting cycle (the normal film speed of 24 frames per second is increased to 25 in order to match the scanning cycle).

The problem can be overcome by making use of the storage principle of the Iconoscope. But at Alexandra Palace an optical non-intermittent projector has until now been employed, whereby the film is moved continuously, and a ring of oscillating mirrors maintains a stationary image upon the screen of the transmitting camera, one frame fading into the next. Actually the projector is a mechanism known as the Mechau, made a number of years ago by the German firm of Leitz, for normal cinema projection.

This system embodied all the faults of the normal studio transmission, plus certain additional problems, introduced by the optical system. New film transmitters developed by Cinema-Television Ltd. and E.M.I. make use of a different principle. Here again the film is fed continuously; but no stationary optical image is formed. Instead, the film is scanned by means of the raster formed upon the screen of a cathode-ray tube, in such a way as to compensate for film movement.

In order to grasp the principle, let us first assume that a non-interlaced scanning system were employed, and that the

mask lines – the black bars between the frames on the film – were also transmitted. Then if, as the film moved regularly, a moving point of light traced out a straight line behind it 405 times every frame, we should in effect have scanned that frame in 405 lines. A photo-cell on the farther side of the film could pick up these signals for transmission.

Unfortunately we have these two complications to consider, that each frame has to be scanned twice, and, furthermore, that the film carries mask-lines each one-fifth the height of the actual picture frame (a legacy from the introduction of sound film), and it is obviously undesirable to waste scanning lines on these mask-lines.

Accordingly, the scanning point does not move in a single straight line, but forms a complete raster, so dimensioned that as the film moves a distance equal to half a frame, the whole of the frame is scanned for the first traversal, and as the film moves the second half-frame, it is scanned again for the second traversal. The size of the raster also takes account of the mask line.

This system, it will be seen, eliminates all the faults both of the normal intermittent system, and of the optical non-intermittent system. It produces an undistorted signal from the film picture, and, furthermore, variations in the mean density of the film are less important. As a result, it may now be said that film transmission is superior to 'live' transmissions.

Recording Vision Programmes

What of the reverse process – the recording of television programmes upon film? Just as sound is recorded upon disc or magnetic tape for reproduction on a subsequent occasion, so there is a need for the vision programme to be recorded on film. Thus the Derby or the Boat Race can be watched by viewers who can look in only in the evening. Viewers have recently seen examples of such delayed presentation.

The first demonstration of the process was given a couple of years ago to the British Kinematograph Society, by Philip Dorté, then director of O.B.'s at Alexandra Palace.

The films shown – records of actual vision programmes – had been filmed with an ordinary intermittent cinematograph camera, and, for the reasons just described, less than half the scanning lines were photographed. Nevertheless, the quality of the films was acceptable.

A different principle has since been developed, almost identical with the film scanner now used. As a result, the storing of television programmes on film – complete with sound – is as simple as the storing of sound broadcasts – if not quite so inexpensive.

Large-screen Reproduction

New viewers at first experience a certain incongruity between the diminutive figures on the screen of the average home receiver – which varies from 7 ins. to 15 ins. in width – and the full-scale voices emanating from the loud-speaker. It is obvious from geometry that viewing a 12 in. picture at a distance of 5 ft is exactly equivalent to viewing a 20 ft cinema picture from a seat in the circle, 100 ft from the screen. But there is some psychological factor which enters into it, and in my view the ideal home receiver should produce a picture of say 30 ins. in width.

Sets are already on the market in the United States capable of producing a picture of this size or larger. Instead of the viewer seeing the actual face of the cathode-ray tube, the image on the face of a small tube is projected upon a screen, using exactly the same principle as is employed in large-screen television in the cinema.

Early attempts at big-screen television made use of mechanico-optical systems, in which rotating mirror drums produced the scanning raster. As far back as 1932, Baird gave a demonstration on a 9 ft screen, reproducing the Derby. In the Scophony system of 1938 the light was modulated by an extremely ingenious application of the piezo-electric principle. But it can be stated quite definitely that no mechanical system is capable of the immense speeds and perfect accuracy necessary for high-definition television.

Pre-war attempts to project the image of the cathode-ray tube employed enormous lenses, which had to be of considerable focal length because of the size of image to be projected – the 4 in. $\times$ 3 in. image of the tube – and of wide aperture to pass sufficient light*; thus a 12 in. $f/1.5$ lens was specially made⁸. Modern systems, both of home and cinema television, employ an optical system borrowed from astronomy, and known as the Schmidt system⁹.

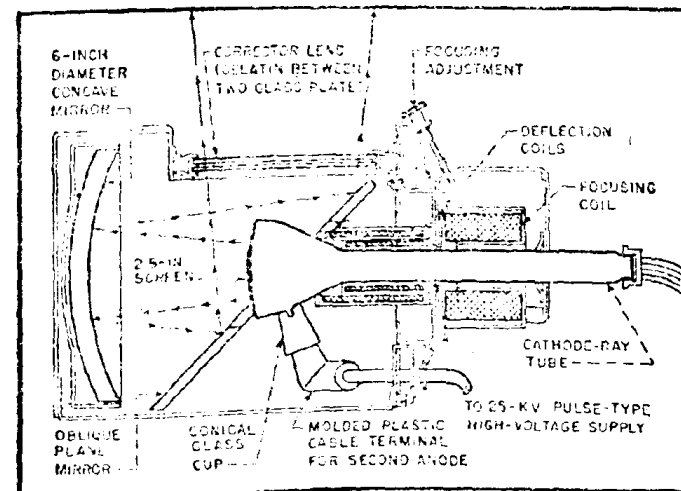


Fig. 5. – Schmidt optical system, as modified for the Philips home receiver. (Reproduced by courtesy of the North American Philips Co.)

Instead of a lens, a large spherical mirror reflects the image on the tube, which is placed at the focal point of the mirror. Such a mirror would of course produce spherical aberrations; to correct these a specially curved corrector plate is used, made of plastic (Fig. 5). Apparatus employing this system, and projecting a picture 3 ft or 4 ft in width, was recently shown at Radiolympia.

* The f value or *equivalent aperture* of a lens is an approximate measure of its light-transmitting power, and is equal to the focal length divided by the diameter of the working aperture.

The 'Cintel' cinema projector employs a 27 in. spherical mirror and an 18 in. correcting lens, giving an equivalent aperture of $f/0.8$ – much in excess of the aperture that would be obtainable with a lens system. For the production of the intensely bright picture on the tube a potential of 50 kv. is used. It is necessary to place the projector about 40 ft from the screen – in the front stalls of the cinema, and not in the projection room.

The equipment was demonstrated at a Bromley (Kent) cinema in December 1948, on a 15 ft. screen, and an image almost indistinguishable from that of a film projector was produced; the incidental brightness was 5 foot-candles, but, by the use of a directional screen, the image showed a high-light brightness of 8 foot-lamberts (the British standard of screen illumination in the cinema is from 8 to 16 foot-lamberts)*.

In the case of material transmitted from Alexandra Palace, the scanning lines were visible, and the picture was at times marred by shading. A programme was later transmitted from the company's own studio at Sydenham, from which it was carried by wire to the Crystal Palace, and thence on a 480 Mc/s radio link; for this a newly-designed camera was employed. The quality of projection left nothing to be desired; in contrast, sharpness, absence of shading and invisibility of scanning lines, it was indistinguishable from a well-projected film picture.

Intermediate Film Process

A totally different system of presenting the television image upon a cinema screen is the use of the so-called intermediate film process, originated by Baird. In this process

* The *foot-candle* is the brightness incident upon a surface 1 ft away from a standard candle. The *foot-lambert* is the unit of reflected brightness; for a matt surface the foot-lamberts are always less than the foot-candles, but a directionally reflective surface has a polar curve which at its peak will have a brightness such that the reflected foot-lamberts are in excess of the incident foot-candles.

the television picture is reproduced upon the screen of a small cathode-ray tube, and is photographed upon film; the film is rapidly processed by means of special plant, in heated solutions, in about 45 seconds, and dried in hot air. It is then fed into an almost normal film projector. By this means, ordinary film can be used, with a powerful arc lamp behind it.

Against these advantages is the cost of film stock and of the skilled labour for processing. For these reasons, the system of direct projection is almost certain to be adopted. Nevertheless, in America both RCA and the Paramount film organisation are marketing intermediate film systems, ready for installation in cinemas.

Cinema Installations

Whether or not London cinemas will actually be equipped for television by the time this article appears is anybody's guess. For the past couple of years negotiations have been proceeding between the Television Advisory Committee and the film renters, exhibitors and producers (known as the R.E.P. Committee) to devise a *modus operandi*. While in America cinema television is proceeding apace, at the time of writing arguments are still going on in this country with the object of overcoming the many problems – almost entirely, it must be emphasised, of a commercial and political, and not a technical, nature. The B.B.C. states that it would willingly see its programmes shown on cinema screens, while pointing out that much of its material is quite unsuitable.

A *quid pro quo* which has been for a long time under discussion would grant exhibitors permission to reproduce B.B.C. transmissions, in return for the industry allowing the B.B.C. to televise certain feature films. The R.E.P. points out that the proposal is one-sided; while the B.B.C. would acquire the right to show its viewers films costing many millions of pounds – at the subsequent expense of exhibitors – it is only an occasional news or sporting event which the exhibitor would wish to present, and then in many cases he

would be faced with copyright difficulties from the promoters of sporting events.

What the film trade demands – but sees at present little prospect of getting – is permission to establish its own transmitting stations, linked by radio or cable with cinemas. In America they have gone about things in a different way; the Society of Motion Picture and Television Engineers has made repeated requests for the allocation of wave-bands for cinema transmission; it appears that under American law such transmissions would not rank as broadcasts, and a broadcasting licence would not be needed. Up to the time of writing, no frequency allocations have been made other than for experimental purposes, and transmissions are being effected over co-axial cables.

The fear has been expressed that, were the film trade to set up its own radio transmitters, there would be nothing to prevent amateur viewers looking in on them. An ingenious suggestion has been made that in such cases the synchronising signals should be transmitted over ordinary telephone lines, and therefore be accessible only to authorised receiving stations.

The Immediacy of Television

Let there be no mistake, television offers the cinema audience something it has never had before. Technically, the picture on the screen may be indistinguishable from that of the film projector; but psychologically there is something totally different.

For some years before the war I saw the Derby and other sporting events on big-screen television at various demonstration centres. One saw staid business men jumping up in their seats with excitement – a sight never witnessed during a newsreel. The B.B.C.'s term for this difference is *immediacy*: one sees news while it is actually happening, and while the result is still in doubt, while in a newsreel one sees merely the pictorial record of an event the outcome of which is known.

Scanning Standards

A matter still the subject of technical argument is whether the existing British scanning standard of 405 lines will be adequate to produce a big-screen picture of quality. Calculation suggests that a definition of 800 to 1,000 lines is necessary to equal the film picture (France has adopted 819 lines); but the consensus of opinion following upon the Cinema-Television demonstration of 1948 was that, given well-tuned-up equipment, with adequate contrast and definition, and provided also that in neither transmitter nor receiver does 'pairing' occur (the superimposing of scanning lines, which effectively halves the definition) the 405-line standard was perfectly adequate.

There is, however, a move to replace interlaced scanning by sequential, since it is felt that not only does interlacing not perform what is claimed for it (the reduction of flicker) but it introduces numerous problems, some of which have been already referred to in connection with film scanning, and others of an electronic nature.

Needless to say, the scanning standard is bound up with the width of wave-band needed for transmission. The band-width at present occupied is 5 Mc/s; for a scanning standard of 1,000 lines it is believed that a band-width of 12 Mc/s would be necessary.

Experience shows that a factor of at least equal importance to definition is contrast, or *gamma*, and enormous improvement in picture quality has been effected by careful control of this factor.

Commercial and Scientific Uses

Television must not be considered merely as entertainment, even though that may be its principal field. It has already been used in research – in the Bikini Atoll atom-bomb tests, for example, television sets were placed in ships near the explosion, where no human being could have lived, and transmitted their images to receivers outside the danger zone. Attempts were made, in view of the brief period of the

explosion, to film the results on a high-speed cinematograph camera, but, as far as I have heard, these experiments were not too successful.

Television has, too, been used to enable a number of medical students to witness the details of a surgical operation. A vision camera was installed above the operating table, and connected with a number of viewing screens.

A commercial television equipment is actually on the market in America. It consists of a transmitter and receiver, connected by cable, and is intended for explosive tests and similar purposes for which direct observation is not practicable.

Colour Television

Television still lacks one thing which the film is at last learning to give us: colour. For a long time to come we are unlikely to see on a television screen (whether in the home or cinema) the beauties of a Technicolor picture.

Basically, colour television is no more difficult – some would say it is easier – than colour cinematography. But it has the one objection of needing to transmit two or three times the signal intelligence, and thus of needing a wave-band in the radio spectrum two or three times as wide as black-and-white.

Three systems of colour television have been demonstrated¹⁰. Two are based upon the additive synthesis of colour – upon the fact that by adding together red, green and blue light in suitable proportions, any desired colour can be produced. In one of these systems the colours are actually projected simultaneously upon the screen, while in the other they are produced consecutively, at such a rate that the eye integrates the three colours. (Present systems of colour cinematography are based upon subtractive synthesis, in which colours are matched by subtracting from white light a suitable mixture of blue-green, or cyan, magenta, and yellow rays. See *Science News* 3, 94 et seq.).

A system of the former type demonstrated some years ago

by Baird employed two colours only, red and blue-green. In one model a twin cathode-ray tube projected an image on either side of a central fluorescent screen; that on one side was red and that on the other side blue-green, and, since the screen was transparent, they were seen in superposition. In 1938 the same inventor demonstrated a two-colour picture on a cinema screen.

In the Columbia process, a disc carrying filter sectors of the three primary colours rotates in front of the transmitting camera, and a similar filter disc rotates in synchronism before the screen of the receiver tube. The three separation images are thus presented in rapid succession to the eye, which synthesises them into the colours as transmitted.

In a process demonstrated by RCA three cathode-ray tubes have screens respectively of red, green and blue, and the signals for each one have to be sorted from the composite received radio signal. By an optical system the three images are superimposed on the screen.¹¹

The same objection of needing greater space in the radio spectrum holds back the development of stereoscopic television, which is perfectly practicable as far as home sets are concerned. It was again Baird who originated a process in which twin images were produced side by side, and viewed through large magnifiers, like an old-fashioned stereoscope. Stereoscopy in the cinema is of course an altogether different matter.

Television in the Film Studio

I have throughout this article referred to the close affinity between television and cinematography. An even closer relationship is predicted for the future. Briefly, the forecast is that in the film studio, instead of film cameras television cameras will be employed on the floor: the pictures will be reproduced simultaneously on viewing screens for the director and other technicians, and will be transmitted to a central recording room, where they will be recorded in much the same way that sound is recorded to-day.

Advantages claimed for the system are greater flexibility, a greater measure of control, and economy in lighting, owing to the fact that the photo-cathode has greater sensitivity than the film emulsion. Equipment for this purpose is already in experimental use in France.

This proposal must not be confused with the system already operated experimentally, by Pye and Cinema-Television, whereby a television transmitter is mounted on the film camera, and operates a number of monitoring screens, so enabling the director and cameraman to see exactly what is being photographed. This system has obvious merits, and is certain to be adopted when practical snags have been overcome.

The possibility has been freely discussed of cinemas becoming mere television receivers - of the programme being radiated from a central transmitting station. Personally I do not expect to see the day when programmes will be transmitted by radio instead of arriving in tin boxes, but I have little doubt that it will ultimately come.

The film industry is in its sixth decade, while high-definition television (if one excepts the hiatus of the war period) is only in its first. The enormous technical strides made in this short period will without doubt continue, until one day the light-sensitive emulsion becomes a thing of the past. When that day arrives, we shall talk neither of cinematography nor television, but simply of the motion picture.

REFERENCES

1. *J. Telev. Soc.*, 5, No. 6, p. 185.
2. *J. Telev. Soc.*, 5, No. 4, p. 132.
3. *J. Telev. Soc.*, 5, No. 5, p. 156.
4. *Brit. Kine.*, 14, No. 4, April 1949, p. 113.
5. *J. Soc. Mot. Pic. Eng.*, 49, No. 1, July 1947, p. 57.
6. *J. Telev. Soc.*, 5, No. 6, p. 170.
7. *Brit. Kine.*, 12, No. 5, May 1948, p. 164.
8. *J. Brit. Kine Soc.*, 2, No. 2, Aug. 1939, p. 111.
9. *J. Telev. Soc.*, 5, No. 3, p. 86.
10. *J. Brit. Kine. Soc.*, 10, No. 1, Jan./Feb. 1947, p. 28.
11. *Ideal Kinema*, Oct. 1949, p. 25.

HOW AN AEROPLANE FLIES



Plate 1: What holds a heavy machine thousands of feet up in the atmosphere, and prevents it falling? The answer in brief is that it is the movement of the air across its wings which gives it lift. The propeller forces the aircraft forwards, and air must meanwhile naturally flow backwards across the upper and under surfaces of its wings, which are specially shaped to take advantage of these air currents. Pictures on the following pages show some of the properties of such air currents. (Crown Copyright Reserved).

SC 111

Plates 2 to 9 are reproduced by courtesy of the Shell Petroleum Co. Ltd. from the film *How an Aeroplane Flies* made by the Shell Film Unit.



Plate 2: A simple experiment: hold a piece of paper to your lips and blow. The paper, which hangs by its own weight, rises to the horizontal when air is blown fast enough across one of its surfaces, while the air on the other surface is more or less stationary. The moving air is in fact at a lower pressure than the rest of the atmosphere, and so the higher-pressure air underneath forces the paper up. This pressure difference is demonstrated in the following experiments.

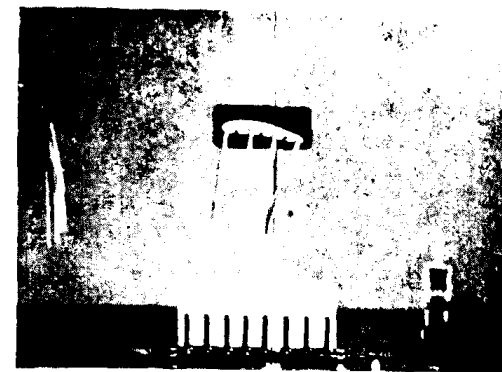


Plate 3: A section of model aeroplane wing is mounted on a stand, centre. There are four small holes (not visible) on its upper convex surface, and three on the lower concave side, communicating with metal tubes inside the wing. Those from the lower surface are seen projecting from the cross section, those from the upper surface are connected by rubber tubes with a row of manometers, filled with dark fluid. To the left is a wind machine; when it starts to blow, the white cotton streamers will flutter in the breeze.

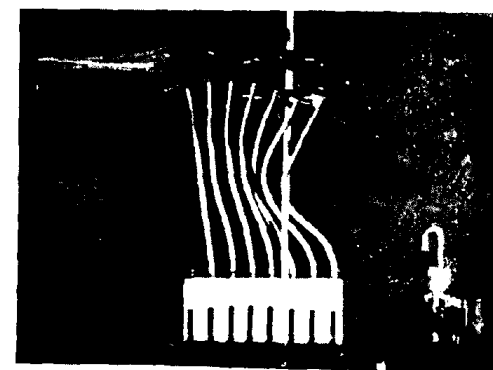


Plate 4: The wind machine in action, blowing across a curved metal plate instead of a wing section (ignore the upper horizontal plate, which makes no difference). Note that the dark liquid rises in the centre manometers above the horizontal line which marks atmospheric pressure. Under the conditions of the experiment this means that the pressure is *less* than atmospheric at certain points just above the curved plate.



Plate 5: A close-up of the model wing section. The holes in the upper surface are visible ...

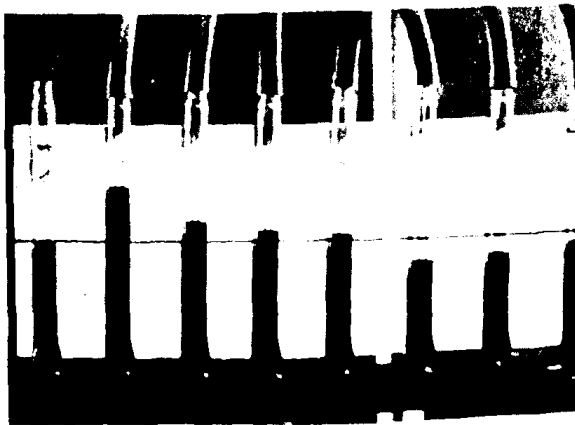


Plate 6: ... and the manometers do this when the wind blows. Numbers 2 to 5 are connected to the upper surface of the wing and register decreased pressure; the last three, connected to the under surface, register a pressure greater than atmospheric. In other words, when air blows across a wing, the wing is sucked up from above, and pushed up from below by the air current.

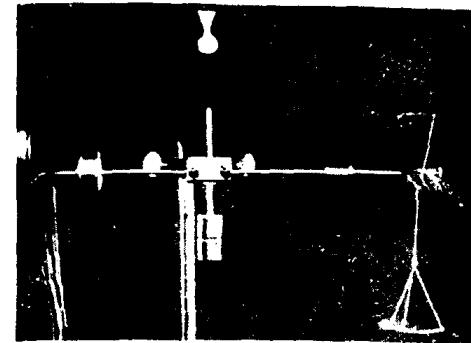


Plate 7: This suck-and-push can actually be measured in terms of a weight which can be lifted. Here is a simple apparatus set up in front of the wind machine. It is a long bar pivoted in the middle, with a fixed weight on one side (nearer the wind), and a wing section set at a certain angle to the air flow, as shown by the graduated scale, on the other. The pointer in the middle shows that in still air the system is exactly balanced with no weights in the pan hanging from under the wing.

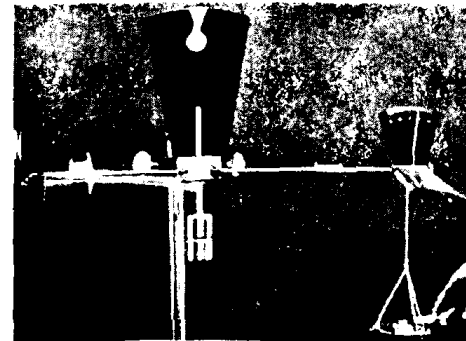


Plate 8: But when the wind blows, the wing is lifted and weights have to be added to the pan to bring the system back into balance. In this way the amount of lift obtained, with the wing set at different angles to the airflow, and with different speeds of airflow, can be measured; note that plate 8 is actually taken with the wing set at a slightly different angle from that of plate 7, but that this in itself does not affect the balance. The greater the angle, the greater the lift, up to a certain point, the 'stalling angle'.



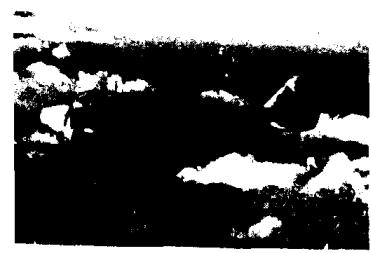
a



b



c



d

Plate 9: At the stalling angle, the air current no longer flows smoothly over the wing but becomes turbulent, and lift is lost. Air flow can be studied by putting chemical smoke into the air stream and watching how it flows over an object. (a) and (b) opposite show how air blown from the left impinges on a disc and a sphere, and becomes turbulent, whereas the flow over a streamlined cigar (c) is perfectly smooth. In (d) the aeroplane tries to approximate in shape to this cigar, for turbulence or 'drag' is an enemy of lift. But there is no space here to illustrate the behaviour on drag of the other factors which govern the aircraft in flight.

THE BACTERIOPHAGE

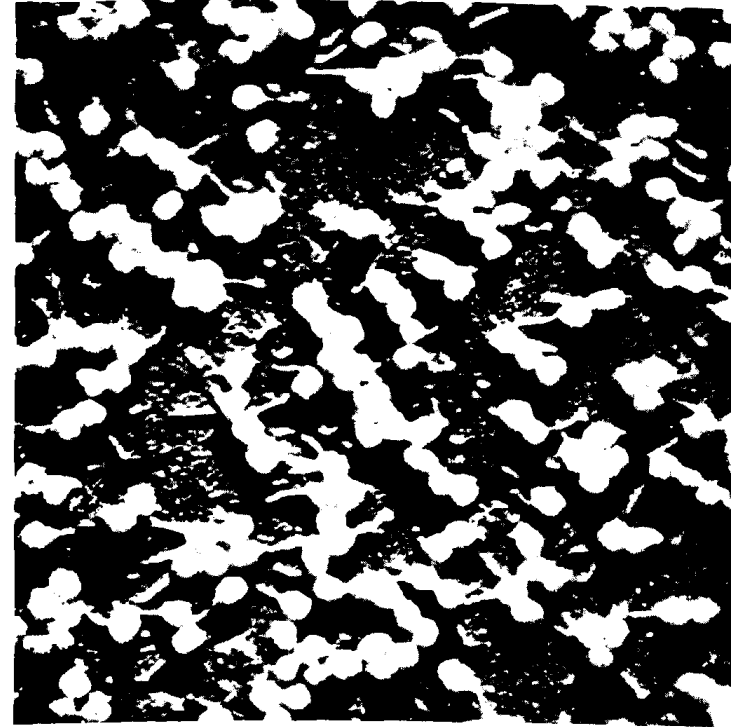


Plate 10: Clusters of the sperm-like elementary particles of one of the even-numbered coli bacteriophages as seen in an ultracentrifugally purified suspension. Magnification $\approx$ ca. 60,000 x.

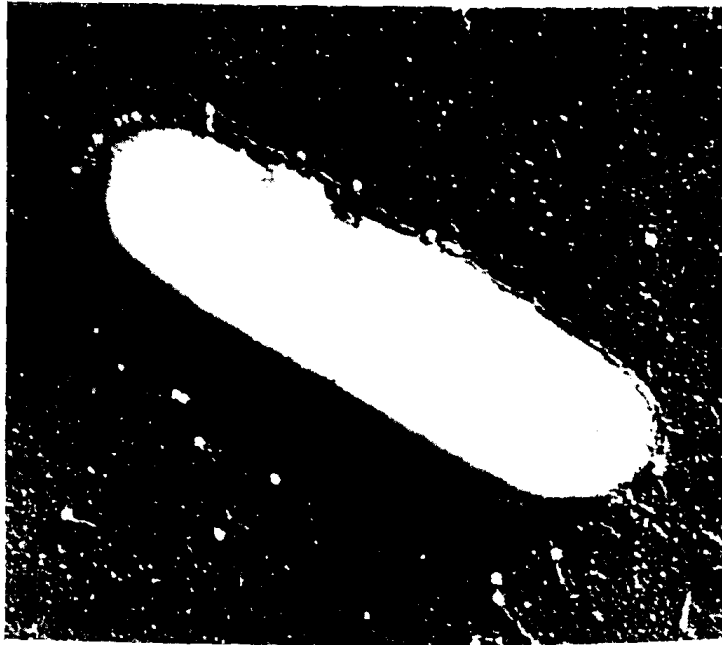


Plate 11: A colon bacillus being attacked by particles of the T1 bacteriophage. Several particles below the organism are easily recognised by their spherical heads and thin curly tails. More particles are adsorbed around the periphery of the bacterium. Magnification *ca.* 35,000 $\times$.



Plate 12: A colon bacillus lysed by the action of T2 bacteriophage, its protoplasm flowing out over the substrate. Spherical particles of this protoplasm can be seen around its edges. Magnification *ca.* 30,000 $\times$.

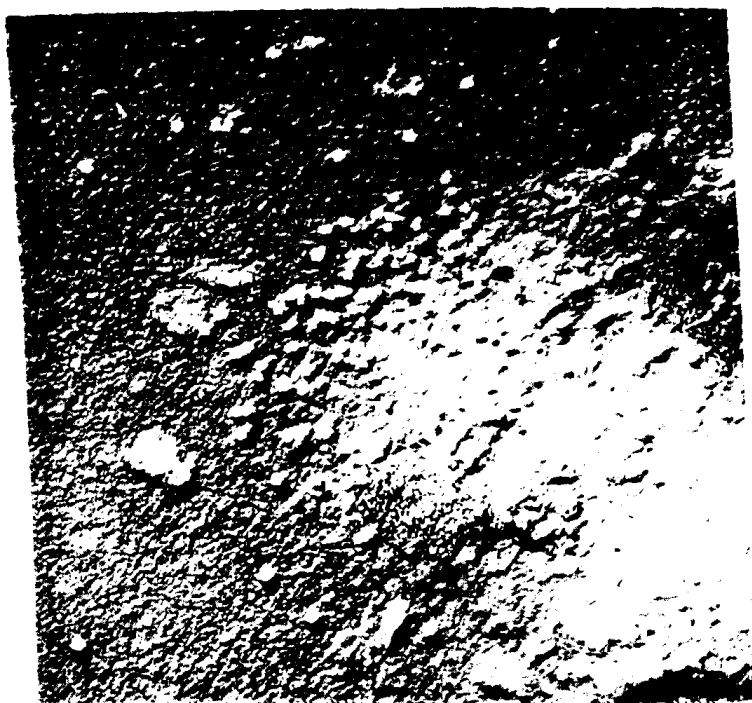


Plate 13: Elementary particles of T3 bacteriophage appearing from and still incorporated in a mass of protoplasm that has issued from a bacterium off the field of this picture. Magnification *ca.* 60,000 $\times$.

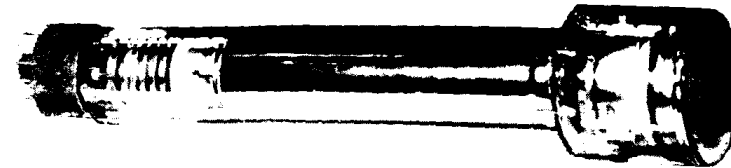


Plate 14: A nearly solid mass of T4 bacteriophage particles having the general shape of the bacterium upon which they have grown. A single particle with its short straight tail is evident above the mass. Magnification *ca.* 30,000 $\times$.



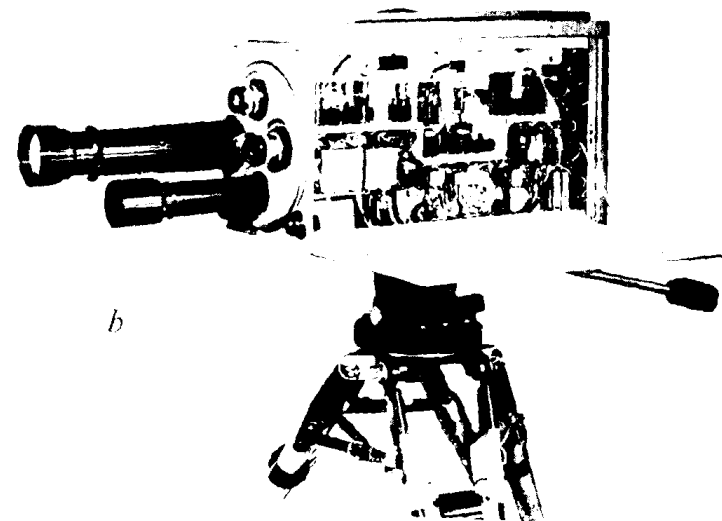
Plate 15: The remains of a partly divided pair of colon bacilli whose protoplasm has nourished so many T3 bacteriophage particles that they have been closely and orderly packed within its confines. After these particles have escaped they have left behind the unconsumed fibrous network seen in this photograph. Magnification *ca.* 25,000 x.

RECENT PROGRESS IN TELEVISION



0 5 10 15 20 25 30cm.

a



b

Plate 16

(a) Image Orthicon

(b) Marconi Television Camera

(the scale applies only to (a))

(By courtesy of Marconi's Wireless Telegraph Co. Ltd.)

MOLECULAR ARCHITECTURE IN PLANTS

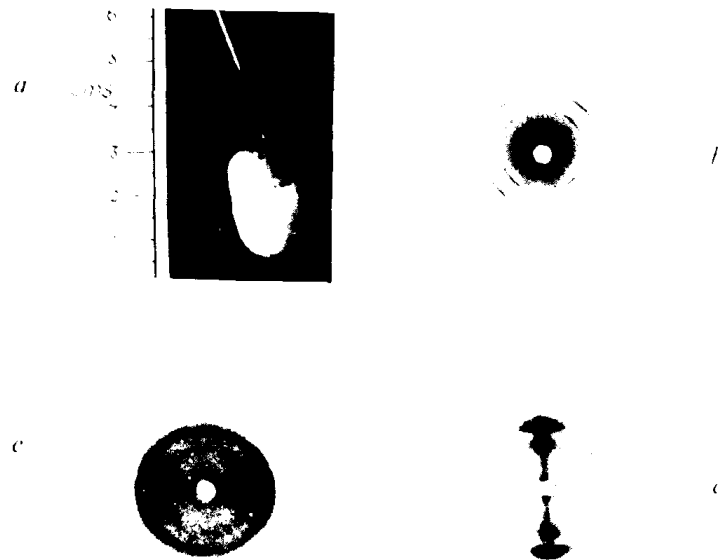


Plate 17: (a) Single cell of *Valonia ventricosa*. The cell has been pierced by a glass capillary, washed out with distilled water, dried and reinfated with air. The glass capillary can still be seen attached. Note the enormous size reached by these cells.

(b) X-ray diagram of a single piece of *Valonia* wall obtained by passing the X-ray beam normally through the wall. Two rows of spots can clearly be seen lying at rather less than a right angles to each other. Each of these rows of spots lies perpendicular to the set of cellulose chains which gives rise to it; hence it can be deduced that in this wall there are two sets of cellulose chains running almost at right angles to each other.

(c, d) X-ray diagrams of the wood of *Pinus sylvestris* (Scots pine) obtained by passing a beam of X-rays through a chip of wood rather thinner than a matchstick; wood from (c) the fourth annual ring. (d) the tenth annual ring. It is impossible to go into detail here but it will be obvious by mere inspection that the wood from the tenth ring is more crystalline, or at any rate corresponds to a better crystal, because it gives a sharper diffraction pattern.

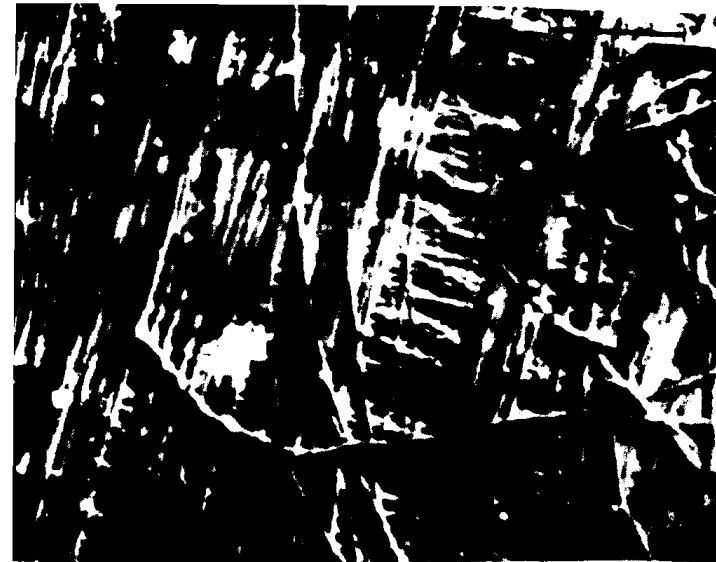


Plate 18: The surface of the wall of *Valonia ventricosa* as seen in the electron microscope. Notice the two sets of apparently endless threads making an angle of about a right angle with each other. The threads are remarkably uniform in thickness, of the order of 300 Å. (Magnification about 20,000 x.)

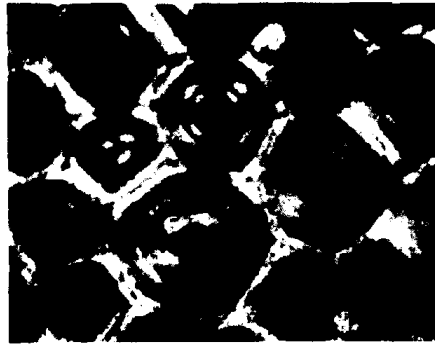


Plate 19: Thin cross section of bamboo fibres as seen under a polarising microscope with the nicols crossed. The cell outline of each cell is marked by a bright line. Within this can be counted three or four dark regions and three or four more bright regions. This can be shown to mean that in the bright parts the cellulose spiral is flat and in the dark parts the spiral is steep. A transverse section of wood would look essentially like this. (Magnification about 800 x.)

THE ELECTRICAL ACTIVITY OF THE HUMAN BRAIN

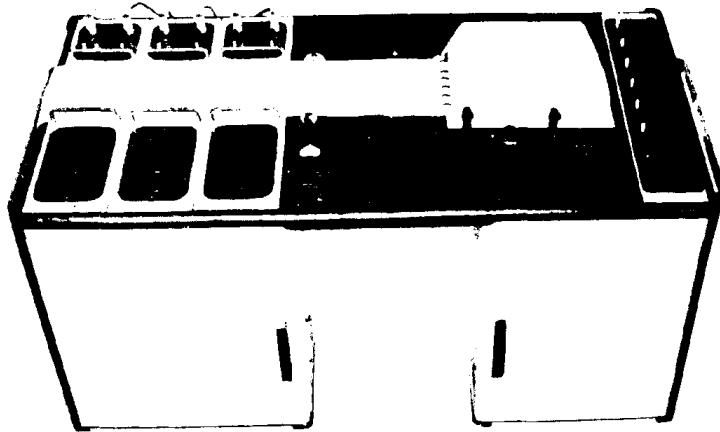


Plate 20: A modern 6-channel Electroencephalograph. The amplifiers are in pairs on the left, and the power supplies on the right. The pens, to the right centre, have an unusually large excursion, and for six channels the paper is 8" wide. (Photo by courtesy of Edison Swan Electric Company).

The Discovery of the Bacteriophage

F. W. TWORT, F.R.S.

I HAVE often been asked: – How did you come to discover the Bacteriophage? It is the kind of question to which it is difficult to give a simple answer, since so many factors played their part in the evolution of an outlook which made the discovery almost impossible to avoid. If now I am to give the picture in true perspective, the more relevant of precursory circumstances must be mentioned, if only very briefly, and the narrative must be given with candour, even at the risk of appearing pretentious.

As a prelude, it may be observed that three of the more important functions of the human brain are revealed in the capacities to observe, to reason logically, and to memorize. For successful research work, it is essential that these features shall be developed to provide a well co-ordinated outlook, and so ensure that the truth shall never be concealed or distorted, should the results of an experiment fall short of expectations or desires. As a child, I was fortunate in this respect; my father, a hard-working country doctor, and my dear mother – with her large family – both inspired my early outlook. Moreover, when later I was put to study medicine at St. Thomas's Hospital and Medical School, I had the good fortune to come under the guidance of men of character and ability, and, under their friendly care, my tuition left nothing to be desired.

During the latter period of my studentship, I developed a keen interest in pathological studies – more particularly bacteriology; but the fortunes of the family were small, so

immediately I had obtained my qualification M.R.C.S., L.R.C.P., it was imperative that I should earn at least sufficient to be able to pay for lodgings and food. Accordingly, I accepted the first paid post available, which was that of Assistant Superintendent of the clinical laboratory at the hospital. Although the salary was very small, it brought me under the kindly guidance of the director, Dr Louis Jenner, which enabled me, in the course of a year, to obtain an effective knowledge of pathological technique. I was then appointed Assistant Bacteriologist to the London Hospital, under Professor William Bulloch, F.R.S., and became responsible for most of the bacteriological work of that hospital. Bulloch soon became one of my closest friends, and, with the never-failing support for over seven years of this broad-minded Scotsman, I was enabled to gain experience in all branches of pathology, and at the same time carry out certain researches in the field of bacteriology.

In 1909, I was appointed Superintendent of The Brown Institution, London University, and thereby gained time and facilities to carry out research work in any branch of pathology or bacteriology in which I judged an advance in our knowledge could be made, provided expenditure could be kept within the limits of the small income available.

It must here be briefly explained that the Brown Institution was founded in 1871, under the will of Mr Thomas Brown of Dublin and London, to provide a hospital 'for the care and treatment of Quadrupeds or Birds useful to Man,' the business of the Institution to be under the direction and control of a Superintendent, and to be conducted in two departments - the Hospital, and the Laboratory. In the Hospital, both in- and out-Patients have been treated, and here the Superintendent has the aid of a fully qualified Veterinary Surgeon. The Institution was set up by London University, and over 260,000 sick animals were treated in the hospital before its destruction by a bomb in 1944.

THE DISCOVERY OF THE BACTERIOPHAGE 35

In the Laboratory various scientific researches have been carried out for different public bodies, as, for instance, the Royal Society, the Local Government Board, the Hydrophobia Commission and the Royal Agricultural Society. Professor Sir John Burdon-Sanderson, M.D., F.R.S., was the first Superintendent, and he was followed by such men as Professor W. S. Greenfield, Professor Roy, Professor Sir Victor Horsley, Professor Sir Charles Sherrington, Professor Sir John Rose Bradford and Professor T. Gregor Brodie; most of whom were Fellows of the Royal Society.

On being appointed Superintendent of the Brown in 1909, I meditated on the position of our bacteriological knowledge in relation to the fundamental biological laws which govern *all* living matter, and I accepted the principles of Evolution as a guiding basis. The reader must bear in mind that the essence of life is to build up organic substances - or fuel - with the help of energy derived from the sun's rays, and then to utilize that confined energy for its various functions by means of oxidation or combustion. However, nature is always striving to pursue an easier course, and a living thing will not build up its own fuel if it can obtain it from some other manifestation of life. Hence the constant struggle between all species, and the gradual evolution of means of aggression and resistance, leading, in the case of the animal kingdom, to a class which is parasitic on the vegetable kingdom, and could not exist without it. On the other hand, green vegetation can be self-supporting, and in the early stages of evolution must have been so, since no animal could have existed to supply food for plant life.

If now we take life in its most primitive forms, such as are represented by thousands of different bacteria, viruses and related types, it is obvious that such as existed in the tree of evolution before the development of the green chlorophyll-containing cell could not have obtained their energy either from animals or from green vegetation, but *must* have been able to build it up for themselves, without

the aid of chlorophyll, when the associated physical conditions were suitable. Their life, in fact, must have been determined by certain fundamental laws of nature which will have to be complied with before true growth in an independent state is achieved in the laboratory. Although many varieties belonging to the bacterial and the virus groups have acquired the capacity of living on the organic materials built up by animals and plants, it is imperative to recognize this as a comparatively recent evolutionary development, which is very unlikely to have cancelled out any of the basic requirements of life which governed the existence and growth of their wild independent ancestors.

I had these considerations in mind when, in 1909, I selected 'The growth requirements of primitive forms of life' as being a fertile field for study in relation to infectious diseases of man, animals, and plants. The problem presented three aspects, these being the chemical composition of the pabulum on which cultivation was to be tested, the nature of the physical conditions to be set up with different media which might be calculated to stimulate the functions of independent life, and the reaction of one form of life on another.

In the first place, consideration was given to the three conditions, Leprosy, Tuberculosis, and Johne's disease of cattle, for, although the causative bacilli of these diseases belong to the same biological group, only the tubercle bacillus, at that time, had been grown on an artificial medium altogether apart from its usual living host. In view, however, of the close relationship between the three bacilli, it was thought probable that the same chemical substances would be required by all for the fulfilment of their lives, but that one of these was of a very special nature, and, although it could be elaborated from standard foods by the tubercle bacillus, the other bacilli had lost that capacity. The problem, then, was to find this special product, or 'Essential Substance', in nature and incorporate it in the medium, when I expected to obtain a growth of both the

lepra bacillus and Johne's bacillus. For various reasons I anticipated finding it stored as a reserve food supply in the fatty covering developed by the tubercle bacillus and its wild ancestors, and this proved to be correct. Moreover, from the growth of Johne's bacillus thereby obtained, a true diagnostic vaccine was soon produced, thus enabling this devastating disease of cattle to be brought under control. Unfortunately, no financial support was forthcoming to allow an extension of the work on essential food factors to be made. The name Vitamin at that time had not been coined, although my 'Essential Substance' has since been named 'Vitamin K.'

However, from the success obtained in the cultivation of Johne's bacillus in 1910, and from the results of many additional experiments which threw light on the production of the 'Essential Substance' and other prime necessities of life, the basic conditions required for the cultivation of viruses became patent. I accordingly transferred my researches to that group of organisms. Now, it has already been emphasized that there must exist types of life which are lower in the scale than those containing green chlorophyll, and which are capable, under suitable conditions, of building up their own nutriment without its aid in the world production of organic foodstuffs. This applies both to bacteria and to that vast field of still more primitive life which includes viruses. We know, in the case of bacteria, that for every variety which has become pathogenic, or disease-producing, there are many parallel varieties which have not acquired the capacity of fighting their way into other living creatures, and there is every reason for believing that the same rule applies to viruses. Many bacteria, it is true, can obtain their food supply from the dead products of other organisms, but, as these products could not have existed when the first living things were created, they must be recognized as a later substitute for some genetic necessity of primitive life. Indeed, it would seem probable that there remain in nature to-day many varieties of bacteria which,

like the viruses, have never been cultivated on artificial media because either the media or the physical conditions employed in association therewith lack this vital factor.

With these considerations as a basis, many experiments bearing on the growth of viruses were started. The selected materials covered a wide range of diseased tissues obtained from man, animals and plants, and such substances as soil, dung, hay and pond water which might be expected to contain non-pathogenic wild varieties belonging to the virus group; and which, if the missing special foods or physical conditions were supplied, should be the easiest to grow in pure culture. It is impossible in this short essay to give even an abridged description of the many experiments which were carried out, and I must confine my observations to the particular series which first led me to the discovery of the bacteriolytic agents which have since been named 'the Bacteriophage.'

For one series of experiments I selected glycerinated calf vaccinia (the fluid used in human vaccination), since tubes of this were easy to purchase for a small sum, and it was well known that the lymph at that time nearly always contained, not only the virus of vaccinia, but a certain number of bacteria, and micrococci such as are capable of producing pustules. Moreover, it was considered possible that some of these microbes, which could be grown easily, might be endowed with the capacity of supplying some 'Essential Substance' for the vitality and growth of the vaccinia virus, in a manner somewhat similar to that of the 'Essential Substance' which I had already discovered as the missing requisite for the cultivation of Johne's bacillus. Moreover, as it was recognized as a fundamental principle that all forms of life were open to attack by some other form, the possibility of a direct infection of the living micrococci by the living virus of vaccinia was considered.

In these experiments there was another very important factor which had to be taken into account. My experiments on the production and storage of the Essential Substance -

or Vitamin K - for the growth of Johne's bacillus, by such allied bacilli as the tubercle and Timothy grass varieties, had clearly demonstrated that production was determined largely by the nature of the medium on which they were grown and on the conditions associated with their cultivation. Further, that a maximum production of the Essential Substance by, for instance, the Timothy grass bacillus in no way ran parallel with the prolific growth of the bacillus obtained on certain highly-nutritious media. It is not possible to describe here in detail the many variations encountered, and it must suffice to place on record that my vaccinia experiments were carried out under the influence of the knowledge already gained.

For many experiments a 1½ per cent agar medium, prepared with some meat extract, was used as a basis. Agar, it may be noted, is a gelatinous material obtained from seaweed, which melts at the temperature of boiling water but remains in jelly formation at body temperature, thereby allowing it to be melted and set in slopes in test tubes after sterilization, and its smooth surface inoculated with material to be tested before the tubes are incubated at body temperature.

The first results were obtained with certain of these agar media. Tubes which had been inoculated with vaccinia lymph and incubated at 37° C. for 24 hours usually showed a certain number of colonies of the white and yellow varieties of micrococci, but some tubes, on closer inspection with a hand lens, showed in addition some minute glassy, transparent areas which failed to grow when sub-cultured. Moreover, on keeping the tubes for a further period, some of the colonies of cocci, which had appeared to be normal, started to turn transparent at the growing margins and, on attempting to make sub-cultures, they failed to grow. Furthermore, when a fresh, vigorous-growing, young culture of one of the micrococci was taken and the pure growth touched with a minute portion of the glassy material, this also, on further incubation, soon started to turn transparent

at the point touched, and the change gradually spread over most of the coccus growth. It was at once clear from these early experiments that I was dealing not with any purely degenerative change in the cocci, but with an acute infectious disease of the colonies which, like most of the contagious diseases to which other forms of life are subject, was much more potent and active on the young than on the old: in fact, the infection had very little, if any, visible action on old cultures of micrococci which had ceased to grow, or on young cultures which had been killed by heat.

The question that had now to be decided was: - What was the nature of this infective agent, and to what extent did other varieties exist in nature? An examination of the glassy material, when stained, under a high-power microscope, revealed only very minute granules, but there was no definite proof that these represented the infective agent. Some, at least, could be assumed to be fragments of the micrococci which had been broken down or lysed. It was on account of the cocci becoming dissolved in this way that I named the contagion the Bacteriolytic Agent. That was in 1915, when I published my first paper on the research in *The Lancet*, and before the name Bacteriophage had been coined.

Additional experiments proved the agent to be exceedingly small, for, if infected colonies of micrococci were made into an emulsion with sterile saline, or water, and the fluid then passed through a fine porcelain filter, the filtrate always contained the infective agent in an active form. The agent, thus freed from micrococci, was now tested on other bacteria, but the tests proved that it had not the same effect on all the many distinct varieties tested. On the other hand, tests on animals gave no indication that it represented the virus of vaccinia, although it showed many characters which suggested that it should be considered as belonging to the virus group. From this standpoint it is important to note that it showed no evidence of independent growth on any artificial medium then being tested, unless the grow-

THE DISCOVERY OF THE BACTERIOPHAGE 41

ing micrococcus was also present for it to live on. On the other hand, it retained its vitality, under favourable conditions, for many months without the presence of its host - the growing micrococcus.

Now, there are many virus diseases of man, animals and plants, most of which are specific for some particular host, and it seemed reasonable to expect that the bacteriolytic agent would be found to exist in nature in many varieties which, for the most part, would prove to be specific for some particular variety or type of bacterium. Accordingly, I explored nature for these other varieties, applying the same general biological rules which had guided me throughout in my researches. The results, put briefly, soon proved that other varieties did exist, that they possessed the same general characters as my original strain, and that they were as specific in their ability to attack bacteria as are the different viruses which produce diseases in animals and plants. As examples, it may be mentioned that different strains were obtained from the intestinal mucous membrane of a dog suffering from acute distemper which acted on a member of the coli-typhoid group of bacilli. Another strain, which infected other members of this group of bacilli, was recovered from material collected from the intestinal tract of a case of infantile diarrhoea and vomiting.

The next most obvious point that required to be decided was - to what extent can the different varieties of 'Bacteriophage' be used as a curative agent for diseases caused by different bacteria? Can a bacterium which is causing a disease be attacked and killed by some specific bacteriophage? Many experiments bearing on this point were carried out, but it would require an extensive essay to relate the technical details of these in true perspective, and any attempt to give a short summary would inevitably fail in its object. A few points, however, may here be noted. While some subsequent workers claimed very favourable results, other workers reported the absence of any beneficial effects. For my part, I am not prepared to dogmatize, for, as already

pointed out, when two forms of life come in conflict with each other, the result of the combat depends very largely upon the physical and chemical conditions under which the battle is fought. In modern bacteriological work, I suggest that sufficient consideration is rarely given to this aspect of a problem, and this applies also in the case of treatment of bacterial diseases with a bacteriophage, notwithstanding the thousands of experiments that have been carried out. Independent workers should meditate on this fundamental fact. In test-tube experiments it is easy to arrange conditions so that a bacteriophage will kill off its host - the bacterium, but that it is equally easy to arrange another set of conditions in which the bacterium will kill or inhibit its invading bacteriophage.

Now, if this is so in the case of a bacterium and its bacteriophage in a test tube, there is no logical reason why the same principle should not apply throughout nature for other virus and bacterial infections which may occur in man, animals and plants. Before, however, the precise conditions which determine the activity and control of such viruses can be properly established, it is of prime importance that means should be found whereby they may be grown and studied on some artificial medium apart from other forms of life. So far this has not been done with certainty because, I am convinced, some additional factor of an exceptional nature is required in order to satisfy their fundamental physiological needs.

An extensive research, based on this conception, was mapped out, and from 1930 to 1939 the Brown Institution laboratories were, by degrees, re-equipped with special apparatus to meet the requirements of the investigation, which in its earlier stages was financed jointly by London University and the Medical Research Council. Progress was satisfactory up to 1936 when the Council withdrew its support, thereby greatly hampering the work. Nevertheless the investigation was continued actively and as extensively as circumstances would allow, and by 1939 I knew

that the technique which was being evolved was correct. Striking results of a positive nature were obtained, but, unfortunately, in 1944 the laboratories of the Institution were destroyed by a bomb, and the University deprived me of my post and all facilities for completing the research.

In conclusion, I would remind my readers that illness resulting from a virus disease in man often ends in death, or perhaps some malformation as in the case of Infantile Paralysis. Moreover, the various types of illness prevailing throughout the animal and vegetable kingdom, such as varieties of distemper in cats and dogs and mosaic diseases of plants, produce annually an incalculable economic loss. Yet my research, aiming at their control and offering, I am convinced, such high prospects of success, has, in its last stages, been foiled.

The Bacteriophage

DR FELIX D'HERELLE

ON the 15th September, 1917, Dr Roux presented to the Académie des Sciences a note (by me) entitled: *On an invisible microbe, an antagonist of the dysentery bacillus*. In the body of the note I termed this microbe *Bacteriophage*.

Since then more than 6,000 notes and papers have appeared on this subject throughout the world, and, far from being exhausted, it still grows in importance, for step by step as more research is carried out it becomes ever clearer that the bacteriophage plays an active part in all phenomena involving germs.

Diarrhoea in Locusts

In 1910, I was in Mexico, in the state of Yucatan, when an invasion of locusts occurred; the Indians reported to me that in a certain place the ground was strewn with the corpses of these insects. I went there and collected sick locusts, easily picked out since their principal symptom was an abundant blackish diarrhoea. This malady had not as yet been described, so I studied it. It was a septicaemia with intestinal symptoms. It was caused by bacteria, the locust coccobacilli, which were present almost in the pure state in the diarrhoeal liquid. I could start epidemics in columns of healthy insects by dusting cultures of the coccobacillus on plants in front of the advancing columns: the insects infected themselves as they devoured the soiled plants.

During the years which followed, I went from the Argentine to North Africa to spread this illness. In the course of these researches, at various times I noticed an anomaly shown by some cultures of the coccobacillus which intrigued

THE BACTERIOPHAGE.

45

me greatly, although in fact the observation was ordinary enough, so banal indeed that many bacteriologists had certainly made it before on a variety of cultures.

The anomaly consisted of *clear spots*, quite circular, two or three millimetres in diameter, speckling the cultures grown on agar. I scratched the surface of the agar in these transparent patches, and made slides for the microscope; there was nothing to be seen. I concluded from this and other experiments that the *something* which caused the formation of the clear spots must be so small as to be filtrable, that is to say able to pass a porcelain filter of the Chamberland type, which will hold back all bacteria.

However, the appearance of these clear patches was inconstant. I sometimes went weeks without seeing a single one, and I could not reproduce the phenomenon at will; I therefore could not study it.

In March, 1915, during the first World War, a large invasion of locusts appeared in Tunisia, threatening to destroy the harvests which were then so vital; I was given the job of starting an epidemic amongst them. As the result of the infection there was a considerable mortality, and, even more interesting, when the following year all the rest of North Africa was again invaded, Tunisia remained free: the illness had continued to rage amongst those swarms of locusts which had survived long enough to move South at the end of the season, and had brought about their destruction during the winter.

In the course of this campaign, I again observed my *clear spots*, and before returning to France I stayed for a time at the Institut Pasteur in Tunis, to investigate their significance.

I showed them to Charles Nicolle, then director of the Institute, and he said to me: 'That may be the sign of a filtrable virus carried by your coccobacilli, a filtrable virus which is the true pathogenic agent, while the coccobacillus is only a contaminant.' So I filtered emulsions of cultures grown on agar and showing the clear spots, and tried

to infect healthy locusts with the filtrate, but without result.

On my return to Paris in August 1915, I was asked by Dr Roux to investigate an epidemic of dysentery which was raging in a cavalry squadron, then resting at Maisons-Laffitte. I thought that the hypothesis put forward for the locusts' illness might be helpful in understanding human dysentery. I therefore filtered emulsions of the faeces of the sick men, let the filtrates act on cultures of dysentery bacilli and spread them after incubation on nutritive agar in petri dishes: on various occasions I again found my clear spots, but the feeding of these cultures to guinea pigs and rabbits produced no disease.

At this time we often got cases of bacillary dysentery in the hospital of the Institut Pasteur in Paris. I resolved to follow one of these patients through from the moment of admission to the end of convalescence, to see at what time the *principle* causing the appearance of the clear patches first appeared. This is what I did with the first case which was available.

The first day I isolated from the bloody stools a Shiga dysentery bacillus, but the spreading on agar of a broth culture, to which had been added a filtrate from the faeces of the same sick man, gave a normal growth.

The same experiment, repeated on the second and third days, was equally negative. The fourth day, as on the preceding days, I made an emulsion with a few drops of the still bloody stools, and filtered it through a Chamberland candle; to a broth culture of the dysentery bacillus isolated the first day, I added a drop of the filtrate; then I spread a drop of this mixture on agar. I placed the tube of broth culture and the agar plate in an incubator at 37°. It was the end of the afternoon, in what was then the mortuary, where I had my laboratory.

The next morning, on opening the incubator, I experienced one of those rare moments of intense emotion which reward the research worker for all his pains: at the first glance I

saw that the broth culture, which the night before had been very turbid, was perfectly clear: all the bacteria had vanished, they had dissolved away like sugar in water. As for the agar spread, it was devoid of all growth and what caused my emotion was that in a flash I had understood: what caused my clear spots was in fact an invisible microbe, a filtrable virus, but a virus parasitic on bacteria.

Another thought came to me also: 'If this is true, the same thing has probably occurred during the night in the sick man, who yesterday was in a serious condition. In his intestine, as in my test-tube, the dysentery bacilli will have dissolved away under the action of their parasite. He should now be cured.'

I dashed to the hospital. In fact, during the night, his general condition had greatly improved and convalescence was beginning.

From that moment, I understood how to reproduce the phenomenon at my convenience; I could study it. The first note, appearing in 1917, was followed by several others, all sponsored by Dr Roux. In these communications I gave an account of experiments touching the nature of the phenomenon of bacteriophagy, its generality, the characteristics of the bacteriophage; I discussed its nature, the peculiarities of its mode of action and, finally, the defence reactions of the attacked bacteria. These various researches, including those I made during a visit to Indo-China in 1920, were finally collected into a single work which Dr Roux permitted me to publish in 1921 as one of the monographs of the Institut Pasteur, under the title *The Bacteriophage, Its Role in Immunity*.

Indifference, Scepticism

All this in the middle of general indifference: at that time there was nobody, I think, except Dr Roux, who took the matter seriously. Some thought I was a visionary, others a humbug; many certainly thought to themselves what one colleague said aloud: 'If this Bacteriophage really exists,

during all the time I have been a bacteriologist I would surely have observed it.'

One anecdote shows to what lengths scepticism went. An American, Miss Harde, who worked at the Parke laboratory in New York, spent several days in Paris in 1919; she came to see me, and I showed her how cultures dissolved in the course of two or three hours after the addition of a trace of bacteriophage culture. She asked me to give her a little of such a culture, to give a demonstration back in New York. Some time afterwards I had a distressed letter from her: she had indeed given her demonstration, but she had been laughed at and congratulated 'on having learned such fine bacteriological conjuring tricks in France.'

In the end, in the face of my insistence, a few bacteriologists became willing to recognise that all my experiments were precise, that they were easy to repeat. But while some, the minority, admitted the correctness of experiments showing that the Bacteriophage has all the physiological characteristics of a living thing, others interpreted them in different ways, pretending they were the effects of a chemical ferment coming, according to some, from the animal body infected by the bacteria, according to others, from the bacteria themselves.

I did not think, then, that a day would come when an electron microscope would be invented, using rays of a much shorter wavelength than those of light, which would allow the photographing, or rather radiography, of particles of some ten millimicrons diameter, a size which from graded filtration experiments I had recognised as that of the bacteriophage I was studying and which I thought was that of all bacteriophages.

Is the Bacteriophage Alive?

A criterion of life exists: it is the possession of a metabolism, which shows itself objectively in the secretion of digestive juices of a chemical nature, diastases or enzymes. Anything which, when placed in suitable conditions, secretes chemical

ferments, is living by definition; if it does not secrete, but remains inert, either it was once alive but is now dead, or we are dealing with something inert by nature.

In 1921 by precipitating cultures of certain bacteriophages with alcohol Eliava and I obtained ferments acting on bacteria in the absence of all living bacteriophage particles. Since then, two of my collaborators, Sertic and Boulgakov, working together, have carried out research in this field for more than ten years. They have made a complete study of the digestive enzymes secreted by the particles of various strains of the bacteriophage attacking bacillus coli. They have shown that the same particle is able to secrete several different enzymes, giving the experimental physiological proof of the living nature of the bacteriophage; and to complete the work there was no need to 'see' the bacteriophage.

But this physiological proof not only means that bacteriophages are living, it proves that they are *completely* alive. Many biologists have seen in filtrable viruses in general, and in the bacteriophages in particular, intermediates between chemical substances and truly living beings. The demonstration of the secretion of chemical ferments by bacteriophages shows that these beings possess a complete metabolism, identical with that of the cells of the most highly developed organisms. The conclusion is that life is one, that no intermediates exist: all things endowed with life are fully alive, and all living things are alive in the same way.

Bacteriophages then are fully living things because they satisfy the criterion of life; but one might believe, *a priori*, that they must be very simple in structure, and therefore very little differentiated. A complete mistake. In 1919 I wrote in a note to the Société de Biologie: 'I have isolated several hundred strains of this bactericidal principle, and I have not yet found two exactly alike.' And in the last collected work published in 1938, I repeated 'Each race shows characteristics differentiating it from all other races.' Amongst these different characteristics are:

Virulence, that is to say, power of attacking bacteria: it varies from hardly detectable action to an effect so brutal that in less than three hours under experimental conditions all the bacteria in a fully-developed culture are completely destroyed and digested.

Extent of action: Certain varieties of bacteriophage attack only rare strains of a single bacterial species, while others can attack, simultaneously, bacteria of different and even widely separated species.

The Temperature at which they are destroyed, which varies from 52° to 82° according to variety.

Vitality: Some types of bacteriophage survive only a few days in a culture, others a few weeks, or a few months, while yet others have such vitality that a culture preserved in sealed ampoules at ordinary temperature is still alive many years later. Thus out of 22 different races of bacteriophage isolated from convalescent cholera stools and left in India in sealed ampoules in 1927, four were still living and virulent in October 1947.

Considering their extremely variable nature, and the diversity of size and shape revealed by the electron microscope, one would be inclined to agree that bacteriophages form a whole complete flora, which might be divided from the point of view of systematics into species and even different genera. But it is not as simple as that. First of all, the characteristics of different races are independent of size; in each of the four morphological types known hitherto, one finds all possible variations. In its other characteristics, a dwarf bacteriophage may be close to a giant bacteriophage, while two dwarfs or two giants may differ greatly from one another.

If we seek the cause of this diversity we see at once that it lies in a propensity to imitation and a power of adaption, both of which are more marked in the bacteriophage than in any other living thing. Experimentally, the characteristics of a bacteriophage can be changed within very wide limits. I have elsewhere shown that this power of adaption and

imitation differs in different strains, and is particularly marked in bacteriophages first isolated from their natural environment. Type collections long maintained in the laboratory by *in vitro* culture become rigid and their characters fixed.

The Bacteriophage is everywhere

Where are bacteriophages found? Wherever there are bacteria; that is, everywhere: in the atmosphere, in the soil, in river water and in the sea; the more bacteria there are, the more bacteriophages: sewage is an inexhaustible source of all kinds of bacteriophage. There is a natural reason for this: bacteriophages are obligatory parasites, they can multiply only at the expense of living bacteria, and they are found only where their reproduction is possible.

But their favourite haunt, where bacteria actually exist in the greatest variety and profusion, is in the intestinal contents of all living things which have a digestive tract. Let us see what happens there. Imagine a man in a state of normal health. Isolate from his stools the bacteriophages there present – a simple job since, thanks to their extreme minuteness, the bacteriophages easily pass through filters of the Chamberland type which hold back bacteria.

The filtrate from the excreta then contains all the bacteriophages and no bacteria. If we let a drop of such a filtrate act on a sensitive culture, the effects (clear spots on agar, more or less complete solution of a culture) allow us to show the presence of bacteriophages and to determine the intensity of their action towards this or that microbe. Such researches show that in the normal man bacteriophages highly active against *bacillus coli* are always present in the intestine. Now let us see what happens in a case of illness.

My first researches were carried out on people suffering from epidemic dysentery (*Shiga bacillus*). Studying them day by day, this is what I found. At the beginning, when blood was present in the stools, *Shiga bacilli* were isolated and the filtrates obtained from the stools contained only

bacteriophages active against bacillus coli. After one, two, three or four days, and sometimes longer, depending on the case, the picture was still the same. Then in a stool filtrate I suddenly found the presence of bacteriophages dissolving dysentery bacilli isolated originally from this same patient; and the bacillus coli bacteriophage was also still present. A few hours later the blood disappeared from the stools and the patient began his convalescence. It is difficult to think that this is mere coincidence, because the phenomenon is constant.

Here now is what I found in cholera. In Indo-China in the course of 1920, and later in India in 1927, I studied cholera patients, several hundreds of them. It is a most interesting disease from the present point of view for two reasons: first because of the high mortality, which can be a hundred per cent and is rarely less than 50 per cent; and secondly because of its brevity: in the great majority of cases in less than four days, even in as little as two, the patient is dead or cured. Under such conditions the cause of cure should be very apparent, and that is in fact what research showed. Every cholera patient was certain to die if a bacteriophage active against the cholera vibrio had not appeared in his stools within 48 hours of the first symptom. I never observed a single exception to this rule. On the other hand every patient in whom a bacteriophage very *strongly* active against the vibrio was present after this time quickly entered the convalescent stage, however grave had been the initial symptoms. Those in whom in the first 48 hours a bacteriophage of only feeble strength had appeared had a prolonged illness and ended either by dying or by being cured, depending on whether the bacteriophage became feebler and disappeared or grew in strength, as shown by *in vitro* test.

In cholera the phenomenon of the bacteriophage *in vivo* is so clear-cut that there can be no doubt of the fact that in this illness it is the sole cause of recovery. But it is not only in intestinal illness that the bacteriophage manifests

its activity. I studied human bubonic plague in Annam and in Egypt; in patients who got better spontaneously, bacteriophages provoking the destruction of the plague bacillus were present in the stools and also in the buboes (swollen glands characteristic of plague).

In 1919 there raged in France among hens a fatal epidemic of an illness recently imported from the United States where it had been known in its endemic form and called 'fowl typhoid.' Before 1919 it had not been observed in France, though it subsequently came to stay. For me it was a wonderful opportunity to study the behaviour of the bacteriophage, for with the animal illness I could experiment at my convenience. I believe indeed that experimentation is significant only when carried out on animals *naturally* sensitive to an infection. This is the principle which Pasteur always observed and its infringement inevitably leads to error.

'Cure is infectious'

The organism of bird typhoid is a salmonella, a close relation of the (human) typhoid germ. The disease was raging in a district which I knew well, on the borders of Champagne and Burgundy. Each day I visited the poultry farms: over several days I did not see a single recovery, every hen infected died, and one by one in the space of three or four days they all succumbed. I analysed numerous specimens of excreta; the salmonella was present, but active intestinal bacteriophage was absent in all birds, whether infected or still healthy. Finally, in one poultry house, a hen got better; in its excreta was a bacteriophage powerfully active against the disease germ. But a curious phenomenon occurred in the sequel. In this hen-house the epidemic ended suddenly: I tested the excreta of the surviving hens and all contained the same bacteriophage; and then it was present everywhere on the farm, notably in the faeces of pigs, horses, cows, dogs. Later I again observed this sudden ending of the disease after a spontaneous cure,

accompanied each time by the same spread of bacteriophage.

One conclusion seemed to emerge: the infectiousness of the cure. A sick hen spreads disease bacteria around her, the epidemic grows. A hen gets better because in her intestine the bacteriophage normally present and growing at the expense of coli bacilli acquires virulence for the salmonella and starts a cure. She now spreads round her the bacteriophages multiplying in her intestine: all the hens still alive at that moment swallow them, and henceforth are protected from infection.

Here are other facts of a similar kind:

In 1927 I studied a violent epidemic of cholera which was raging in the Punjab. In villages which were not yet infected I could find no evidence either in the well water or in the bodies of flies (the two usual agents transmitting the infection) of cholera vibrios or of bacteriophages with some activity against them. If the epidemic appeared in a particular village, always following the arrival of some person incubating the disease, and he was the first patient, from that moment the vibrios were spread in the surroundings and one found them in the flies and in the well water, but never any bacteriophages. One patient got better spontaneously, and bacteriophages virulent for the cholera vibrio began from that moment to be found around the convalescent: first of all in the well of his own house. Cures increased, and the epidemic ceased when these bacteriophages were widely spread in the district. An epidemic of cholera spreads by the spread of vibrios from sick patients; it dies out as a consequence of the diffusion of bacteriophages from convalescents.

Phage therapy is an All-or-Nothing therapy

These two phenomena, the cure of individuals and the ending of epidemics, can be reproduced experimentally, and this is the best way of confirming the conclusions I have just mentioned. In fact we can grow these powerful bacterio-

phages isolated from convalescents in the laboratory and obtain as much of them as we wish. In the test-tube as in the patient it suffices to give them bacteria sensitive to their action for them to multiply at the expense of these latter.

But note that, as we have seen, each race of bacteriophages possesses its special characteristics: to produce in a patient a bacteriophagy powerful enough to destroy all the disease bacteria in a few hours, powerful races must be used to start with.

Phage therapy is an all-or-nothing therapy. If bacteriophage cultures of feeble or moderate virulence are used, the pathogenic bacteria will resist its action and cure will not result. Every bacteriophage therapy must, therefore, be the result of a selection which demands long researches. Experience has further shown me that one must use for their *in vitro* culture only recently isolated bacteria (excluding type collection strains) isolated at the start of the illness, avoiding the use of bacteria contaminated with chronic bacteriophages. These contaminants are always present in strains isolated from patients in whom a spontaneous bacteriophagy caused by bacteriophages of feeble activity has had time to start.

The first illness on which I tried bacteriophage therapy was avian typhoid. First I tried individual treatment: all the hens treated in the first 12 hours survived. Then I tried prophylactic treatment, mixing bacteriophage cultures active against salmonella with drinking water in poultry yards where the epidemic was beginning: hens already infected were cured, and the epidemic extinguished. I sponsored similar trials by various vets. in several districts, and their results were similar to mine.

Bacteriophage treatment spreads

I then made tests on patients suffering from toxic dysentery (Shiga). This was at l'Hôpital des Enfants-Malades under Professor Hutinel. Depending on the case, between 24 and 26 hours after the administration of 2 c.c.s. of shiga-phage

by mouth, blood disappeared from the stools and the patient began his convalescence. These results were published in 1921. Various bacteriologists in Germany, in the United States, and in Brazil, tried to verify them. All declared that their results were completely negative. What had happened? Brazilian experience will teach us.

Da Costa Cruz of the Instituto Oswaldo Cruz of Rio de Janeiro published in 1922 a statement that the administration of bacteriophages to these dysentery patients had not had the slightest action. But this author went on studying bacteriophages out of academic interest, and he learnt little by little something of their various properties; in particular, that virulence varies from one bacteriophage to another. In 1923 he re-tested their use in treatment, this time using fairly virulent bacteriophages, and he then got results similar to mine.

Little by little the treatment of bacillary dysentery became common. Eliava in Russia, Compton in Egypt and Morrison in India have used it to throttle epidemics. During the last war its use was general in the German and Japanese armies, and also, more especially, in Russia, where it has been systematically employed in civil medicine as well as in the army; and where nine institutes are occupied in the study and preparation of diverse therapeutic bacteriophages.

I mentioned that I had studied bubonic plague. I had made several preliminary tests in Annam in 1920, but it was in Egypt in 1926 that I really tested the treatment. Applied according to the rules, within six hours from the injection of a single cubic centimetre of plague-phage into the bubo the temperature dropped to normal, the general symptoms disappeared and convalescence began, a rapid convalescence because the buboes had not become infected. Following publication of these results experiments were made in various countries: in some, results were negative; in others, they were similar to mine. During the last war in Russia treatment of plague by bacteriophage was applied as a routine in the south where the malady is endemic.

Then it was the turn of Asiatic cholera during an epidemic in the Punjab in 1927. As a first experiment I treated 74 patients, all gravely ill; of this number 26, treated with 4 c.cs. phage by mouth in the first six hours after they had passed the first stool characteristic of the illness, all entered convalescence in the following 24 hours; 41 were treated between 6 and 24 hours after the start and the mortality amongst them was 10 per cent; and finally seven received the bacteriophage later and the mortality was 14 per cent. The overall mortality for the treated was 8 per cent. In 124 controls untreated or treated by other methods the mortality rose to 63 per cent, which was the general mortality rate elsewhere in the course of this epidemic.

As a result of these researches two laboratories for the preparation of therapeutic bacteriophage were built in India, one in Patna under the direction of Prof. Asheshov, the other in the Pasteur Institute of Shillong in Assam under the direction of Colonel Morrison of the Indian Medical Service. The first applied prophylaxis on a large scale by the addition of bacteriophage culture to the wells in every village where cases of cholera appeared, with the result that while in villages where the usual prophylactic measures were in force (addition of potassium permanganate to the wells and vaccination of the population with anti-cholera vaccine) the mean duration of epidemics was 27 days, which is the same as the mean duration in villages where cholera is allowed its head, in several hundred villages where the bacteriophage was applied without any other measure the mean duration was 48 hours.

In Assam, Morrison applied this method of treatment in all cases from the start. For this purpose a store of ampoules was built up in each district where it was decided to apply the method. He has thus succeeded in keeping districts with over a million inhabitants free from any cholera epidemic, and the mortality from cholera has been reduced to a low figure, whilst in neighbouring districts acting as controls

the epidemics have continued, as in the past, to claim hundreds of victims each year. Since that time the treatment and prophylaxis of cholera by bacteriophages has spread to China and Japan.

Between bacteriophage and bacteria is ceaseless struggle

I have described only diseases with which I myself have been principally occupied. Many other researches have been made along the lines that I have explored. Gratia has foreshadowed the treatment of staphylococcus and streptococcus infection, and Raiga has tested it out. Out of more than 6,000 cases he has been able to avoid the necessity for surgery in 97 per cent ... Hauduroy has made the first tests on typhoid and in urinary infections, a method of treatment which has spread. Tsulukidze at Tiflis treats peritonitis, following intestinal perforation, with therapeutic bacteriophage. In the most severe cases, all of which occurred in the course of typhoid, and which are practically always fatal, he has lowered the mortality by 20 per cent.

But it would take me hours to present the full picture of the ceaseless struggle which goes on in nature between bacteriophage and bacteria. I would have to write of the role of bacteriophage in the natural purification of river water, in the bacterial illnesses of plants; of fields of lucerne which perish when bacteriophage breaks out amongst the bacteria in the nodules which fix atmospheric nitrogen. I would have to describe how certain industrial fermentations languished when bacteriophage attacked the bacteria producing them, and inversely how bacteriophage is used to prevent harmful fermentations which interfere with the industrial production of certain higher alcohols. I would have to tell you of the researches of Mazé of the Institut Pasteur, who has shown how bacteriophages, after they have become adapted to the destruction of lactic ferments in pigs' intestines, are carried into cheese factories by the hands of the workers and also, no doubt, by the atmosphere, and interfere with the ripening of cheese.

In nature every time that bacteria do something, a bacteriophage interferes and destroys the bacteria, or provokes a modification of their action. Thus the aim of microbiologists is to study these actions and reactions, and to select the bacteriophages which attack harmful bacteria in order to bring about their rapid destruction wherever this is useful.

Translated by J. L. Crammer

The Nature of Bacteriophage

RALPH W. G. WYCKOFF

BACTERIOPHAGES are the viruses of the bacterial world. Like all viruses they have been grown only in association with the living cells that are their hosts and like all but the biggest viruses their elementary particles are too small to be seen with optical microscopes.

When bacteriophages were discovered a generation ago by Twort and by d'Herelle it was hoped they could be used to control and cure many of the human ills caused by the bacteria on which they feed. Unfortunately, these expectations did not materialize, though d'Herelle and others insisted, and have continued to insist, that, when properly employed, bacteriophages can have great efficacy in the control of epidemics of certain infectious diseases. Because they did not fulfil the claims of their most ardent advocates, however, they rather quickly became discredited and for years were scarcely more than one of the curiosities of bacteriology. In recent years they have emerged from this oblivion, not indeed as a practical tool of the medical art, but as one of the most effective instruments for probing into the fundamental nature of the viruses of which they are a representative.

There are several reasons why bacteriophages are well adapted to such investigation. For one thing their bacterial hosts can be grown in quantity with only a small expenditure of time and money, and on them comparatively large quantities of bacteriophage can be produced. This ease of preparation, combined with the short life span of the bacterial hosts, means that a hundred experiments can be made with bacteriophage while one is being carried through on most plant and animal viruses. It is another advantage

THE NATURE OF BACTERIOPHAGE

61

of bacteriophage that many different kinds with different properties are found in nature. They attack many sorts of bacteria and have sizes that run almost the full gamut between objects nearly big enough to be seen under the optical microscope and those not much larger than the very small viruses of foot-and-mouth disease and infantile paralysis.

Though the renewed interest in bacteriophage is an aspect of the greater attention given in recent years to viruses in general, it can lead to results of wider importance. One of the chief problems of biology to-day, as it was a couple of generations ago, is the search for an adequate scientific understanding and definition of the term 'living'. In the intervening years we have learned many of the attributes of living matter and have come to understand many of the chemical mechanisms it utilizes, but the very fact that we still use the term more or less intuitively shows that the kernel of its meaning continues to escape us.

Viruses are important to this question of the nature of the living process because some are endowed with essential attributes of living matter, and either are amongst the simplest of living things or are part of a bridge stretching unbroken from the animate to the inanimate, if indeed such a bridge exists. There is an increase in complexity of chemical composition on passing from the world of macromolecules to that of independently functioning microorganisms; and it has become the habit to conclude that along with this goes a transition in properties whereby the larger and chemically more complex objects in this scientific Limbo are more fully alive than the smaller and simpler ones. This may be the case, but the determination of whether it is so or not is an experimental problem that has not yet been satisfactorily answered. It can, however, be approached with a unique directness through the many experiments that can easily be made with bacteriophages.

Recent work with the bacterial viruses has been devoted mainly to a few questions. These deal (1) with their chemical composition and properties, (2) with the shape and size of

their elementary particles, (3) with the way they are produced from the bacteria that are their hosts, and (4) with the genetics of the many mutants that appear as they grow and multiply. All these lines of research bear in most direct fashion on the central problem of their fundamental nature. Many of the facts brought out by the chemical and morphological studies favour the view that bacteriophages increase by a sort of self-reproduction resembling that manifested by bacteria themselves. The many new facts turned up by genetic experimentation have, on the other hand, usually been interpreted in terms of hypotheses that consider new bacteriophage particles to arise more directly from bacterial protoplasm. Additional experimentation is needed to resolve these seeming conflicts, but research on bacteriophages is made especially attractive at the present moment by the fact that the electron microscope is a tool capable of resolving this conflict through direct observation.

With the greatly extended vision of this instrument the elementary particles of bacteriophages appear as highly organized sperm-like entities with heads enclosed in definite membranes and tails of different lengths and thicknesses. Most work has been done on the many strains of bacteriophage that attack colon bacilli. These coli bacteriophages fall into one of three groups as seen under the electron microscope. Those designated as T2, T4 and T6 have ellipsoidal heads and the short, stiff tails shown in Plate 10. The sizes of their heads (60×80 millimicra)* are such that, packed side by side, it would require at least fifty to be as long as a bacterium and half a million to make an inch. These three types of bacteriophage have more in common than their shape, and differ from the odd-numbered bacteriophages which have spherical heads and long, thin curly tails (Plate 11). The small-headed pair, T3 and T7, are very closely related; the other two, T1 and T5, have larger spherical heads and are less alike. The characteristic shapes of the bacteriophages that have been

* One millimicron = $1/1000$ micron = $1/1,000,000$ millimeter.

recognized under the electron microscope make them easier to work with than viruses which, like that of poliomyelitis, have spherical particles not readily distinguished from the many particles of similar size and shape present in healthy tissues.

Determination of the chemical composition of bacteriophages, as of that of other viruses, is dependent on our ability to isolate them in a pure and unaltered state, and in amounts sufficient for analytical procedures. Fortunately, these requirements are met by several of the coli bacteriophages. The results thus far obtained by Beard and his collaborators indicate that they are extraordinary in their large content of nucleic acid.

There are, it will be remembered, two kinds of nucleic acid found in nature, desoxyribose nucleic acid, originally obtained from the thymus gland, and ribose nucleic acid first prepared from yeast. Some viruses contain one, some the other; none has as yet been proved to contain them both. The virus of influenza seems to possess as little as 2 per cent nucleic acid, and most have considerably less than 10 per cent. Bacteriophage, however, is more than 40 per cent desoxyribose nucleic acid. Small amounts of the ribose acid have also been found in the analyzed samples, but this is probably from bacterial protoplasm which has not been completely eliminated.

Another important constituent of the bacteriophage, besides protein and carbohydrate, is 2 per cent lipid that is all neutral fat. It has been especially instructive to compare this composition with that of the bacteria upon which the bacteriophage grows. These bacteria contain very little (2-5 per cent) desoxyribose nucleic acid but about 20 per cent of the ribose-type acid; no neutral fat, but about 8 per cent phospholipid. These inversions in composition are strong evidence that bacteriophage as it multiplies draws neither its fat nor its nucleic acid from its host, but instead can utilize directly constituents of the medium in which the bacteria grow.

Experiments by Cohen with a radioactively labelled supply of phosphorus seem to confirm this for the nucleic acid. In his experiments, when bacteria made radioactive by being grown in a medium containing radiophosphorus were transferred to a non-active medium before infection with bacteriophage, the bacteriophage that was produced was not radioactive; whereas newly-formed bacteriophage was radioactive when inactive bacteria were lysed in a medium containing radioactive phosphate. On the face of it this result indicates that the bacteriophage independently synthesizes this major constituent of itself directly from inorganic phosphate drawn from the medium, and not by transformation of its host's supply of the element.

It is hard to describe briefly or to analyze the results of the genetic experiments that have been made. The most significant are based on the fact that whereas odd- and even-numbered coli bacteriophages cannot grow together in the same bacterial cell, this interference, as it is called, does not operate to prevent the simultaneous infection of a bacterium with pairs of even-numbered bacteriophages such as T2 and T4. Mutant strains occur frequently in all bacteriophage cultures, but they appear to be especially numerous and of different sorts in cultures involving these mixed infections. Common mutants differ from parent strains in the appearance of the cleared areas, or plaques, they produce among bacteria growing on agar plates; other mutants are able to lyse different strains of colon bacilli, which thus can be employed to recognize them.

In one experiment which shows a change in the properties of bacteriophage arising from mixed infection, it was found that the number of clear plaques obtained by plating the yield from such an infection with a mixture of bacteria susceptible to either strain of bacteriophage exceeded the number obtained by plating with bacteria susceptible to one strain only.

Another experiment pointing to an exchange of genetic material between strains infecting a single bacterium can be

carried out using for example T2 and T4r (a mutant of T4 having the characteristic *r* which gives quick lysis); the new particles produced by this infection will include T4 as well as T4r. An especially striking experiment suggesting genetic exchange to eliminate lethal factors has recently been carried out by Luria. He found that related strains, each of which had previously been irradiated with ultra-violet light till unable to multiply, 'recovered' the ability to grow and reproduce when mixed.

Though experiments of this type point to the wealth of new knowledge about bacteriophages that results from a genetic approach, the phenomena are so complex that it is not easy to reach unequivocal interpretations of the facts themselves. The magnitude of these complexities is suggested by the work of Anderson, who has found that traces of the amino-acid tryptophane or related substances are required to bring about the adsorption of T4 and T6 to sensitive bacteria, and that this effect of tryptophane can itself be neutralized by the indole which is produced as the bacteria grow.

It has not proved possible to 'revitalize' bacteriophage after x-irradiation in experiments analogous to those of Luria with ultra-violet light, but the careful analysis of survival curves of bacteriophage irradiated with x-rays in the interval between mixture with sensitive bacteria and the first rise in bacteriophage titre has been shown by Latarjet to give new information about the way multiplication takes place. During this interval the apparent sensitivity of bacteriophage to the radiation changes, and about midway through the period the shape of the curve of dosage-against-survival changes in a way related to the rate of bacteriophage multiplication. The results obtained have been thought to show that the new bacteriophage arises through a very rapid differentiation of the bacterial protoplasm.

Much of the quantitative experimentation with bacteriophage now being carried out makes use of the device of Delbruck and Luria for observing a so-called burst time

and burst size. If appropriate dilutions of bacteriophage and bacteria are incubated together, and titrations made of the number of bacteriophage particles present in samples taken every few minutes after mixing, there is no increase for an initial period, that is 13 minutes for T3 and 21 minutes for T2. At the end of this interval, which is the burst time, the number of determinable particles rises very rapidly to a value that is often many times the initial count and that remains fairly constant if enough bacteriophage was used in the first place to infect all the bacteria present. If the percentage of bacteriophage particles that have been adsorbed to susceptible bacteria is determined and if proper dilutions are used, then the magnitude of the increase will measure the average yield of bacteriophage per infection; this is the burst size.

The form of these curves is obviously determined by the growth of bacteriophage. It is generally considered that bacteriophage becomes adsorbed to susceptible bacteria when mixed with them, and that then, either remaining on the surface or penetrating within the cell, induces the generation of new bacteriophage particles within the bacterial protoplasm. At the conclusion of this multiplication, whatever its mechanism, the bacteriophage-filled bacteria are supposed to burst and liberate the newly-formed particles into the medium. The electron microscope makes it possible to check the general correctness of this interpretation and to examine the steps through which bacteriophage is produced. The first step of adsorption is readily verified by photographs that show the approach of bacteriophage particles and their attachment to the bacterial membrane (Plate 11). What happens next, however, is more complicated and depends very much on the age, or rate of metabolic activity, of the attacked bacteria. If the bacteriophage is T2 or T3, and if the bacilli belong to an especially young and actively-growing culture such as is obtained by making several preliminary transplants to fresh medium at hourly intervals, many ruptured cells are found in preparations

from mixtures which have been incubated for not more than five minutes before being chilled and formalinized. Fluid protoplasm evidently flows from such broken bacteria and engulfs the invading particles (Plate 12).

Especially dark objects that are larger and more numerous the longer the incubation within the burst time appear in these protoplasmic masses. The amount of enveloping protoplasm diminishes as these masses increase, and after the longer incubations they often stand out clearly as clusters or microcolonies of newly-formed bacteriophage (Plate 13). What is thus seen with very actively growing bacteria suggests that lysis, in the sense of cell rupture, may take place before the proliferation of bacteriophage is complete. On this basis the rapid rise in measurable bacteriophage that occurs at the end of the burst time would be due not so much to rupture of the bacterial cell wall as to the escape of the newly-formed bacteriophage from the depleted mass of protoplasmic jelly which has enclosed and nourished it. These results and many observations on bacteriophage-infected colon bacilli growing on agar plates indicate that bacteriophage may not require an intact bacterial cell for its growth. They increase the difficulty of maintaining the hypothesis that bacteriophage arises through a sudden differentiation of modified bacterial protoplasm. The hundreds of electron micrographs already made of infected bacteria show the steps to be expected if new bacteriophage actually develops by a process of self-reproduction and division, but there is need for much more study of this multiplication.

It is harder to be sure of the interpretation to be given to what is seen after the infection of older bacterial cultures. As with young organisms the cell walls readily rupture, but the protoplasm they enclose is so viscous that it retains the shapes of intact bacilli. Bacteriophage seems to grow by penetrating and partially consuming this mass so that towards the end of the period of multiplication there are found many masses of new bacteriophage

particles which have still the outlines of the original cells (Plate 14).

An explanation of this different behaviour of young and old cells is to be found in the visible fine structure of their protoplasm. The protoplasm that has flowed from the young bacillus of Plate 12 consists almost entirely of minute spherical particles. These spheres are present in older cells but in them they are intermixed with filaments that seem to increase in amount with the age of the culture. It is reasonable that the greater viscosity of the protoplasm of older bacteria should be due to the development of this fibrous skeleton. The spheres gradually disappear from an infected cell as bacteriophage multiplies, but not so the filaments, and these are especially apparent in the residues of older lysed cells from which the new bacteriophage has more or less completely escaped (Plate 15). The implications of this fundamental change in the structure of bacterial protoplasm with age go far beyond the problems being discussed here.

The possibility of bacteriophage's multiplication elsewhere than in an intact bacterium suggests many new fields of investigation. One deals with attempts to observe with the electron microscope bacteriophage multiplication in non-viable bacteria. Recently we have been successful in seeing such multiplication from bacteria rendered incapable of further growth (inactivated) by treatment with formaldehyde. The interpretation of what is seen in these experiments has an evident bearing on what is meant by the term 'dead.'

All the lines of investigation outlined above are so far from having attained their full development that we may be sure their further pursuit will result in a steadily deepening understanding of the life history of bacteriophage and thereby of the viruses it typifies.

Viruses and the Veterinary Surgeon

G. P. WEST

Foot and Mouth

FOOT-AND-MOUTH disease is not ranked among the killing diseases. In countries where slaughter is not compulsory, affected animals usually recover, though with the so-called malignant form – such as was experienced in Italy and spread over Europe in 1918 – the mortality may be as high as 50 per cent. Recovery is often not complete, however, and the fever results in wasting and a reduced milk yield. Pigs, sheep and goats also become infected, and, as can be imagined, even if there are not many deaths, any outbreak is bound to affect milk and meat production in the area very seriously. Animals which have recovered are not immune for long, so that the disease, once introduced, is liable to flare up again.

Isolation is not practicable, for foot-and-mouth disease is one of the most contagious known. Quite apart from the movement by road or rail of infected animals, the virus responsible for the disease is easily conveyed by people on their clothes, by lorries and cars, birds and rats. Any attempt to confine the disease to a single farm usually meets with failure, and the risk of the disease spreading increases as the number of infected farms rises. Thus the whole of France was involved within a few weeks in 1937, and by the end of that year 63,000 outbreaks had occurred in Belgium. In the following year Germany was involved, and the loss there amounted to the equivalent of £80 million. In the United Kingdom, round which the sea acts as a partly, if not entirely, effective barrier, and where invasions

by the virus can be stamped out, the slaughter policy is the most economic.

There are, unfortunately, not less than three types of the foot-and-mouth disease virus, and Sir Daniel Cabot, when Chief Veterinary Officer to the Ministry of Agriculture, stressed the fact that the presence of viruses of two different types in one locality had frequently been encountered. This means that a vaccine which produces or confers immunity against only one type of virus may be useless in preventing an outbreak of foot-and-mouth disease. Even if a vaccine is prepared which immunises against two types of the virus, infection with a third is an ever-present danger.

The practical difficulties in the way of producing a vaccine from all three types of virus have proved virtually insuperable, apart from the huge expense. Germany started producing vaccine which would give protection against two types of the virus, but soon had to give this up and revert to a vaccine useful against only one type. One of the difficulties has been that the virus cannot be cultured, as bacteria can be, on such media as meat broth; it will grow only in living tissues. Although vaccine is produced in Denmark at a cost of 3s. 6d. per dose, it has been stated that it would take about three months to make enough for all the herds in the country, and would involve the sacrifice of 500 head of cattle per week. This would mean decreasing the Danes' small meat ration by a further 10 per cent merely in order to produce a quantity of vaccine sufficient to cope with a severe epidemic. And the resulting immunity lasts only a little over six months.

It may be wondered how, despite our quarantine regulations, outbreaks of foot-and-mouth disease originate in these islands. Migrating birds have been blamed, but the most common method of introduction is undoubtedly through imported chilled or frozen meat in which the virus can remain alive for several months. Pieces of infected meat or bacon rind sooner or later find their way into pigs' swill. In 1944, 78 out of 93 outbreaks began among

pigs, and in practically every case (and contrary to law) swill had been fed raw on the premises.

Illness in Dogs

Of the diseases of dogs, Canine Distemper and 'Hard Pad Disease' are, from the scientific point of view, two of the most interesting. It is just over twenty years ago that Laidlaw and Dunkin, with funds provided by the *Field*, completed their research into the cause of distemper – a disease somewhat similar to influenza in human subjects – and isolated the causative virus. Subsequent work by Mr Thomas Dalling, the present Chief Veterinary Officer to the Ministry of Agriculture, enabled a method of immunisation to be perfected; or rather, two methods.

In one case a dose of vaccine, containing the killed virus, is followed at an appropriate interval by the injection of liquid containing the living virus. With the other method, the dose of living virus is followed by a dose of anti-distemper serum. In both cases the result is the setting up of what is called active immunity, and after a certain time the inoculated dog can safely mix with infected dogs without becoming ill. Passive immunity, of short duration, may be conferred by the administration of anti-distemper serum only, and this is a useful method of protecting dogs at shows, etc.

In naturally occurring cases of distemper, the effects of the virus which causes the initial illness are soon followed by those of 'secondary invaders' – bacteria which take advantage of the animal's lowered resistance and cause such complications as bronchitis, pneumonia and enteritis. It is against these secondary invaders, incidentally, that some of the vaccines on the market act. They will not prevent distemper, only the complications; the use of one of the methods described above affording the only true preventative available in this country.

The distemper virus is not characteristically 'neurotropic.' That is to say, unlike the virus of rabies, it is not specially

addicted to causing damage to nervous tissue. Nevertheless, in a certain proportion of cases – mainly puppies 3 or 4 months old, or dogs approaching middle age – the nervous system is damaged, and the dog dies, following a series of convulsions. This offers an interesting point of comparison with 'Hard Pad Disease.'

The latter differs from distemper in several respects. Apart from the fact that six-day-old puppies have been known to suffer from 'Hard Pad', the typical distemper picture of dry nose and thick nasal discharge is replaced by that of a cold, excessively wet nose. Sometimes, but not invariably, a peculiar hardening of the pads of the feet takes place – in some cases so marked that the dog makes a tapping sound as he walks across the floor. Convulsions may soon begin to occur – a symptom of encephalitis (inflammation of the brain) and the disease end fatally.

Is this a new disease? Some think so, for 1944 was the first year in which the Wellcome Research Laboratories detected significant changes in their microscopical findings in pathological material sent to them by veterinary surgeons in practice. And to-day, in some areas, a typical case of distemper is rarely or never seen, having apparently been replaced by 'Hard Pad.' Immunologically, the two diseases are distinct. The virus isolated from cases of 'Hard Pad' will not set up distemper, and a dog resistant to distemper, either following recovery from a natural infection or as a result of inoculation, is not immune from 'Hard Pad', though more resistant to it than a dog which has never had distemper. The hardening of the pads has been set up experimentally in ferrets, but never the encephalitis; while dogs artificially injected with the 'Hard Pad' virus have usually developed a transient fever lasting a fortnight and then made a complete recovery; only rarely has hardening of the pads been reproduced, and never the encephalitis.

These findings indicate that some factor – in addition to the virus – is necessary to set up the encephalitis, and the identification of that unknown factor is the task of current

research. What of the two viruses? They have been described as 'not unrelated', and many believe that the Hard Pad virus is really the distemper virus which has undergone mutation to become a separate entity.

The Origin of the Solar System

SIR HAROLD SPENCER JONES, F.R.S.

THE solar system, as we know it to-day, consists of the Sun as the parent body, nine large planets, more than 1,500 minor planets or asteroids, 30 satellites belonging to six of the planets, and a large number of comets. It contains, in addition, a considerable amount of meteoric matter, circling around the Sun under the binding spell of its gravitational attraction, and of which we have evidence in the form of shooting stars and meteorites. The system shows a number of regularities of arrangement. The planets all revolve round the Sun in the same direction and the planes in which they move are inclined to one another at rather small angles. The satellites of the planets, with the exception of a few of the smallest, revolve around their primaries in this same direction. The Sun and the planets rotate on their axes in this sense also, while their equatorial planes are generally inclined at small angles to the orbital planes. The orbits, both of the planets and of their satellites, have small eccentricities.

There is, moreover, a certain regularity in the arrangement of the planets. With the exception of the two outermost planets, the mean distances are in close agreement with a curious empirical law discovered by Bode in 1772. At that time there appeared, according to this law, to be a missing planet between Mars and Jupiter; the discovery of the asteroids, which may perhaps be the fragments of a disrupted planet, filled the gap. Comparable laws for the distances of the satellites from their primaries have been

THE ORIGIN OF THE SOLAR SYSTEM 75

found. The planets can be divided into two groups, with distinct characteristics. The inner group, Mercury, Venus, the Earth, and Mars, have rather small masses, high densities, low speeds of rotation and few satellites. The outer group, Jupiter, Saturn, Uranus, and Neptune have large masses, low densities, rather fast rotations, and many satellites. Pluto, which is anomalous and about which not a great deal is as yet known, must be left out of consideration.

These conspicuous regularities cannot have come about by chance. The probability, for instance, of a common direction of orbital motion of so many bodies having been brought about by chance is infinitesimally small. The system must have had a definite origin. Many hypotheses have been suggested to explain how the solar system came into existence. The problem has proved to be one of the most difficult in cosmology and there is still no theory that is completely free from objection.

The various theories can be divided into two groups. The first group attempts to account for the solar system as the result of a natural process of evolution from a primordial mass of gas or nebula. The second group assumes that there has been an interaction with some other celestial body, causing catastrophic change instead of slow evolution.

The early hypotheses, which preceded Newton's enunciation of the law of gravitation, such as the vortex theory of Descartes, can be left out of consideration as not worth serious discussion. The oldest hypothesis which is deserving of mention was suggested independently by the theologian, Swedenborg, and the philosopher, Kant. They supposed that the Sun was originally at the centre of a nebula, which was in rotation round the Sun under its gravitational attraction. The mutual collisions between the separate particles of the nebula caused it to flatten out into a disk. The matter in this disk gradually condensed round the denser aggregations, giving rise to a number of local nebulae, each in rotation. It was supposed that smaller condensations

gradually formed in each of these, which gave rise either to a planet or to a planet with a system of satellites. Kant supposed that the larger the planet, the greater would be its gravitational attraction and the larger, therefore, would be the number of its satellites. The rotations of the planets would be in the same sense as their revolution round the Sun, while their orbital planes would be approximately in the plane of the Sun's equator. Thus the main regularities in the solar system seemed to be accounted for.

A hypothesis, which is in some respects analogous, was suggested by the great mathematician, Laplace, in 1796, in his popular book *Exposition du Système du Monde*. Though unaware of Kant's theory, Laplace also supposed that the Sun was originally at the centre of a nebula, which was assumed to be hot and slowly rotating. The nebula cooled gradually, condensing in the process. As it contracted, its rotation necessarily became more rapid, in order to conserve the total angular momentum.* When the centrifugal force at the equator became greater than the gravitational pull, a ring of matter split off round the equator. The main mass continued to contract, and it was assumed that several rings were thrown off at intervals and in succession. Laplace supposed that each ring then condensed into a smaller nebula, though he did not explain how this was brought about, and that in some of these a similar process was repeated, thus giving rise eventually to a system of planets, some of which had systems of satellites. The central mass eventually condensed to form the Sun. This hypothesis, like that of Kant, accounted qualitatively for the main regularities in the solar system. Laplace put it forward merely as a suggestion, explaining that it did not merit the same confidence as a theory which is capable of mathematical verification.

But there is a fatal objection to both these hypotheses, which is concerned with the distribution of the angular

* Angular momentum is the product of the mass of a body and the rate of its rotation.

momentum in the system. In a system which is not acted upon by outside forces, the total amount of angular momentum must remain unaltered, even though frictional or viscous forces may have caused some of the energy to be degraded into heat. The distribution of angular momentum, as between one part of the system and another, may, of course, change. In the solar system, almost the whole of the angular momentum (98 per cent) is in the orbital motions of the major planets, though they contain less than one-seventh of one per cent of the total mass; the Sun, which has practically all the mass, has only 2 per cent of the angular momentum. The puzzle is not why the planets have so much angular momentum, for any bodies revolving in the same orbits would have the same amount of angular momentum per ton, but why the Sun has so little. If the solar system had been formed as Kant or Laplace supposed, we should expect to find the angular momentum distributed in a manner similar to its distribution in the satellite systems of Jupiter and Saturn, where the rotation of the planet provides by far the greater part, the portion contributed by the orbital revolutions of the satellites being small. If all the mass and angular momentum of the solar system were concentrated in the Sun, the period of rotation would be about 12 hours, instead of about 27 days. There would be no risk of breaking up in the way Laplace supposed, for the centrifugal force at the equator would be only about 5 per cent of the force of gravity.

If, however, the system has in the past been acted upon by outside forces, so that it has not in fact always been isolated, there will have been changes in its angular momentum. Various attempts have been made to devise a workable scheme by which the planets could have been formed and endowed with their large angular momentum, as the result of the interaction of an external body. In the planetesimal hypothesis, proposed by Chamberlain and Moulton about 1900, it was supposed that at some remote epoch another star, in its motion through space, chanced

to pass very near the Sun – within a distance of a few diameters. The two bodies, mutually attracted by their gravitational pull, swung round each other in hyperbolic orbits and separated again. Near the closest approach an enormous tide was produced on the side of the Sun facing the star, which, aided by the expansion of the hot compressed solar gases, became so great that a succession of huge eruptions occurred, accompanied by the ejection of matter. As the star passed by, the ejected matter was pulled after it by its attraction and caused to move in eccentric orbits round the Sun. It was supposed that this matter would condense into innumerable liquid drops, which would quickly solidify; they would be moving in the same direction round the Sun and more or less in the same plane. The denser portions formed the nuclei of the planets which, as they circled round the Sun, progressively grew by accretion, sweeping up, by their gravitational pull, the planetesimals in their vicinities. The residual matter formed a resisting medium which would gradually reduce the eccentricities of the orbits to small values. The satellites are supposed to have grown from smaller nuclei close to the large planetary nuclei.

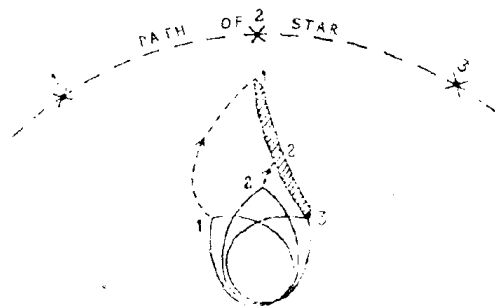


Fig. 6. – Ejection of filament of matter from the Sun by a passing star. 1, 2, 3 denote successive positions of the star, and the corresponding positions of the point of ejection from the tidally distorted Sun. The diagram is drawn for the instant when the star is in position 3.

The objection was raised by Jeffreys that the collisions between the small bodies would be so frequent and would occur with such high velocities that the particles would be volatilised long before accretion could have any important effect. A modified form of the theory was developed by Jeans and Jeffreys. They supposed that the Sun was much more distended than it is now and that the passage of the star caused a continuous stream of matter, in the form of an elongated filament, to have been drawn out from the Sun (Fig. 6). This filament would be unstable and would break up, under its gravitation, into a number of separate fragments which would move in elliptical orbits round the Sun. These fragments condensed into planets and Jeans pointed out that, as the filament would be thickest near its middle, the largest planets would be expected to be found in the middle range of planetary distances, as is the case. While still only partially condensed some of the fragments might, under suitable conditions, be broken up on their first return to the vicinity of the Sun, so producing the satellites. As in the planetesimal theory, the resistance of the residual matter, which is supposed to have been gaseous, reduced the eccentricities and inclinations of the orbits of the planets. Some of this matter would escape from the system; some would fall back on the Sun and give it a rotation in the same direction as the orbital motion. Similarly, the planets were supposed to have been set into rotation by matter, which had acquired momentum from the star, falling back upon them.

Though this theory appears plausible, it will not stand detailed examination. The rapid rotation of the major planets is difficult to explain. When Jeffreys investigated this question quantitatively he found that they would need to have reabsorbed far more matter than is comprised in all their satellites, which does not seem possible. Jeans showed that the idea that the satellites were formed from the primitive planets by the tidal action of the Sun would not work. The distribution of angular momentum still remains

a difficulty. Though the passing star would have imparted angular momentum to the filament, Russell showed that, if sufficient angular momentum was transferred, large inner planets and small outer planets would be expected.

Jeffreys accordingly modified the theory by supposing that, instead of a near approach of the Sun and the star, there had been an actual collision – not a head-on collision, for that would not produce rotation, nor a grazing one, for that would not remove enough matter. He supposed the collision to have been about midway between these two extremes, a ‘side-swiping’ collision, as Russell termed it. The two bodies, drawn by their mutual gravitation, would approach each other with enormous speeds. Great tides would be raised on them and matter would begin to be ejected before the actual collision occurred. During the collision the heavy cores of the two bodies would swing round each other in a sharp curve and, separating, move off into space. The nearer portions of the Sun and star would intermingle and become greatly compressed; the shearing motion would give rise to intense turbulence and produce rapid rotation. As the two bodies separated, a filament of hot gaseous matter, having a rapid whirling motion, would be drawn out between them. The subsequent course of events would be much as in the case of a near approach, but the rapid rotations of the major planets are accounted for. It is not possible to work out in detail the exact course of events, but by a rough estimation Jeffreys concluded that the total mass of the planets would be of the right order of magnitude.

But the theory is still open to several objections. It was assumed that at the time of the formation of the planets the Sun was much more distended than it now is. But from the present knowledge of stellar evolution it can be inferred that when the solar system was born, a few thousand million years ago, the Sun was almost in its present state. If we suppose that the colliding star was of about the same size and mass as the Sun, the angular momentum per ton of the

matter from which the planets were formed can roughly be estimated; on the most favourable suppositions, it proves to be less than one-tenth of the average angular momentum of the planetary system. The essence of the difficulty can be seen in this way: the matter from which Jupiter was formed was drawn out to a distance about one hundred times greater than the distance between the Sun and the star at their nearest approach; the matter from which Neptune was formed was drawn out to about 500 times this distance. In order to obtain sufficient angular momentum it is necessary to suppose that matter dragged out to these distances by the departing star was left moving in directions almost at right angles to the direction of motion of the star.

But this question of angular momentum is not the sole difficulty. There is another and even more serious one. Not far down in the Sun's atmosphere the pressure becomes greater than one million atmospheres and the temperature more than one million degrees. The collision would still further heat the matter which is drawn out into the filament. It has been shown by Spitzer that gravitational attraction would be inadequate to prevent the rapid dissipation of matter ejected under such high pressure and with such high temperature. The ejected material would, of course, rapidly cool; but by the time it had cooled, there would not have been enough left to form the planets.

It must therefore be concluded that no collision theory which supposes the planets to have been formed from the Sun can be made to work. To overcome the angular momentum difficulty, Russell suggested that the Sun might originally have been a binary star and that its companion, supposed to be a good deal smaller than itself, revolved around it at a distance comparable with the distances of the present major planets. By supposing that the intruding star collided, not with the Sun itself, but with its companion, the angular momentum difficulty can be avoided; for the angular momentum of the planets was not produced by the collision, it was already present in the angular momen-

tum of the orbital revolution of the companion. The collision merely distributed it differently amongst the surviving parts of the system. For such a theory to work it is, of course, necessary to devise some way by which that portion of the companion which was not left behind in the form of the planets was got rid of.

This problem has been investigated by Lyttleton. It is extremely complicated and not susceptible of exact analysis. Rough estimation is all that is possible. He succeeded in showing that, under certain conditions, it would be possible for both the colliding star and the Sun's companion to escape from the neighbourhood of the Sun, while leaving a sufficient part of the ejected matter to revolve around the Sun under its gravitational attraction. But, though the angular momentum difficulty can thus be overcome, the theory must, in order to succeed, be able to satisfy two further requirements. The physical condition of the ejected matter must be such that condensation into planets is possible, and the differences in density and mass (and therefore also in chemical composition) between the inner and the outer planets must be explained.

Lyttleton has attempted to avoid some of the difficulties of accounting for the condensation of the present planets from the ejected material by supposing that they were formed by a two-stage process. He suggests that the first stage might have been the formation of two or more large bodies, each of sufficient size to hold together under their own gravitation. If such a body, in the course of cooling and contracting, acquires too rapid a rotation for stability, it will break up, not into a binary system nor into a planet and satellite, but into two bodies moving independently. Lyttleton has shown that the mass-ratio of the two bodies so formed must be at least six to one. As the mass of Jupiter is less than six times that of Saturn, it is necessary to suppose that there must originally have been at least two bodies formed in the first stage. He further suggests that, when instability set in, there might not have been a clean

division into two separate bodies, but that secondary disturbances may have developed, so that small portions may have continued with each of the main bodies to form their satellites, while others may have become independent of them and have produced the terrestrial planets and the Moon.

There is, however, the difficulty that the rate of rotation required for instability to occur in any of the contracting rotating masses is appreciably faster than that of any of the planets. To meet this difficulty Lyttleton argues that the rotation of the inner planets would have been slowed somewhat rapidly by the friction of the tides raised on them by the Sun, and that the major planets would have gradually grown in size and mass by drawing in much of the surrounding uncondensed material through which they moved; their rotation would, of course, have been gradually slowed down at the same time, while their mean densities would have gradually decreased by the accretion of material which consisted largely of the lighter elements.

Other theories of a catastrophic nature have been proposed which do not assume external disturbance from an outside star. The important part which fission may have played caused Lyttleton to suppose that the Sun was originally the principal component of a triple system, whose other two components were initially close together. Their masses slowly increased by accretion of interstellar matter until eventually the coalescence of the two companions was brought about. The angular momentum of the combined body was then too great for stability, so that the system split up and separated into two bodies moving independently. Under suitable circumstances both of these bodies could escape from the system, leaving merely a fragment, part of the 'splash' which accompanied the cataclysm, as a captive of the Sun. He supposes that the planets were formed from this fragment.

Hoyle has attributed the formation of the solar system to the outburst of a supernova. He supposes that the Sun was once a member of a binary system and that its com-

panion developed into a supernova. If the material thrown off in the course of the outburst were thrown off symmetrically, the two stars would continue to form a binary system. But the evidence of the Crab nebula, which is believed to have been formed as the result of a supernova outburst in the year 1054 A.D. indicates that the emission may have been asymmetrical. An asymmetry of only a few per cent would be sufficient to break up the binary system, if the separation before the outburst was comparable with the distance of Jupiter from the Sun. Much of the material would be shot off with a velocity exceeding 1000 km/sec. and would be moving much too fast to be captured by the Sun. But, as the outburst died away, the speed of ejection would have progressively decreased and sufficient material may have been captured by the Sun to have formed the planets. The amount required is less than 1/10,000th of the total mass ejected at the outburst of 1054 A.D. As the angular momentum per ton of this material would be comparable with that of material moving in a circular orbit round the Sun at the distance of Jupiter, the angular momentum requirements can be satisfied. The direct rotation of the planets is explained as in other catastrophic theories.

None of these theories is entirely satisfactory and, in particular, they are all open to Spitzer's objection that extremely hot material from the interior of a star would have dissipated away into space before it could begin to condense. Some new attempts have consequently been made to account for the solar system by the gradual evolution of a primordial system. Two of these theories are worth mentioning.

An interesting theory has been developed by Alfvén, which is based upon the existence of a general magnetic field of the Sun. It assumes that electromagnetic forces have played a prominent part in the formation of the solar system. Alfvén points out that the force exerted by the Sun's magnetic field on an electrically charged particle can be very much greater than the Sun's gravitational pull.

Suppose, for instance, that a proton is moving along the earth's orbit with the same velocity as the earth: the electromagnetic field arising from the Sun's magnetic field (assumed to be about 50 gauss) is some 60,000 times greater than the gravitational force; even in Pluto's orbit, it is 250 times greater. Alfvén supposes that the Sun, in the course of its motion through space, entered a nebula or cloud of interstellar gas and was surrounded by it. The nebula is assumed to consist initially of neutral atoms. The gravitational attraction of the Sun will cause the atoms to fall towards the Sun with increasing speed, so that the whole cloud in the vicinity of the Sun becomes heated up. When the energy of the atoms becomes equal to the energy required to ionize them, Alfvén supposes that they will become ionized, so that there will then be a mixture of negatively charged electrons and positively charged ions. When this occurs, the whole system of forces is transformed because the electromagnetic forces now predominate.

The magnetic field of the Sun induces electric currents in the ionized cloud, and an effect similar to that of the damping of a metal plate between the poles of a magnet takes place. The cloud is accelerated in the direction of its rotation at the expense of the retardation of the Sun. Angular momentum is transferred from the Sun to the cloud. Alfvén calculates that about 10 per cent of the momentum would be transferred to the cloud in some 100,000 years; he finds also that the braking effect is greater in the Sun's higher latitudes and it is, in fact, well-established that the Sun rotates less rapidly near its poles than near its equator.

Once an atom is ionized, its movement towards the Sun is stopped. The positively charged ion is compelled to move along the magnetic lines of force until it reaches an equilibrium position in the equatorial plane of the Sun. With reasonable assumptions about the atomic weight and the ionization potential of the atoms, Alfvén is able to calculate the distribution of mass in the equatorial disk, and to

show that there will be a maximum concentration at a distance which is about equal to the distance of Jupiter from the Sun. The mass distribution is found to agree roughly with the mass distribution for the four major planets. He supposes that, having reached the equatorial plane, the ions and electrons will recombine and that condensation then takes place, first planetesimals and then planets being formed.

It seems reasonable to suppose that, if the planets were formed in this way, they would themselves be magnetized, and that the primary process would operate again on a reduced scale on any matter left over, so producing satellites. The critical distance at which this would occur for Jupiter is found to be of the same order as the distance of its four major satellites. In the case of Saturn the critical distance proves to be less than Roche's limit, expressing the least distance at which a satellite can exist; so no condensation would take place and the ring system of Saturn is thus accounted for. For Uranus and Neptune, the critical distances are of the same order as the radii of the planets themselves, so that no satellites are to be expected. The primary process, as imagined by Alfvén, can thus account for the massive outer planets, for the inner satellites of Jupiter and Saturn, and for the rings of Saturn.

For the terrestrial planets and for the outer satellites of the major planets a somewhat different process is suggested. The plausible assumption is made that the nebula consisted not merely of gas but also of solid particles resembling meteoric dust. Alfvén supposes that the dust particles, because of their mass, would be able to penetrate the magnetic field much further than the atomic particles. They come sufficiently near to the Sun to be vaporized; the resulting gas is then ionized, and the electrically charged particles are repelled along the innermost lines of force. Some of the matter condenses to form the terrestrial planets, while a portion reaches the outer planets and, according to Alfvén, gives rise to their outer satellites.

The theory appears at first sight to be able to explain many points of detail in the structure of the solar system and to provide rough quantitative estimates of the distribution of mass between the planets. The difficulty about the distribution of angular momentum is avoided by supposing that the Sun rotated much more rapidly when it entered the cloud than it does at the present time. But the theory is based fundamentally on the existence of the general magnetic field of the Sun; recent observational results, obtained since Alfvén formulated it, throw serious doubt upon the reality of the Sun's magnetic field. Moreover, a detailed investigation by Ter Haar seems to prove that ionization could not occur in the way Alfvén supposed. No satisfactory explanation of how the equatorial disk condenses into separate planets is given. Interesting as the theory is, it does not seem that the solar system can be accounted for in the way Alfvén has supposed.

A different theory, depending upon hydrodynamical considerations, which is, in effect, a modification of Kant's theory, has recently been advanced by von Weizsäcker. It is supposed that the Sun, before it possessed any planets, passed through a relatively dense region of the interstellar cloud. This local cloud is assumed to have had the same chemical composition as the Sun, consisting predominantly of the light elements, hydrogen and helium. The Sun, in its passage through the cloud, collected an envelope, consisting of an aggregation of atoms and dust particles, all moving round it in more or less independent orbits under its gravitational attraction, the planes of the orbits being distributed in a more or less haphazard manner. The mass of this envelope is assumed to have been about one-tenth of the Sun's mass.

Internal frictions within the envelope will change the shape and orientation of the orbits of the various particles so that they are eventually reduced to orbits of nearly circular shape in the vicinity of the Sun's equatorial plane, resulting in a disk whose diameter is comparable to the

present diameter of the solar system and whose thickness is less than one-hundredth of its diameter. The temperature of the material at any distance from the Sun is determined by the Sun's radiation and is therefore about the same as that of a planet at the same distance.

At this stage in the evolution of the disk, the particles are still moving in practically independent gravitational orbits, so that the parts near the Sun have greater angular velocity than those which are more distant. Viscous forces now come into play, whose effect is to tend to equalise the angular velocities, slowing down the faster moving inner parts and speeding up the slower moving outer parts of the system. This process brings about a gradual transfer of the angular momentum from the inside to the outside. The cumulative effect is the gradual dispersion of the envelope into interstellar space, the portions with higher than average angular momentum, including the greater part of the lighter elements, being lost, so that there is left a slowly rotating central mass, surrounded by a faster rotating gaseous cloud, whose rotational velocities are ultimately determined by the central mass and follow Kepler's third law. In this way, Weizsäcker accounts for the present distribution of angular momentum in the solar system.

But meanwhile another process is supposed to occur, which forms the most interesting and novel part of the theory. It is argued that turbulence would set in and that particles of equal mean motion would collect into vortices. In order to provide a type of motion in which the vortices obstruct each other as little as possible, it is suggested that they would arrange themselves into a succession of rings, giving a system of great stability. The most stable arrangement is obtained when there is an integral number of vortices in each ring. Weizsäcker gives reasons for assuming that there were five vortices in each ring. With such an arrangement of vortices, the ratio of the radii of two consecutive rings is constant, so that the radii follow a law similar to Bode's law connecting the distances of the planets

from the Sun. In the neighbourhood of the circles between successive rings of vortices there will be steep velocity gradients, so that large viscous stresses will be produced and turbulence will develop. Secondary eddies, which may be regarded as in the nature of roller-bearings, will form on the circles separating the main vortices, the rotation in the eddies being in the opposite direction to that in the vortices.

Weizsäcker shows that the conditions for condensation will be more favourable in the roller-bearing eddies than in the vortices. The argument depends upon the mean free paths of the different particles. For the larger particles, the mean free path proves to be greater than the size of the roller-bearings, so that there will be a range of particle sizes which cannot be carried along by the roller-bearings but which can be carried along by the vortices. Condensation will therefore start in these roller-bearings; nuclei will first be formed, which will grow by accretion, as impinging atoms or particles stick to them, and then, in the later stages, growth will occur by gravitational attraction. Weizsäcker further supposes that the eddies in each ring will eventually come together and form a single planet. He is very vague as to the way in which this comes about, the mechanism not being explained (Fig. 7.)

During their formation and for some time after it, each planet will be surrounded by an extensive atmosphere. It is assumed that a somewhat analogous sequence of events will follow in these atmospheres and that satellite systems will be formed, but no detailed discussion of this process is given.

Some of the details of the process of planet formation on the basis of this theory have been discussed by Ter Haar with very interesting results. For condensation nuclei to form it is necessary that the vapour pressure of large particles should be less than the pressure in the gas, so that more atoms will condense on a particle than will evaporate from it. The processes involved are analogous to those involved in the formation of raindrops. He shows that the

nature of the particles that condense is determined mainly by the temperature. There would be practically no ionization of the matter by the solar radiation, consequently the distribution of temperature in the cloud will be controlled by the Sun's radiation; the temperature will decrease rapidly outwards from the Sun. In the outer regions compounds like water, ammonia, carbon dioxide, etc. can condense, but in the inner regions only metals and other inorganic compounds, which are heavier and less abundant. The first stage of the condensation, therefore, leads to heavy bodies in the inner regions and lighter bodies in the outer.

The second stage in the condensation is growth by impinging particles sticking to the nuclei, and the final stage is growth by gravitational capture. This third stage is important mainly in the case of the outer planets. Thus the inner condensations begin by having higher densities than the outer; gravitational capture, mainly of the

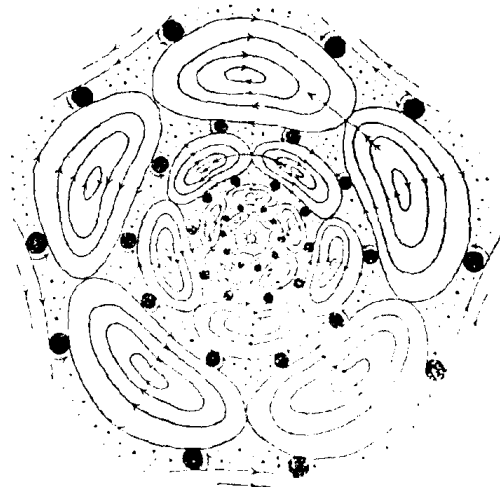


Fig. 7. - Arrangement of vortices and roller-bearing eddies as pictured by Weizsäcker.

lighter elements, by the outer condensations, accentuates the difference.

Assuming that only one planet is formed on each circle, Ter Haar shows that the condensation process provides a mass distribution in the system which agrees as well with the observational data as can be expected from general considerations. The only serious disagreements are that the mass of Mars is much smaller than the theory requires, and that there is no planet between Mars and Jupiter. He suggests that some catastrophe may have occurred in that neighbourhood, and that it produced the asteroids from a previously existing planet.

Ter Haar remarks that the outer planets would have grown much more rapidly than the inner planets; the growth of the inner planets would have stopped before that of the outer planets, the cessation of growth being determined in each case by the dissipation of the envelope. Weizsäcker estimates that the density in the envelope would have been reduced to the density of the matter in the vicinity of the Sun in about 200 million years.

The properties of the satellite systems are then considered. The satellites can be divided into two groups: the first five satellites of Jupiter, the first eight of Saturn, four of Uranus, and one satellite of Neptune comprise the first group, with orbits of small eccentricity, all approximately in the equatorial plane of the primary. These are termed the 'regular' satellites. The remainder, including the Moon, the two satellites of Mars, and the other satellites of the outer planets are termed the 'irregular' satellites. It is suggested that the regular satellites were formed within their planetary atmospheres by condensation, and that the irregular satellites were condensation products, which were captured by the planets at a later stage. It may be mentioned that the distances of the regular satellites from their primaries follow laws analogous to Bode's law. The satellites being small, gravitational capture did not play any role in their building-up process; it would, therefore, be expected

that their densities should be higher than the densities of their primaries, but lower than the densities of the inner planets, which agrees with observation.

There are so many analogous features in the satellite systems and the planetary system, including a similarity in the ratio of the mass of the primary body to that of the secondaries, that a question of some interest is why the outer planets have been left with a fairly rapid rotation, while the Sun has a slow rotation. The slowness of rotation of the inner planets can be ascribed to tidal friction, caused by tides raised by the Sun. In the case of the outer planets, the large vortices would in the beginning have supplied angular momentum to their atmospheres, and the angular velocities in these atmospheres would correspond to Kepler's third law; the rotational velocities of the planets would have been given by the same law. When the planetary disks started to disintegrate, the transfer of angular momentum caused the planetary rotations to slow down. The present rotational speeds correspond to a loss of about 70 per cent of the angular momentum, so that the outer planets were, in fact, appreciably slowed down before their envelopes were dissipated.

The theory of Weizsäcker, as modified and amplified by Ter Haar, has some notable successes, but it is not without its difficulties. Although the transfer of angular momentum will slow down the solar rotation, the process, on detailed investigation, seems scarcely capable of accounting for the present slow rotation, so that the distribution of angular momentum in the solar system still remains somewhat of a difficulty. Nor is any explanation given of how the roller bearing vortices in each ring collect together to form a planet; Weizsäcker admits that the process is difficult to visualise.

Our survey of the various theories of the origin of the solar system leads to the conclusion that there is as yet no theory which can explain satisfactorily all the features of the system. The most recent theory, that of Weizsäcker, has

had the most striking successes and, being more amenable to mathematical treatment than any catastrophic theory, there is a hope that the outstanding difficulties may be removed by further investigation. If the solar system was actually formed in the way he supposes, planetary systems should be of frequent occurrence, particularly in the Milky Way, where the interstellar clouds are abundant. The one restriction is that the temperature of the central star must be below a certain value for condensation to be possible, so that none of the very hot stars should have planetary systems. If the solar system was formed as the result of catastrophic action, arising from the close passage of another star, planetary systems must be of extreme rarity. The chance that during the lifetime of the earth any star, taken at random, has collided with another star is less than one in a hundred million. If observational evidence could be obtained of the frequency of planetary systems we should at least be able to exclude theories of one type or the other. But of that there appears to be little hope at present.

Molecular Architecture in Plants

R. D. PRESTON

JUST as we cannot hope to understand the growth of a crystal without a good deal of information concerning its molecular structure, so we cannot attempt any understanding of the growth of plants without the most profound knowledge of the organisation of the protoplasm, the living stuff of which plants and animals are composed. It is with this particular aspect of botanical research in mind that the following pages have been written. Not that we shall be considering directly protoplasm itself; but, by studying the dead (or scarcely living) cell wall which surrounds plant cells, we shall see that something can be added to our store of knowledge of protoplasmic structure. At the same time we shall reveal some of the entrancing beauty of plant life, going down to the very smallest limits of size.

During the past few years many lines of physiological research in botany have been concerned, directly or indirectly, with the structure of protoplasm. Thus the permeability of protoplasm to dissolved substances (solutes), organic and inorganic, the electrical properties of plant tissues, the surface tension of naked cells, the dissection of cells with micromanipulators, as well as the more direct evidence resulting from observation in polarised light, determinations of viscosity, etc., have all given information, if somewhat meagre, upon this point.

It has for a long time been recognised that the main framework of protoplasm is protein in nature, and, now that so many protein molecules are recognised as long thin

MOLECULAR ARCHITECTURE IN PLANTS 95

threads, it is generally considered that a network of polypeptide chains, probably folded into special configurations, confers upon protoplasm its unique properties. In fact, however, our ideas on protoplasmic structure owe more to argument from the known details of the fibrous proteins as presented in the main by Professor W. T. Astbury, than to precise investigation of protoplasm itself. It is, of course, peculiarly difficult to attempt structural determinations on protoplasm, if only because any interference with it leads immediately to the loss of the special organisation associated with life. Any pointers towards its structure are, therefore, very welcome.

Now plant cells, unlike animal cells, clothe themselves in a sheath or envelope consisting of a number of polysaccharides and their derivatives, among which the most important is cellulose. This is a fact of which use can be made. Cellulose, though chemically very different from the fibrous proteins, consists like them of long molecular chains (Fig. 8), but the links are derived from glucose and not from the amino acids. Further, these chains, which may be $10,000\text{\AA}$ * long, or more, lie parallel to one another for some part of their length and spaced regularly the same distance apart, forming small 'crystals' (or regions of special order) whose borders merge into neighbouring regions where the chains do not show this pronounced regularity. These 'crystals' correspond to the 'micelles' originally proposed by Nägeli as the basis of structure and, though they can no longer be thought of as completely separate entities as they were by Nägeli and even later workers down to the most recent times, it is convenient to retain Nägeli's term, if only for reasons of tradition. The regions of random chain orientation may then be called the 'intermicellar spaces.' These two terms will be used with these meanings throughout the following pages. Such, then, is the framework of the cell wall in plants, and it is within the intermicellar system that there is imbedded the bulk of the other substances in the

* $1\text{\AA} = 1/100,000,000\text{ cm.}$

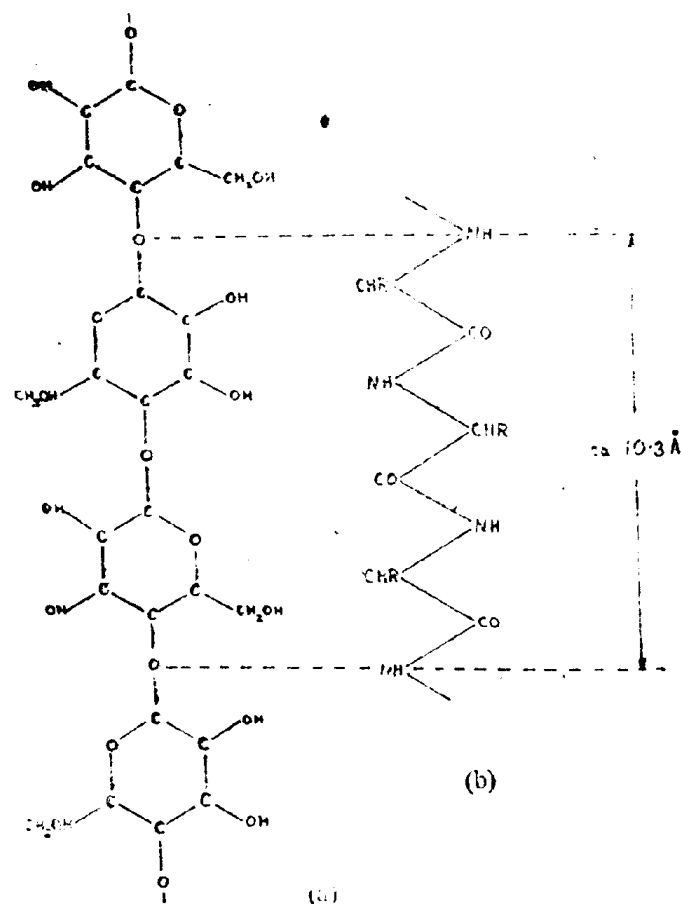


Fig. 8. - Diagrammatic representation of part of the structure of cellulose and of a protein. C=carbon, H=hydrogen, O=oxygen, N=nitrogen, R=one of the organic radicles of the amino acids.

(a) *Cellulose*. Four glucose residues are shown, each $C_6H_{10}O_5$ (i.e. glucose, $C_6H_{12}O_6$, minus one molecule of water) linked into a chain

all - pectin, hemicelluloses, lignin, etc. - with which we shall not be concerned here.

One further feature of cellulose which is very striking in terms of the structure of the fibrous proteins is this. If we pass from any chosen point on a chain and travel along the chain, then we come to a precisely similar point at a distance of 10.3 Å, having passed through two glucose residues. This is exactly the length occupied by three amino acids in the fully extended polypeptide chains in protein (see Fig. 8), and a similar coincidence is found with the folded protein chains of hair, etc. This resemblance may, of course, be quite fortuitous, but it is just possible that it gives us a pointer which we should not ignore. For the crystalline regions of cellulose, within which the cellulose chains lie strictly parallel, themselves also lie more or less parallel to one other and, as we shall see, this common direction of orientation bears some quite specific relation to the size of the cell, to its rate of growth, or to both. It seems very probable that this very special organisation of the chains is not accidental but results from the vital activity of the cell; and how can this happen, except through the agency of the proteins in the protoplasmic surface lying between the protoplasm and the cell wall? In other words, the study of chain orientation in naturally occurring cellulose very possibly yields information concerning proteins in the underlying protoplasmic surface. This is an exciting possibility since it is this very surface which controls the rate of

with hundreds of these residues. The total length of two residues in the chain is 10.3 Å, as shown in the figure, and the breadth of the chain in the plane of the paper is of the order of 5 Å. It should be noted that the molecular chain is, in fact, not planar as this simple diagram suggests; alternate -OH groups lie first above then below the plane of the paper, but the chain is nevertheless wider in the plane of the paper than it is thick at right angles to this plane.

(b) *Protein*. Three amino acid residues are shown, each -CO-NH-CHR-, the total length being about 10.2 Å. Note, as for glucose, that the molecular chain is not planar; the radicles R stand alternately above and below the plane of the paper.

entry or exit of minerals, sugar, water, etc., to the cell, and there is at present no other way of studying it directly.

There are, of course, many other reasons why it is important to determine the exact details of structure in cellulose, for instance its economic importance, and a great deal of work has been done. It will be possible in this article to mention only a small fraction of the research, choosing particularly whatever may throw light on protoplasmic structure.

Most cells are of microscopic size and therefore not easily handled or put before the mind's eye of a non-microscopist. Let us start with one of the few really big ones. The seaweed *Valonia* is a native of the warmer seas; one particular species - *V. ventricosa* - which is found off the Bermudas, takes the form of single cells ranging in size up to a pigeon's egg or even larger (Plate 17a), and each cell is anchored to the rocks at one end by a few filaments which grow out there. These cells or vesicles have all the structures of the cells of more normal size. The outer membrane, the cell wall, is comparatively thin (a fraction of a millimetre) and covers a thin layer of living protoplasm. The rest of the space in the cell is occupied by a solution of salts, sugars, proteins, etc., and is not usually regarded as part of the living system.

The cell is so large that the wall can be cut into individual pieces with a pair of scissors and these pieces examined with ease in the X-ray spectrometer or under the microscope. It turns out to have a most fascinating structure. From the X-ray diagram (Plate 17 b) it has been deduced* that in any one piece of wall the long thin chains of cellulose lie along either of two directions, almost at right angles to each other, or occasionally in a third direction, but in no

* Space will not allow any attempt here to explain the interpretation of X-ray diagrams. Those interested should consult the standard texts (e.g. Bunn, C. W.: *Chemical Crystallography*, Oxford, 1946). A simplified account will be presented shortly by the writer (*Molecular Architecture of Plant Cell Walls*, Frontiers of Science Series, Chapman and Hall).

others (cp. Plate 18). Now the wall is not homogeneous in thickness microscopically; it consists of many layers (as many as forty have been counted), one over the other like leaves in a book though, of course, much thinner; and it would be easy to conclude that each layer is distinguishable because each is composed of chains all lying in one direction, the difference between the layers being a difference in chain direction. This, however, is not so. It is a most difficult point to establish in *Valonia* just what factors are involved in this layering, but in other algae, which show a similar structure, it is very clear that the layers are alternately cellulose-rich, pectin-poor; and cellulose-poor, pectin-rich. Thus both sets of cellulose chains occur in one and the same cellulose-rich layer, and the lamellae within this layer, each of which has only the one set of cellulose chains, must be incredibly thin.

This attempt precisely to establish the nature of the lamellae may at first sight seem like splitting hairs; but it should be clear on reflection that it is most important to know how thick the individual lamellae are in which there is only one set of chains. For consider what is happening in these algae. Within each lamella the cellulose chains lie parallel to one another and these chains, or at any rate their constituents as they are produced by the protoplasmic surface, might become oriented in a particular direction by the chains already laid down in crystalline order, by a sort of crystallisation process, just as new material on a growing crystal is oriented by the atoms and molecules already arranged on the crystal surface. This may be precisely what does happen. A time comes, however, when, for some reason, a halt is called to this process; and thereafter the chains which are laid down become oriented at an angle to those already present. The fact that this angle is not constant rules out any possibility of 'twinning' such as occurs in some crystals, e.g. quartz crystals, so that we have here a very peculiar phenomenon indeed. The most queer thing about the whole business is this, however: when a

halt is again called to this process which we have likened to a process of crystallisation, then the next set of cellulose chains are not just laid down at any angle to those in this last lamella; they are laid down exactly parallel to the chains in the last layer but one, in spite of the fact that this particular layer is now much too far removed from the protoplasmic surface to have any orienting influence on new material being deposited. This is repeated probably hundreds and possibly thousands of times throughout the wall thickness.

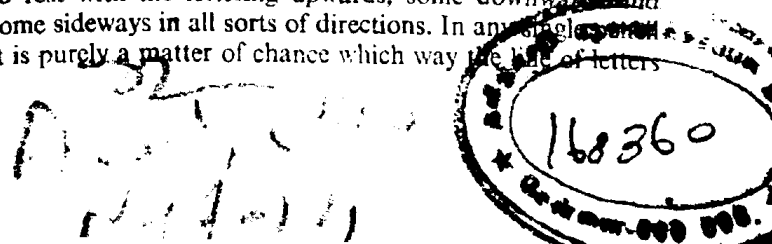
It must be remembered that this whole picture was deduced from diagrams such as Plate 17b, that is, by theoretical interpretation of a diffraction pattern and not by mere visual examination of a pictorial photograph. It was very exciting, therefore, when it became possible with the advent of the electron microscope to see structure in such walls – not yet, it must be confessed, in such detail as to confirm completely all the deductions from X-ray diagrams. An electron micrograph of the surface of the *Valonia* wall is illustrated in a beautiful photograph in Plate 18. It will be realised that the threads visible in the photograph, some 300 Å thick, are much coarser than the molecular threads (some 5 Å) and even than the 'micelles' (some 50 Å) and that these latter structures are not seen. Nevertheless the threads made visible in this way must be built up of cellulose chains arranged precisely in the manner deduced in the X-ray technique. A similar structure has since been revealed in other related seaweeds, and in some fresh-water species, and the search for an explanation of this organisation is a matter of first-class importance.

It is difficult to see how such beautiful regularity of structure could possibly arise except through the agency of a mechanism sited in the surface of the protoplasm where it abuts on the cell wall. If this is so, then the switch in chain direction from point to point in the wall thickness implies, since the different wall lamellae are laid down one

upon the other by the protoplasm, some corresponding periodical switch in the properties of the protoplasmic surface. It should be possible to make the relevant observations to test this hypothesis if the algae concerned could be grown in the laboratory. This has proved difficult in a city laboratory, particularly because the plants have to withstand a long period of transport in very adverse conditions and therefore arrive looking very sorry for themselves! In the last few weeks, however, thanks to careful packing and to air transport, plants have arrived which are now showing signs of growth, and we hope for the best.

Meanwhile some progress has been made with other algae which grow in the temperate zone and can therefore be obtained round the coasts of Britain. In some ways these are less satisfactory to work with than *Valonia*, but they can be cultured fairly easily and at the moment it seems very clear, from the observations made over the past eighteen months, that changes in the environment are associated with the switches in wall structure and, therefore, presumably with putative changes in the cytoplasm; but until this has been fully confirmed it is perhaps best to leave it at that for now.

Let us return for a moment to the electron micrograph of *Valonia* cell walls, for there are some further peculiarities which we have not yet noticed. First, what do we imagine to be the cross-sectional shape of the threads visible there; are they round like wires or flat like ribbons? This is a matter of some interest in view of some deductions which can be made from the X-ray diagram and which we have not yet discussed. We know that the threads must have the same 'faces' pointing the same way. We can form a mental picture of the meaning of this by considering a number of pencils with, as is usual, lettering down one edge only. If we throw these on to a table then some will come to rest with the lettering upwards, some downwards and some sideways in all sorts of directions. In any one pile of pencils it is purely a matter of chance which way the face of letters



faces. We can arrange them, if we like, with all the lines of letters facing upwards, but this is clearly a very special arrangement. This is, however, precisely the condition existing in the threads in the *Valonia* wall. The threads have no 'lettering', of course, by which we can actually see this arrangement, but their internal crystalline structure is such that certain molecular planes lie parallel to the surface. In fact, if we think of the cellulose chain shown in Fig. 8 as a thin ribbon, then this ribbon lies in the wall with its length, of course, in the surface but the plane of the ribbon tilted at an angle of about 45° to the wall surface – and every single chain tends to take up this position. Now this would easily be understandable if the threads visible in the electron-micrograph were themselves ribbons much thinner at right angles to the wall surface than parallel to it. But what if, as one might at first sight suppose, they are actually rod-shaped? As yet we do not know what shape they are and the next step is obviously to take a look at these threads sideways.

In *Valonia* then, we have a very special and beautiful structure whose existence we can hardly doubt is due to the vital activity of the protoplasmic surface and whose explanation must yield information of real importance to an understanding of this activity. Let us just notice one thing in passing. If a (comparatively) huge wall surface such as this were formed of cellulose chains lying in one direction only then it would tend to split as the cells proceeded to grow. By laying down two sets of chains at right angles the cell is, however, clothed in an envelope which is mechanically strong in all directions and the possibility of splitting, with its fatal results, is therefore avoided. The cells are clothed by a strong 'fabric', using a principle developed by man many, many thousands of years after nature had first used it, for precisely the same purpose in the weaving of cloth: with the difference that man, with his clumsy tools, has found it necessary to interweave the fibres in the separate lamellae.

This leads us to the second of the two peculiarities mentioned above. We must remember that this alga, surrounded by a mechanically strong coating, is nevertheless growing. As yet no one knows a great deal about the growth of *Valonia* except that it is slow. But in the other algae with essentially the same structure, such as *Cladophora* which can be studied round the coasts of Britain, we do know a great deal about growth, and it is quite clear that the cell can get bigger and bigger – in the case of *Cladophora*, longer



Fig. 9. – Diagrammatic representation of one cell in a filament of *Cladophora*. The lines represent the run of the cellulose chains on the front face of the cell. One set of chains forms a flat spiral running round the cell, and the other a steep spiral of opposite sign.

and longer, for this is a filamentous alga (see Fig. 9) – not only in spite of what must be a very considerable barrier to expansion, but without effecting any change in the direction of the threads. This is a peculiarity which we shall meet again in a moment in other plants, and perhaps more forcibly since these other plants will be more familiar. In *Cladophora* one of the sets of threads is so arranged as to form flat spirals round and round the filament. This filamentous cell grows longer and longer, but the spiral is not pulled out steeper and steeper as it would be, for instance,

if we took hold of the two ends and pulled. The cells grow, therefore, in such a way that the structure of the walls remains uniform.

The same thing is true for the growing cells of land plants of a similar kind, and about whose growth we do know a great deal more. Suppose we remove the bark from a tree early in spring; we shall find it very easy to do this at this time of the year and we must all of us have done it at some time in our lives, and perhaps noticed the slippery feel of the new wood. At this time the living layer in the stem – the cambium – lying between the bark and the wood, is actively growing and it is its high water content in spring, giving it the consistency of a soft jelly, which enables the bark to be slipped off so easily. The slipperiness of the wood surface arises because this living layer remains on the surface of the wood. It is very easy to skim it off with a blunt knife or even with a spoon; in fact on one occasion in the spring of 1939, when my colleagues and I had access to two elm trees which were about to be felled, we collected in all several pounds of cambial tissues – sufficient to do a complete chemical analysis in triplicate, using macro-methods! The details of this analysis need not concern us here, although the fact that this growing tissue did contain cellulose is one we might just notice, particularly since there was some doubt at about that time whether very young growing cell walls did actually have any cellulose. One other significant feature of the analysis was that, when this cellulose was extracted from the dried and powdered cambium, the extract set into a sheet rather like Cellophane. This contrasts strongly with the cellulose of the secondary wall of the mature tissues which, under identical conditions, is recoverable only as a powder. It suggests some physical difference between the organisation of the cellulose in growing and in mature tissues, and this is a point we shall return to again.

Microscopically the strips of cambium taken from a tree can be seen to consist of innumerable long thin cells,

usually some 15 – 20 μ * broad and often about one hundred times as long, blobs of protoplasm surrounded by a delicate cell wall and held together by intercellular films of pectin. Under suitable conditions, and even in the fresh condition, these can be studied by the X-ray method, and it is thereby confirmed, if in fact confirmation is needed, that this tissue does contain cellulose. The cellulose chains can be deduced to lie nearly transversely to the length of the cell, forming, as it were, flat spirals round the cell whose windings, as shown by other methods, actually make an angle of some 10° to the transverse. As far as we can tell it does not matter how long these cells become – and each cell continues to grow in length for many years, particularly in coniferous trees – the chains of cellulose maintain their transverse orientation. This appears to be true of the majority of elongating cells, and we have just seen that a similar situation arises in some algae. There are, therefore, at least two things we want to know about these cells. Firstly, how does it come about that the chains of cellulose take up this almost transverse orientation and, secondly, granted an initial transverse orientation, how does it happen that this orientation remains transverse during elongation of the cell?

Taking the first point, there seems at present no settled body of opinion. Until a few years ago it was thought possible that the orientation was imposed by the protoplasm of a cell by a very simple mechanical means. In the production of filaments of artificial silk a solution of cellulose is squirted through a narrow orifice; as the chains approach and pass through this orifice, they tend to line up parallel to the lines of flow and therefore to each other, and it is understandable that they should do so. Now the bearing of this on our problem is this: the protoplasm of those cells which we can observe without disturbing them by cutting into surrounding tissues is in movement, commonly streaming round and round the cell. Under a microscope, the movement appears reasonably fast, but we must re-

*1 μ = 10⁻³ mm.

member that a microscope magnifies distances and therefore velocities and that, in fact, the true velocity is not more than a few millimetres a minute. Nevertheless it was thought

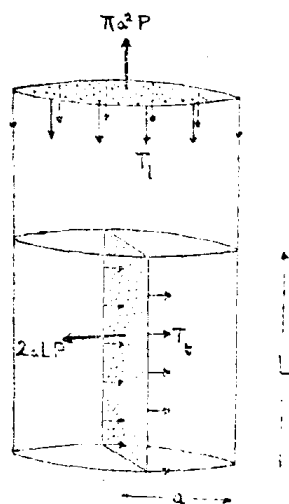


Fig. 10. - Diagrammatic representation of the tensions in a cylinder under internal pressure, P .

Upper part of figure. Consider a plane septum across the cylinder (stippled area). Since this must be in equilibrium the pressure over its surface must balance the tensions in the cylinder wall around its periphery. Hence

$$2\pi aT_1 = \pi a^2P$$

$$T_1 = Pa/2$$

Lower part of figure. Consider half the cylinder cut away and a plane septum inserted longitudinally over the cut face (stippled area). Since this must be in equilibrium, then

$$2aLP = 2LT_2$$

$$T_2 = Pa$$

Hence T_2 is twice as great as T_1 .

Note that this anisotropy of tension arises through anisotropy of shape. The more nearly a real cell approaches the spherical in shape the more nearly will T_2 be equal to T_1 .

possible that this movement might align the particles of cellulose before they were deposited in the wall.

In the cells with which we are most concerned streaming is never seen, though that might be due to the drastic things we have to do to the cells in order to see them at all. The theory has, however, not stood the test of time in other ways. No single case has ever been reported of a correspondence between direction of streaming and direction of orientation of cellulose chains, and until such a correspondence is shown the idea remains, of course, sheer poetry. Further, streaming mechanisms cannot hope to explain the complicated structures found in *Valonia*.

A more recent and equally elegant idea was put forward independently by two workers, Castle in America and Van Iterson in Holland. We can best grasp their suggestion again by an analogy. Consider a steel cylinder of compressed gas, and suppose that we were so unfortunate as to drop it in such a way that it burst. Now the experience of such events shows that the cylinder will not shatter nor blow off one end; it will usually split down the middle like a pea pod. The explanation of this is not far to seek and can be found in any good text-book on mechanics, the fact being that in any cylindrical object under pressure from within the tension in the wall of the cylinder across the cylinder is twice as great as that along the cylinder. Fig. 10 will make this clearer. Now, returning to our cells, this is precisely their condition. They are cylindrical, and they are under pressure from within, due to intake of water; and the pressures set up in them are by no means negligible, for they can amount to 10 atmospheres (150 lb. per square inch) or more. Hence here, too, there will be a transverse tension twice as great as the longitudinal one, and it is assumed that the chains of cellulose are aligned transversely for this reason.

There are, unfortunately, a good many things wrong with the application of this principle to cells, and it is mentioned here principally to show that no possible mechanical ex-

planation is being overlooked. We need not enumerate all the criticisms; it will be sufficient to ask ourselves how such an idea could possibly explain orientation in a cell like *Valonia*, which is not, nor ever was, even remotely cylindrical. It seems inevitable that we are being driven back for an explanation to an orienting mechanism of a molecular nature in the very surface of the protoplasm.

As for the maintenance of the orientation during growth, again there are some ideas, but precious few facts. It seems at the moment to be generally conceded that growth in plant cells occurs largely through the stretching effect on the cell of its internal hydrostatic pressure. If this were all, however, then we should expect the wall structure to change just as it does if the cell is pulled. There seems no fundamental difference between pushing from inside and pulling from outside. But we have already remarked that in growth the flat spiral of cellulose does not open out. There is surely something very fundamental that we are not taking into account in our simple picture, and this is where it becomes significant that the protoplasm and the wall in growing cells may really not be separate entities but may interpenetrate each other. It seems highly probable that in growing cells the wall is not a passive organ of the cell, but is itself actively growing and perhaps even controlling or limiting growth. There are now very many examples of growing cells where the growth processes do not in fact seem to put any strain on the walls.

A close association of this kind between the wall and the protoplasm of a growing cell seems all the more probable when we come to examine a living cell after it has ceased to expand. Consider, for instance, what happens to a cambial cell. These long, thin cells have one more peculiar habit not mentioned before. Like all other growing cells they increase in size (by increase in protoplasmic content) until at some stage each cell divides into two. Now we might expect that such a cell would divide transversely across the middle to produce two cells of the same width but of about half the

length, and this does happen with some cells. With the cambium of coniferous trees, however, which we can choose as our particular example on account of the relative simplicity of what takes place, this happens only infrequently. By far the greater number of divisions occur by splitting in two *longitudinally*, to produce two cells of about the same length but of only half the width in one direction (see Fig. 11). Further, the plane of division is always tangential to the stem, i.e., parallel to the stem surface, never radial. Of the two resulting cells, one of which, therefore, lies deeper in the stem than the other, one will undergo changes leading eventually to its death, while the other will continue to grow on as a cambial cell and undergo further similar divisions.

Let us follow the history of the deeper-lying cells which mature. They take in water rather rapidly and swell up radially (see Fig. 11) so as to increase in radial dimensions some 5 or 6 times. Again we might imagine that this rapid expansion would cause the chains of cellulose to become quite transverse, but again this does not happen. Once the cell has reached its full size, however, a new phenomenon, or an old one under a new manifestation, occurs. The cell proceeds to deposit from within, on the delicate primary wall, a new wall layer, or layers, so that the wall becomes very thick. The same kind of thing happens also in the development of cotton fibres and of hemp, jute, sisal, flax, bamboo and in fact all other fibres of commerce. Sometimes the cell wall remains rather thin, as in the coniferous tree during spring, or in sisal or kapok; sometimes it almost fills the cell as in some species of bamboo. In either case, the peculiar fact is that in these new wall layers, the so-called secondary wall, the chains of cellulose are no longer transverse. Commonly they lie much more nearly longitudinal and it seems to be quite a general rule that the longer the cell has become and the narrower it is, the more nearly do the chains lie longitudinally. This is true with a remarkable degree of accuracy.

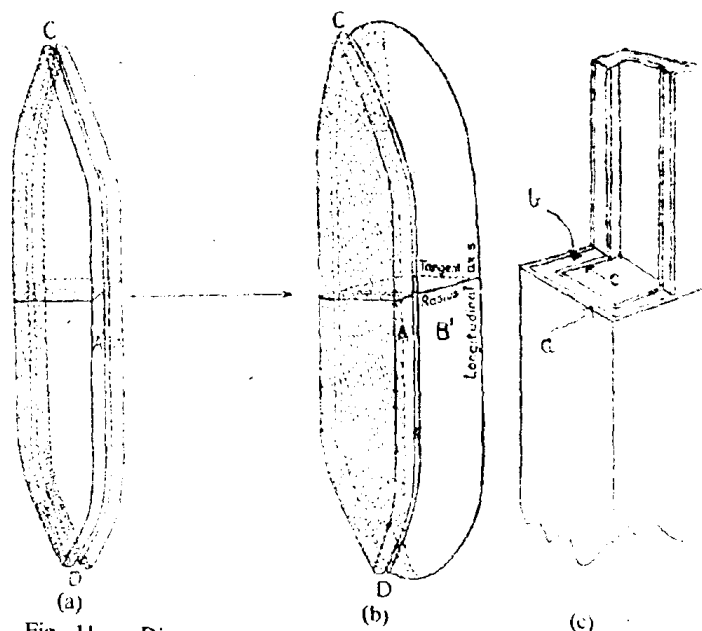


Fig. 11. - Diagrammatic representation of the development of a tracheid of conifer wood from the living cambium. All the cells are drawn much too short, for clearness of figure and highly magnified.

(a) A single cambial cell as seen from the outer surface of the cambium. The cell, CD, is shown just after one division to form two daughter cells A and B.

(b) The half cell A is now growing and regaining the size of the original cambial cell. The cell B has expanded radially to form a much broader cell B'. As drawn, this cell appears blunt-ended; it is, however, some twenty or thirty times longer, compared with its width. The ends in the real cells therefore are tapering.

(c) The cell B' now proceeds to lay down a thick secondary wall from within, to form a tracheid of the wood. This is in three layers: a is a thin outer layer, the first to be laid down; b follows, a much thicker layer, as a rule; then comes c which is again thin. In the two thin layers the molecular chains of cellulose lie in a fairly flat spiral, making an angle of some 50 degrees to the length of the cell, and the micelles which they form are not so perfectly aligned parallel to each other as they are in the thick layer, b. In this thick layer the chains lie almost longitudinal, the angle at which they lie depending on the dimensions of the cell.

Let us continue with the case of the cambium as one particular example. The cells which are cut off on the inside of the cambium are those which compose the wood with which we are all so familiar. Now, since the cambial cells are growing longer year by year and in each year are, in effect, cutting off replicas of themselves continually, i.e., forming the tracheids which constitute the wood, it follows that the tracheids become longer and longer as we pass outwards across the annual rings from the pith towards the bark. Now if we take a chip of wood from the various annual rings from any one level in a conifer stem we find that the X-ray diagram changes gradually as we pass outwards from the pith. The change is really very obvious and a good example is given in Plate 17 (c, d). The change shows that the spiral round the cell is becoming steeper, and the molecular spiral, therefore, here does become steeper as the cell elongates. It should be emphasised that these spirals are not visible microscopically.

Occasionally there appear on the wall markings called *striations* which lie parallel to the cellulose chains, and then we can see that the striations behave in exactly the same way. If we measure the average length, L , of the tracheids in any one annual ring and the average angle, θ , which the spiral winding makes with the length of the tracheids then we find the following relation to hold:-

$$L = K \cot \theta$$

where K is a constant. This was shown to be true a long time ago, in 1934 in fact, but the correct explanation has not yet been given. The equation corresponds to the condition of a spiral elongating at constant girth, but it must not be forgotten that while the cell was growing - as a cambial cell - its cellulose chains were in a flat spiral which remained flat. It was only after elongation had ceased that this new wall layer was deposited, and its chain direction bears no relation to that in the growing cell. Indeed, to make matters still more complicated, this secondary wall is not itself homogeneous in chain direction. It consists of three con-

centric layers, a narrow outer and a narrow inner one with a thick layer sandwiched in between (see Fig. 11).

The presence of these layers is clear in an ordinary microscope, but their real features are brought out only if we look at a very thin transverse section of the wood under a polarising microscope. To explain even superficially why this is so would take us much too far into the theory of light, but all we need in fact know is that, under the conditions usually used in the polarising microscope, any cellulose chains which are lying flat on the microscope stage will appear bright against the dark background and any which are lying at right angles to this, i.e., along the direction of propagation of the light, will appear dark; any which lie at some intermediate angle will appear more or less bright according to the angle. The photograph presented in Plate 19 will make it clear that, in the cells which we are considering, the cellulose chains of the outer and inner layers lie more or less in flat spirals (remembering that this is a cross section of the cells) and in the central layer more or less longitudinal.

Strictly speaking all the evidence we have been discussing until now refers to the central dark layer. The other layers with a different spiral angle nevertheless still follow a length relation of the same type, and this makes it quite certain that elongation of itself has no part to play in an explanation of this remarkable fact. To clinch this point once and for all, we know now that if we take any single tracheid from any annual ring in a conifer stem, and measure its length L and the spiral angle θ of its central layer, then a similar relation holds (Fig. 12) though the tracheids have not developed from the same cambial cell.

The same sort of thing is known to be true for bamboo fibres, for sisal and for cotton, and is probably quite general. Even when changes in cell length are induced experimentally, there is no breakdown in the general principle. Thus, if we apply clamps to the stem of a conifer during the growing season, then the tracheids grown below the clamps are much

shorter than they are in the neighbouring unclamped cells which have developed in the same year – and the molecular spiral is flatter!

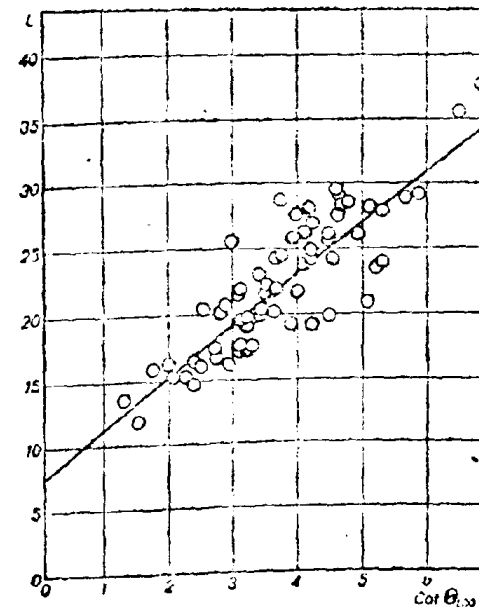


Fig. 12. – *Picea*. Correlation between tracheid length and the cotangent of the spiral angle corrected to standard breadth ($\theta_{1.00}$).

Although, therefore, the cellulose chains in these mature cells must be laid down by the protoplasm just as were the chains in the growing cell, the orientation is different and the change calls strongly to mind the remarkable structure in *Valonia*. The structure of the walls of the cells of these higher plants is different only in degree and not in kind from those of the lowly algae. In higher plants the switch seems to occur just at the time when the cells cease extension growth, and involves also one further difference. Whereas during length growth the direction of the cellulose chains

bears no obvious relation to cell length, after the adult size has been reached the orientation seems to be completely dependent on the dimensions attained.

Now we do know that, after cell enlargement is over, the connection between the protoplasm and the wall is a very loose one; for it can be severed merely by placing the cell in a strong sugar solution. This is not the case in growing cells. Since, therefore, in the adult cells where chain orientation is related to length, the protoplasm is separate from the wall, then in elongating cells, where there is no such relation, it may not be. Thus we have further support for the conception that the protoplasm-wall organisation may be quite different in growing and non-growing cells.

From this point of view it may also be significant that the cellulose micelles in the growing wall are much smaller than they are in an adult wall, as is readily deduced from the corresponding X-ray diagrams and is in harmony with the 'glassy' nature of the cellulose extracted from cambium. They are also much further apart, as we may deduce from the fact that the amount of cellulose in unit volume of the wall is much less than it is in the secondary wall. Even from the structure of the micelles themselves, therefore, without any reference to their orientation, we can see that the conditions obtaining during growth are very different from those during maturity. It may, therefore, be very significant that much of our knowledge of the physiology of cells, and indeed the bulk of the conceptions which we apply to growing cells, are in fact based on observations of adult cells.

It is easy to see why this should be so. In order to define conditions as precisely as possible in any experiment it is necessary that the tissues under examination should remain constant in kind and in quantity throughout the investigation; and it is very clear that such observations made under standard conditions must form the first step in any investigation. It is equally clear that, sooner or later, we must, in spite of the difficulties involved, make a concentrated

study of tissues which are actually growing. A beginning has in fact been made in several laboratories, and we await the results with much eagerness.

Meanwhile there can be no doubt that these beautifully regular patterns which modern physical techniques have thus revealed in the walls of plant cells result from some very special intrinsic property of the surface of the protoplasm. In the elongated cells of flowering plants it looks very much as though the protoplasm at any single point 'knows' not only how wide the cell is at that point but also how long it is. More precisely and less teleologically, we can say that the organisation of the protoplasm at any point in a cell is a function of cell dimensions, so that in some degree the protoplasm of a short cell must be different from that of a long cell. It is not possible to go further than this at the moment, for that would be to trespass into the realms of pure speculation; the practical applications of the principle revealed are very clear, but that is not in our sphere either. It is certain, however, that these studies eventually lead, directly or indirectly, to a better understanding of the activities of protoplasm and, therefore, of the growth of plant cells. Who can say what they might mean also in terms of animal growth?

BIBLIOGRAPHY

Some aspects will be found more fully discussed in the following:
Preston, R. D. (1938) *Ann. Bot.* N.S. III, 507.
Frey-Wyssling, A. (1948) *Submicroscopic Morphology of Protoplasm and its Derivatives*. Elsevier.

The whole subject of plant biophysics will shortly be reviewed by the author in a book published by Elsevier.

Correspondence

THE article 'The Moon and the Weather' by K. G. Collier in *Science News No. 11* wishes to introduce the layman to the *scientific method*. At its end, however, the paper also claims to have illustrated the *elements of research*. No doubt, the first object has been attained; the elements of research, on the other hand, embrace features not otherwise prevalent in scientific activities and these the paper seems to have left out.

The scientific method in general is a way of seeking an answer to a given question; scientific research, in contrast to this, has to find the question first. It approaches a vague complex of insufficiently understood phenomena and tries to extract some scientific questions, i.e., questions which can be examined by the general methods of scientific analysis. As long as the questions are put in a way not relevant to the unknown solution, the answers, scientifically as they may have been reached, will lead nowhere.

Let me make this clear in the present case. The problem is: has the moon any significant influence on the weather? In order to extract from this a scientific question, an 'analysis of the problem and definition of its ideas' was carried out. The result of this may be summarized thus:

1. The moon's influence on the weather is to be understood as a change of weather with a change of the moon.
2. A change of weather means a break into a fine or a wet spell of specified duration.
3. A fine spell is defined as one where high barometric pressure co-exists with long periods of sunshine.
4. A wet spell is defined as one where low barometric pressure is accompanied by precipitation, specified in detail.
5. The meaning of change of moon is defined in detail.

With these stipulations the scientific question could be shaped. The answer to it, which was found subsequently, by observation, statistics, theory of probability and logical deduction, will not be criticised here. But suppose that these five stipulations were quite arbitrary and could be replaced by an infinite number of others:

Example A - Replace assumption (1) by: The influence of the moon is to mean a change of weather at full moon or new moon, provided this change occurs between 0 a.m. and 6 a.m. local time.

By introducing thus the position of the sun, all investigations would have to be repeated.

CORRESPONDENCE

117

Example B - Maintain stipulation (1) and replace (2) by: A change of weather means the breaking up of a depression or of an anti-cyclone, even if these are not accompanied by a change from rain to sunshine or sunshine to rain.

Example C - Replace (2) by: A change of weather is to mean the change from dry and sunshine to precipitation and clouds or vice versa, even when these are not accompanied by a change in the barometric pressure.

Example D - The influence of the moon on the weather is to mean the tendency to cloud formation with or without rain in spite of a high barometric pressure, within 4 hours of the change of moon, provided this takes place between 5 p.m. and midnight local time.

These counter-proposals may not appeal to the pupil. They could serve, however, to show that the above stipulations were arbitrary and could be replaced by innumerable others. If, therefore, no definite influence of the moon on the weather could be established in the reported case, should the pupil really come to the conclusion that 'the moon has no influence on the weather' (p. 37)? Will this not frustrate further research on the problem? Should not, instead, the pupil conclude that the moon has not just that influence on the weather which was implied at the outset by the adopted stipulations? Is there any rational argument which will show that none of the possible alternative stipulations, most of them not yet thought of by anybody, is worth an examination? By ending the exercise with a generalization such as the statement above, the pupil may attribute an exaggerated kind of competency to science. This only prepares the road to disappointment and suspicion.

At this juncture, the research worker will not only refrain from drawing a final conclusion; he will go back to his start and will change his stipulations. But how - the pupil might ask - can the stipulations be found which would lead to a positive answer, provided there is one?

We cannot know for certain whether the moon exerts an influence on the weather; no research worker knows for certain whether the thing he seeks really exists. Great scientific discoverers have exhibited instead most surprising traits, such as the intuitive talents of an artist, almost fanatical perseverance, keen faculties of observation; irrational motives have played a part as often as has induction. All these features would hardly be useful if the elements of research and the general method of science were one and the same thing. They are not:

Any scientific analysis, rigorously as it may be carried out, stands and falls with the arbitrary stipulations which must precede it.

Should this fact really be withheld from the pupil?

R. J. JARAY.

Mr Jaray has read into the article a purpose more ambitious than I was attempting. He asserts 'At its end the paper claims to have illustrated the *elements of research*.' What in fact I said was it 'contains several of those very elements of research which fascinate and discipline the scientist's mind.' I think my point remains good.

All the same I have found Mr Jaray's comments very interesting. I like especially his insistence on the difficulty of arriving at the right questions to ask. I agree with him that it would be highly undesirable to lead the pupil to 'attribute an exaggerated kind of competency to science.' This point about the limitations of the scientific method is one I made myself (p. 32). But in fact I have found in my experience with non-scientists (in Sixth Forms and adult classes) that the great difficulty is not to prevent the pupil getting an exaggerated idea of the competency of the method but to get across any appreciation of the method at all. Simplification, great simplification, is essential; I know my own classes would have been utterly befogged if I had introduced the qualifications and complexities Mr Jaray suggests. They were already sufficiently impressed with the complexity of the problem and the arbitrariness of our criteria in the preliminary process described (so briefly) on p. 36. Nevertheless, I am glad that he has stressed the need to point out, at the end of such an experiment as I have described, the tentative nature of the process and the conclusions. I hope that he, or other writers, will come forward with other problems equally amenable to simple scientific analysis by the layman.

I am glad that the article provoked this and other criticisms, even if not strictly relevant to its purpose. On reading it again in the light of the comments I think the cause of the misunderstanding was probably not my formulation of conclusions on pp. 37-38, but my use of the word 'superstition' on p. 32. It was evidently a mistake to mention this word, and not being a meteorologist, I am quite willing to withdraw it! The aim of my article was not to prove any general law about the moon and the weather - the numbers were quite inadequate for that - but to illustrate an educational method.

K. G. COLLIER.

The Electrical Activity of the Human Brain

DR. WILLIAM A. COBB

THE possibility of recording the electrical activity of the human brain is becoming widely known as a result of the campaign against epilepsy, and other less disinterested propaganda. From the remarks of the uninitiated when brought into contact with this technique it is clear that a complete misconception of its objects and limitations is usual, fostered not a little by sensational journalism. A simple account will be given of some of the phenomena observed, the methods of investigating them, and some applications of the results. Space does not permit discussion of the hypotheses of a subject which is still only a step removed from total empiricism.

In *Science News* 5 Dr W. A. H. Rushton described the mechanism of conduction in a single nerve fibre. The electrical unit of conduction is the 'action potential', which is due to the rapid passage along the surface of the nerve fibre of a zone having negative polarity with respect to the rest of its surface. When suitably recorded this has a duration of a few milliseconds and a rapid rise and fall, for which reason it is sometimes called a 'spike potential.' It is from such units that the vast mosaic of nervous action has to be built up. Messages in the nerve can be varied only by changes in the grouping of the fibres which are conducting and in the frequency at which each fibre transmits action potentials; generally speaking, increasing frequency implies increasing intensity; for example, of pain or muscular contraction. When a brief electrical stimulus is applied to an intact whole

nerve there is a high degree of synchronisation of the action potentials of individual fibres and the recorded response is relatively large and simple. If, however, the stimulus is of the normal physiological type, the stretching of a muscle for example, then the action potentials are quite unsynchronised and the recorded picture is a blur of many random spikes; the more of these there are, and the less they are synchronised, the greater is the likelihood that, in their algebraic sum as recorded from the whole nerve, little more than a thickening of the recorded trace will be observed.

The brain consists of very large numbers of nerve fibres with their nutrient cells,* and more or less supporting cells. Some of the fibres are gathered together in bundles bearing messages to or from distant parts of the body, but a great part of the brain mass consists of short fibres connecting brain cells with immediate or more distant neighbours. There is thus a bewildering network, not by any means fully mapped, of fibres passing in all directions. The situation is not unlike a telephone system in which the exchange, with all its possibilities of interconnection, is the brain, and the groups of wires on their poles along the road-side are nerves. It is quite possible to tap one of these wires – a single nerve fibre – and listen to the conversation: to tap all the wires of a group at once results in bedlam, with perhaps an odd word half-recognised now and then. When we come to the exchange, simple though it is compared with the brain, the confusion is still worse, for we are not allowed inside, and can only listen with an ear to the wall, so to speak, to *all* the messages coming in or going out.

Had you asked a physiologist of 20 years ago what possibility there was of recording the electrical activity of the human brain through the intact skull and scalp he might have replied, a little patronizingly perhaps, that there was none, using some such analogy as that of the telephone

* The relative importance of the fibre and the cell has long been discussed. A current view is that the cell exists merely to maintain the fibre.

system to convince you that it was not worth bothering to make the attempt. Even time, however, was not needed to prove him wrong, for since 1925 Hans Berger, a German psychiatrist, had been working away (with apparatus by modern standards very insensitive) to satisfy himself that he could detect on the human scalp electrical changes which had their origin in the brain beneath.

Quite in the best romantic tradition he satisfied none but himself and, though paper after paper was published, he received scant attention. Although what he had done was contrary to every expectation, the main reason for the lack of interest in his work was probably that he published it in psychiatric journals, largely unread by the physiologists who could, and finally did, appreciate the need for testing his results. It was not until 1934 that Professor E. D. Adrian of Cambridge published the first of a series of papers, with various collaborators, fully confirming the main findings of Hans Berger and starting an ever-increasing flood of observations on the changes in the 'brain waves' in health and disease. The clinical applications followed rapidly on Adrian's first publication, Grey Walter in England noting changes associated with cerebral tumours, and the possibility of locating the latter thereby, while in the United States F. A. Gibbs, with Davis and Lennox, observed electrical phenomena which were characteristic of epilepsy.

We must now consider the nature of the potential changes which, when recorded, collectively form the Electroencephalogram, as Berger termed it. (The related terms can be deduced by analogy with the Telegram: – graph for the 'machine,' – graphy for the subject as a whole, and so on. Habitual users eliminate these ponderous words and speak of the E.E.G., or sometimes, familiarly, of 'brain waves'.)

The most prominent feature of the E.E.G. is an almost rhythmic voltage-fluctuation at about 10 cycles-per second, with a peak-to-peak amplitude of about 40 μ v,* arising in

* 1 μ v = one millionth of a volt.

the posterior parts of the brain. It is sometimes still called the Berger rhythm, though Berger himself disclaimed the title, and it is more usually now termed the alpha rhythm. In some people it is nearly sinusoidal and may be remarkably constant in amplitude, but strict regularity of rhythm, form, and amplitude is exceptional; the latter in particular is subject to great changes from moment to moment, and deep, almost regular, amplitude modulation is a striking feature of some records. Almost all subjects show, at times, a complete absence of the rhythm, but there is great variation from person to person in the proportion of time for which it is present and the greatest amplitude which it may reach. Except in children, a greater amplitude, peak-peak, than $100 \mu\text{v}$ is very exceptional, while it is not at all uncommon for the mean amplitude to be $5 \mu\text{v}$ or less. While the frequency is much more constant, fluctuation in the individual of ± 1 c/s is quite normal, and the range from subject to subject is accepted as being from 8 to 13 c/s.

The most characteristic feature of the alpha rhythm has yet to be mentioned. The description given above applies to the record from a subject lying quietly at rest, with his eyes shut. If he opens his eyes the alpha rhythm, after a brief interval, disappears, to return again, after a similar interval, when he closes them (Fig. 13). In most subjects this 'blocking' effect is not absolute, and when the eyes have remained open for some time, alpha rhythm tends increasingly to appear, though of less amplitude than with the eyes closed. Other factors, such as a sudden loud noise, a painful stimulus and, in particular, intense mental effort will reduce or abolish 'alpha' activity; the efficiency of any one of these 'arousal stimuli', as they have come to be called, varies greatly with the individual make-up, though that of seeing, and especially of seeing something interesting, is generally by far the most powerful. In contrast, it is said that sound has an equivalent effect in the blind, although the attempt to see in the recently blind is still potent.

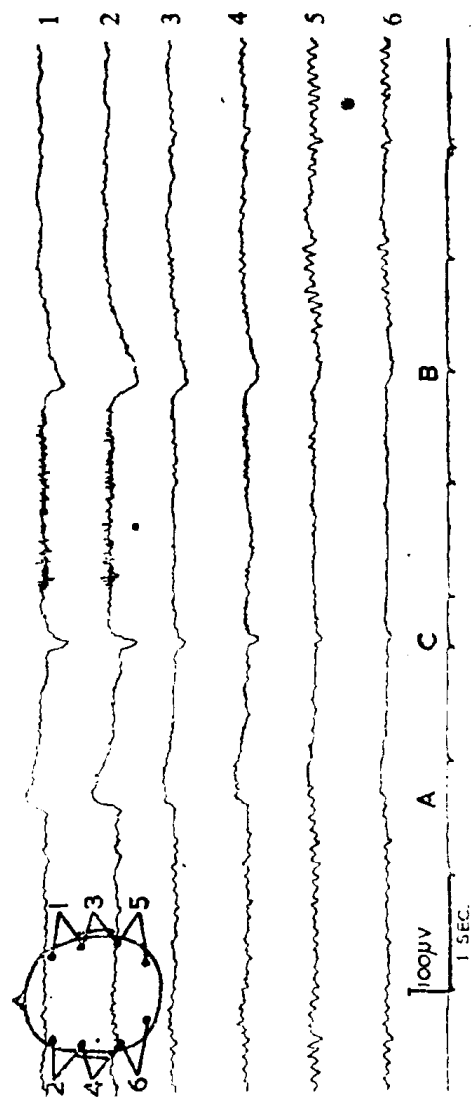


Fig. 13. — The E.E.G. of a normal subject, showing the occipital 'alpha' rhythm at 10 c/sec. The little diagram in the upper left-hand corner is the head seen from on top with its nose in the air, showing the position of the eight electrodes on the scalp, and how they are taken in pairs to give the six simultaneous records. At A the eyes are opened and the alpha rhythm disappears; at B they are closed and it returns. This record also shows two types of so-called artifact, due to muscle action-potentials, and to eye-movement. The former can be seen as a series of small spikes in the first two traces, just before the eyes are closed. The latter are also most evident, as would be expected, in the frontal records; at A there is an upward deflection corresponding with eye-opening, at B the converse change associated with eye-closing, while at C there is the quick double change characteristic of a blink.

Berger regarded the alpha rhythm as a function of the whole brain, while Adrian showed it to arise in areas very near to, but not identical with, those concerned with sight in the posterior regions of the brain; both views have had support, but the latter is almost universally accepted. Although arising far back, the rhythm spreads forward considerably and can be recorded in most subjects well in front of the plane of the ear, there being a steady decrease in amplitude from behind forward. The nature of the spread is still a matter of doubt; it may be purely electrical, as in an electrolytic solution, or it may be by conduction along nerve fibres, or both. An origin from both cerebral hemispheres can be demonstrated and the two sides usually show fair, though not perfect, synchrony; amplitudes, on the other hand, can be far more independent of each other.

The nature of one of the misconceptions mentioned in the introductory paragraph may now be evident: the E.E.G. is far from being a record of the brain at work; its most striking feature is derived from the brain at play, or at most marking time. Adrian's interpretation of the alpha rhythm was that it represented resting activity, the idle cells 'beating' at their natural period, in unison; during periods of visual or mental activity the influx of stimuli from outside would result in each individual cell working at the rate appropriate to its job, and the sum of this unsynchronised activity would no longer be recordable.

In addition to the alpha rhythm the normal adult E.E.G. may contain more or less of other features, the most common of which is a frontal rhythm at about 22 c/s. This may be entirely absent, continuously present, or something in between, but is always of low voltage (by E.E.G. standards). Waves at 6 or 7 c/s are also commonly seen and regarded as normal if their amplitude is small. Both the fast (beta) and slower (theta) rhythms have pathological medical significance when of larger amplitude, in contrast with the alpha rhythm, the amplitude of which alone seldom has any special meaning.

ELECTRICAL ACTIVITY OF THE HUMAN BRAIN 125

Even when there is no beta or theta rhythm and the alpha rhythm is blocked, there is not complete absence of all activity; there remains an irregular background, perhaps to be regarded as the 'noise level' of the brain, of the order of 2 or 3 μ v.

Now that some idea has been given of the content of the E.E.G. it is appropriate to consider the methods of recording it which are used, and which are to some extent a limiting factor to further advance. The earliest investigations were made with a string galvanometer, a technique deriving from the well-established practice of electro-cardiography. This instrument had two great disadvantages in its new application: it was too insensitive to deal satisfactorily with potential changes only 1/20 part of those derived from the heart, and it was a current-operated device. If current is drawn from the brain cells the potential across them immediately falls, and a faithful picture of their behaviour is impossible.

The introduction of valve-amplifiers, with high input resistance, was therefore an essential first step to progress, and the main lines of their development were quickly established by Matthews, Tönnies and others, not only for the purposes of the E.E.G., but as general physiological amplifiers. A 'gain' of the order of 2 million times is required, with good stability, absence of background noise, and the ability to discriminate against extraneous electrical interference. The latter is achieved by the use of a balanced push-pull input; the output is also push-pull, so it is usual to construct all the four or five stages in this way; they are condenser coupled, thus overcoming the problem of drift, an over-all time constant of about 0.3 secs. being quite adequate for most E.E.G. recording. The electrical 'noise' inherent in amplifier components, and particularly valves, is of the same order of magnitude as the smaller changes in the E.E.G., and at present sets the practical limit of useful amplification. By careful design, and choice of components, this random variation can be brought down to about 1 μ v

peak-to-peak, but few amplifiers will maintain this low level without constant attention.

The E.E.G. amplifier was at first made to drive some form of mirror oscillograph, the deflections of which were registered on moving photographic paper. This was also the case with the cathode ray oscillograph, which superseded the mirror type and which is still used for investigations requiring special accuracy rather than long continuous records. The use of photography for recording has been generally abandoned as laborious and costly; moreover, the record is not instantly available to allow comparison of one electrical change with another. Such immediate records can be written out in many different ways, most of which are in use with commercial oscillographs; the most obvious and generally used is that of ink on white paper, which, although inclined to be messy, has the advantage of giving good contrast and using relatively cheap materials. The stylus may be driven by a moving-iron, moving-coil, or even crystal, oscillograph. It is this component which has given designers the greatest trouble; the requirements of linear voltage-response, linear frequency-response from 0 to (say) 100 c/s, an excursion of at least 2 cm. peak to peak, and robust construction, all within small size, have proved difficult to achieve. Although there are now a number of satisfactory designs, there is still room for improvement.

The record obtained is a graph of voltage changes against time, the time-axis being provided by the movement of the paper tape on which the record is written out. It is evident that one such record can indicate only the changing voltage differences between two points, and to study the behaviour of various parts of the brain simultaneously a number of recording systems ('channels') is necessary. This number could advantageously be large, were it not for the difficulty of interpreting large numbers of traces simultaneously, and the high cost (about £200 per channel) of the apparatus; most modern machines therefore compromise with six or eight channels (Plate 20). That it is something of an achieve-

ment to keep this number in constant daily use may be realised by imagining a similar number of radio sets, all together in a cabinet rather smaller than a cottage piano, all tuned to different stations and the stations changed every few minutes throughout the day; the interference between the programmes must be negligible, the mains-hum inaudible, and the amplification many times greater than is asked of a domestic radio.

The electrodes, with which the brain's electrical changes are picked up, are usually either metal disks about 1 cm. in diameter, sealed to the scalp with collodion, or small gauze pads soaked in concentrated salt solution and held on the scalp by some form of cap - one designed for hair-waving rather than 'brain-waving' has been much used in Britain. Many different patterns of electrode arrangement have been devised, each conferring some real or imagined advantage, and no sort of standardization has yet been reached, or indeed, by many, desired. On the whole, with increasing numbers of channels, larger numbers of electrodes are used, tending to give a more detailed picture at the expense of more complicated technique. The latter is by no means always justified by the former, and increased detail carries with it a further burden for the interpreter. Most routines involve 16 electrodes or thereabouts.

The potential changes at the electrodes are led by a multiway cable, usually to enter the machine by way of a complex selector switch, with which it is possible to connect any electrode with either input-grid of any of the amplifiers, either in prearranged groupings or independently. This input switch has associated with it a device for applying small rectangular test voltages to the amplifiers, to check their performance and to serve as a voltage calibration. A usual value for this test signal is 50 μ v and the 'gain' is commonly adjusted so that it gives rise to a deflection of 5 mm.

A typical modern record, then, consists of six traces written in ink on white paper about 6" wide, with additional

traces for recording time and random events. The paper moves at 3 cms/sec., and the record from a patient may last for 20 minutes or more. When it is considered that each of the six traces is at least slightly different from the others, and that no one second of activity is exactly like another, it will be realized that the task of visual interpretation of the 40 yards of records, which represent 20 minutes, may be formidable. In practice little attempt is made at precise analysis; the main frequencies are noted, with their spatial and temporal distributions, and their relationships to one another and to changes in the subjects' attention, wakefulness and so on.

A further complication confronting the interpreter is the presence of electrical changes not derived from the brain, to which the general, and not wholly appropriate, term of 'artefact' is given. Their sources are many and their forms diverse; some are instantly recognisable, others mimic the cerebral activity with great fidelity. To mention a few: from extraneous sources comes interference from A.C. mains, diathermy, electric motors, etc., and the brief spikes from switch contacts; from the recording system there are transients due to bad contacts in plugs and the like; from the patient large slow artefacts associated with movement, due to redistribution of static charges, rocking of insecure electrodes and so on. From the patient also come the electrical counterparts of certain physiological changes which in other circumstances are themselves the subject of recording, and which are not then regarded as artefacts; of these the electromyogram from the scalp muscles, the electrocardiogram from the heart, and changes associated with eye-movement are the chief (Fig. 13).

With good apparatus, an electrically screened recording room, careful technique, and cooperative subjects it is possible to obtain records which consist almost solely of cerebral activity, but this perfection is hard to maintain, and is indeed impossible when dealing with unselected patients. Part of the skill of interpretation is, therefore, the

ELECTRICAL ACTIVITY OF THE HUMAN BRAIN 129

ability to recognise and reject changes which are not of cerebral origin.

Some years ago the English E.E.G. Society attempted to lay down limits within which the record could be considered normal, a task which proved to be too difficult and which was abandoned in favour of the easier delimitation of abnormality. The great variation in content of the records from normal subjects has already been indicated; if this is regarded as being of the nature of personal idiosyncrasy there are still other causes of variation to be allowed for, such as age, wakefulness and emotional tension.

To take the first, there is a steady process of maturation from birth to somewhere about the 16th year. The E.E.G. of the infant is a rather arrhythmic affair, showing irregular swings of voltage, a second or more in duration and about 100 μ v in amplitude. These tend to become more rhythmic

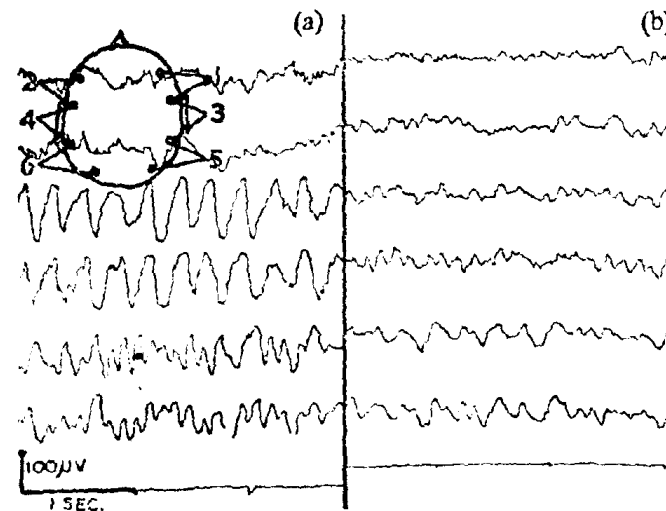


Fig. 14. - (a) Normal 4-year old child. The alpha rhythm is well established, and a $3\frac{1}{2}$ c/sec. rhythm in the post-central region is a prominent feature. (b) the same child one year later, showing great decline in the $3\frac{1}{2}$ c/sec. activity.

and less slow; to them are added rhythms at 3-7 c/s which are the most evident feature from about 2-6 years (Fig. 14). Some time during this period the alpha rhythm appears, gradually increasing in amplitude and persistence and usually being well established by 8 years, at which time the slower rhythms have diminished considerably. The adult form is often achieved by 12 years, though there may still be some instability and a greater tendency to respond to changes in the blood supply to the brain than there is in later life. That there is a correlation between age and the maturation of the E.E.G. is clear, but 'age' is not necessarily to be regarded as the time elapsed since birth: mental age is a familiar concept, and sexual age easily understood. The correlation does not seem to be with any one of these alone, and it may be that all in varying degree, and others as well, are significant.

The record in deep sleep is not unlike that of the infant and it has been suggested that, as sleep lightens to wakefulness, the E.E.G. recapitulates its maturation. Though this idea is rather fanciful it does indicate the variation of content which occurs with depth of sleep. Deep sleep is hardly likely to go unnoticed by the observer who is recording the E.E.G., but the early stages, characterised as 'drifting', are easily missed by observer and subject alike, and yet may modify the record considerably, tending to abolish alpha rhythm and introduce runs of theta rhythm, and of waves at about 14 c/s.

Despite the numerous factors whose effects may have their reflexion in the E.E.G., and there are many others which have not been mentioned, the record from any subject has considerable individuality. Suppose that your E.E.G. has been recorded on two different days; one of the records is mixed with those of half a dozen other subjects at random; then, using the first record as guide, there would be a good chance of finding its mate correctly. There are also family resemblances, probably about as close as family facial likenesses and, like them, reaching the closest simi-

larity in uniovular (identical) twins. It has been said that the E.E.G. is second only to finger prints in its reliability for making the distinction between uniovular and binovular twinning.

In all that has been so far described there is little indication of why it is that an ever-growing number of expensively staffed and equipped laboratories is set up to record the E.E.G. of patients. In fact, the normal E.E.G., as a study in physiology, has received but scant attention compared with the enormous outpourings, chiefly from the U.S.A., on its aberrations.

The general types of abnormality will be described briefly. They are of two main kinds: rhythms* which are unusually fast or unusually slow, and waves which have unusual forms or associations one with another. It is also true, for example, that rhythms which are normal in childhood may be abnormal in later life, or which are normal in sleep may be abnormal in wakefulness.

It is in the diagnosis and investigation of epilepsy and of cerebral tumours that the E.E.G. has made its greatest clinical contribution, and these subjects will be taken to illustrate the type of information which can be obtained.

Epilepsy can broadly be classified as Idiopathic or Essential on the one hand, and Symptomatic or Acquired on the other. The former, or a predisposing cause, is inherited, and its treatment, though increasingly successful, is purely palliative; in the latter, on the other hand, the fits are due to some local brain disease such as a tumour, or a scar resulting from past injury or infection; the ideal treatment, not always possible, is removal of the cause. Furthermore, fits are very frequently the first indication of the presence of such a cause, and consequently to make the distinction between the two types of epilepsy is of the greatest importance. It is also often of the greatest difficulty, and while clinical, radiological, and electrical methods all

* A term of convenience only, not acceptable to the physicist. Strict rhythmicity is almost non-existent in the E.E.G.

have their successes it is fortunate that their spheres do not precisely coincide, and the failure of one method may be the success of another.

An invariable accompaniment of the type of epileptic fit classically described as *petit mal* is an outburst of complex waves at about 3 c/s, each consisting of a surface negative spike, followed by a hump-shaped wave of the same

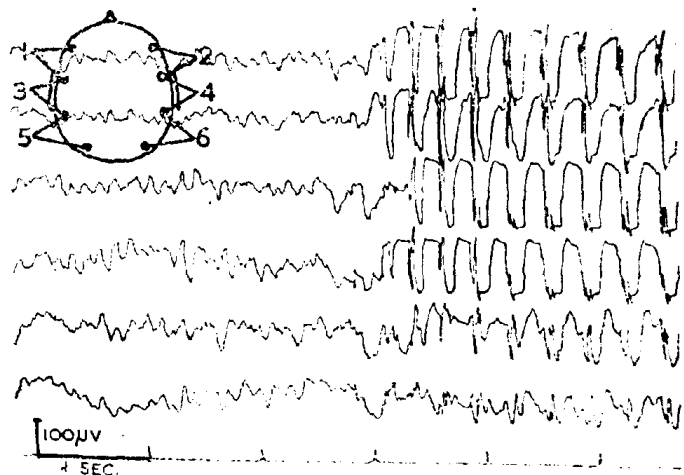


Fig. 15. — *Petit mal* epilepsy in a 10-year old child. The activity normal for this age is suddenly replaced by high-voltage complexes recurring $3\frac{1}{2}$ times/sec., and each consisting of a wave followed by one or two spikes. The flat tops of some of the waves are due to the limited movement of the pens, which are unable to record the full amplitude. This outburst lasted some 20 seconds and corresponded with a brief partial loss of consciousness or 'absence.'

polarity (Fig. 15). They also occur in the inter-seizure periods, when the outbursts tend to be briefer, and they occur in subjects whose fits are not only, or never, of the *petit mal* type. They are diagnostic of epilepsy, and when occurring regularly and synchronously from both hemispheres, are characteristic of Idiopathic Epilepsy.

ELECTRICAL ACTIVITY OF THE HUMAN BRAIN 133

Recording of major fits (*grand mal*) is somewhat uncommon and the form of electrical manifestation variable, usually including some high voltage* spikes and always much slow activity, but seldom any strictly repetitive complex wave-form. The electrical diagnosis of idiopathic *grand mal* depends largely upon the appearance of the record between fits, when it may contain 'spike and wave' complexes or variants of this form, or groups of waves with a 'paroxysmal' quality, standing out sharply from their neighbours by reason of a sudden change in amplitude or frequency, or both.

In symptomatic epilepsy the findings are of two kinds: those due to the causative disease, to be described later, and those which are a direct indication of an unduly irritable area of brain (due to injury in its widest sense), from which the epileptic attack may originate. Jasper and Penfield of Montreal have made the latter their special study, and have shown that the characteristic manifestation from such an irritable area is an irregularly repeating surface-negative spike with a duration of about 1/30 second and an amplitude between 50 and 500 μ v, which can be accurately located if it arises from an accessible part of the brain (Fig. 16). When its source is not on the upper convex surface of the brain it must travel some distance before it can be recorded, and in these circumstances it becomes broadened out, to have a duration of perhaps 1/5 sec.; it is then termed a 'sharp wave.'

By no means all the records from epileptics show the characteristics described and indeed about 25 per cent of them are quite normal, while a similar proportion, perhaps more, are abnormal without showing any of the typical wave-forms. If, however, the recordings are repeated sufficiently often, and under appropriately varied conditions, the majority, possibly all epileptics, will be found to have abnormal records. This variability can be correlated with

* As always, magnitude is relative: to the electroencephalographer 'high voltage' means at most a few hundred microvolts.

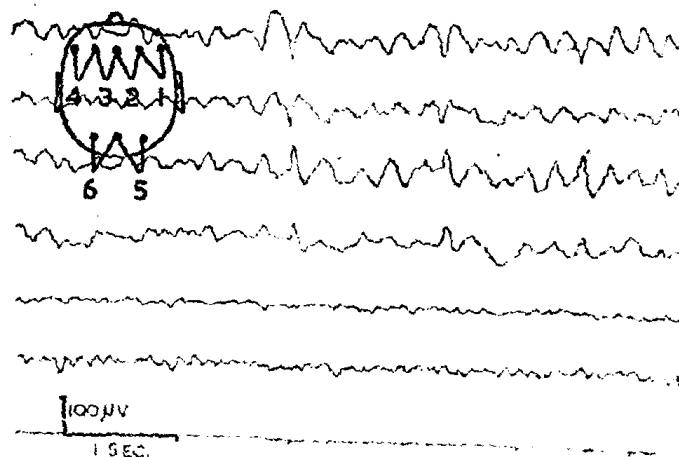


Fig. 16. - An irritable focus, the cause of epilepsy in a girl of 17. In the posterior regions the E.E.G. was relatively normal, while in the frontal regions there was an area from which 4 c/sec. waves arose. From the centre of this area there were many single 'sharp waves' of which three are illustrated.

the great variability of the clinical manifestations: a subject may go for months without a fit and then have a large number in quick succession; his E.E.G. might be normal during the former phase and highly abnormal during the latter.

A wide variety of organic brain diseases cause E.E.G. changes which are essentially similar, the mechanism being probably the same, though minor variations allow distinctions to be made. In a diseased area brain cells may be either healthy or dead, or something in between. In the first case electrical activity will be normal and in the second absent, so that it is only the sick and moribund cells which can produce abnormalities. These are in the form of waves which are unusually slow, from less than 1 per sec. to about 3 per sec., usually very irregular in period, voltage and form,

and with an amplitude of some 50 to 200 μ V (Fig. 17). In the case of a cerebral tumour the tumour-tissue itself is inert and it is from the surrounding brain cells, distorted and depleted of blood supply by pressure, that the slow waves arise by which the site of the tumour can be located. As with epilepsy, it is not every brain tumour or other gross cerebral disease which puts its recognisable seal on the record; many, in fact, give no indication of their presence.

Electroencephalography, therefore, as a clinical tool, presents three main problems to the interpreter: first, to get more information into the record; secondly, to extract more of the information which undoubtedly is already there, shrouded in the great complexity and variability which the

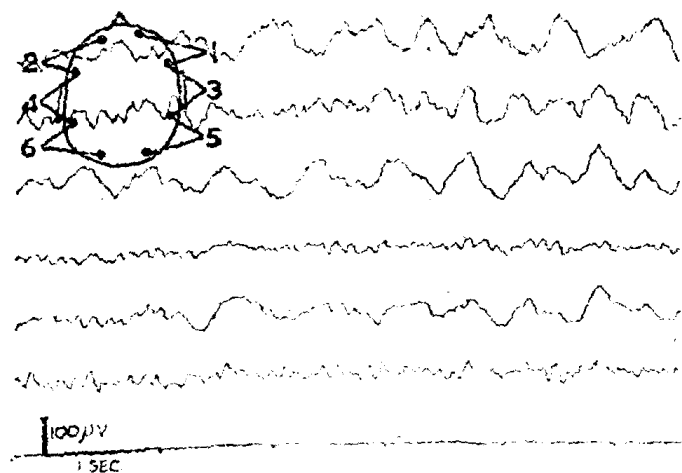


Fig. 17. - Cerebral Tumour in a man aged 39. The E.E.G. on the left side is relatively normal, while on the right there are large, slow, irregular waves. Traces 1 and 3 are, very roughly, mirror images of each other, indicating the origin of the slow waves at or near to the electrode common to these two channels. This method of 'phase reversal' is the most reliable way of locating the source of any disturbance.

record possesses; and lastly, bound up with the first two problems, to reach a better understanding of the fundamental processes which give rise to the potential changes.

An obvious first step towards putting more into the record might seem to be to amplify the brain waves still further; unfortunately this is probably not possible, and, if it were, would almost certainly add still more indigestible detail to the half digested E.E.G. of to-day. It is unlikely to be possible because the present amplifiers are working at a 'gain' at which potential changes due to the random movements of electrons in valves, and similar uncontrollable factors, are little smaller than the smallest features of the E.E.G. as we know it.

A more promising approach is being rapidly developed and extended in a number of directions. This is to record the brain waves under various normal or near normal conditions, and during the application of various 'stresses'. To instance a few of the former: it has been, from the beginning, an almost invariable routine to make the subject 'overbreathe' (breathe more deeply and rapidly than his bodily requirements demand) for about three minutes. This causes a change in the acidity of the blood, and, in nearly all subjects, will result in some changes in the E.E.G. if performed sufficiently vigorously. These changes are sometimes of diagnostic value, and frequently so in *petit mal* epilepsy; the record which was normal or nearly normal at rest will usually contain 'spike and wave' complexes during or immediately after over-breathing.

Recently it has been shown that some subjects who have normal records while awake may have very abnormal ones when asleep, including wave-forms which are characteristic of epilepsy. This is not altogether surprising, as many epileptics have their attacks predominantly or only when asleep.

An example of a different type is the provision of appropriate emotional stimuli: patients who will respond with epileptic phenomena to such stimuli are uncommon but

interesting. In one woman the E.E.G. became more and more abnormal as she listened to a record of 'Eine Kleine Nachtmusik' and finally, towards the end of the second side, she had a major epileptic attack. Another patient had had several attacks, all following the ringing of a telephone bell; attempts to repeat this in the recording room were a complete failure; it may have been that the pitch or rhythm of the bell was wrong, or possibly the conditions of the recording room were too unusual for the sound to make the appropriate emotional impact.

Most of the 'stresses' which have been applied are chemical in nature. For example, drugs of the camphor group will cause fits in all subjects if administered in sufficiently large doses; recently a technique has been devised of giving repeated small doses until the E.E.G. first begins to show changes, and in this way the site of origin of epileptic activity may be demonstrated without the undesirable result of precipitating a fit.

A most interesting new technique of physical stimulation has been devised by Grey Walter of the Burden Neurological Institute. It has been known since 1934 that if a light is flashed rhythmically in the eyes, at around 10 c/s, responses at a similar rate can be recorded from the posterior parts of the head. If this is done at a frequency appropriate to the subject, as found by trial and error, and particularly if the occurrence of the flashes is made to have a certain phase relationship to the responses, then the latter may increase in amplitude and spread, or develop into characteristic epileptic forms, sometimes culminating in an epileptic attack. The exact frequency and phase required to produce such a response may be highly critical.

The second problem to be discussed is the interpretation of the record, and, in particular, how more can be learnt from it than simple inspection and plotting of frequencies will allow. The possibilities of, and need for, some form of frequency analysis were early appreciated, although it was obvious that the variation in amplitude of the presenting

rhythms made it impossible to represent the whole E.E.G. by any single expression, as may be done for a continuous unvarying rhythm. However, if any arbitrary, and necessarily brief, period of time is considered it is possible to plot frequency components in relation to this arbitrary fundamental. This may be done graphically with infinite tedium, or, as A. M. Grass devised in 1938, by mechanically multiplying the frequencies to bring them within the range of a commercial type of wave-analyser. These are precise methods, and in practice not very useful, as they produce answers for only small sections of the record, a few seconds at the most, and time and cost do not allow analysis of the whole, or even of an appreciable part of it.

A first approach toward greater usefulness is to make the analysis in discrete steps of frequency, rather than continuous. This allows the use of a number of circuits each tuned to a given frequency in the relatively narrow wave-band with which the E.E.G. is concerned. At any moment the amplitude of the output of any one of these circuits will represent the energy at the corresponding frequency in the primary E.E.G. at that moment, and it can be presented as a constant frequency - variable amplitude trace. To do this for some 20 or 30 discrete frequencies is possible in the laboratory, but too cumbersome for routine use, and some condensed form of presentation is necessary, again sacrificing precision to utility. Such a presentation has been devised by Walter, and is in increasing use. An epoch of 10 seconds has been chosen, during which the energy at each frequency is stored in condensers; during the succeeding 10 seconds the stored energy is released, frequency by frequency, to write a histogram on the same paper as the primary record, on which it is superimposed. Meanwhile, a second set of storage condensers is being charged, the two sets alternately receiving and discharging every 10 seconds. With a display of this type it is not possible to distinguish between a rhythm of amplitude x present for the whole epoch and a similar rhythm of amplitude $2x$

ELECTRICAL ACTIVITY OF THE HUMAN BRAIN 139

present for half the epoch; nor does it convey any information on the phase relationships of the several frequencies. The apparatus must not, therefore, be divorced from the primary record but in conjunction with it, by virtue of its relative simplicity, can add significantly to the value of the E.E.G.

Although the human E.E.G. recorded from the scalp has made, and no doubt will continue to make, its contributions to the better understanding of cerebral physiology and pathology, it is largely from experiments on the exposed animal brain that most advances can be expected to come. However, the human brain is not infrequently uncovered in the operating theatre and its electrical activity can then be examined, with fruitful results not only in the more exact location of tumours and, in particular, of epileptogenic foci, but also in the precise topographical representation of function. Incidentally, it is confirmed that the E.E.G. as recorded from the scalp is a true picture, the main difference being a ten- to one hundred-fold increase of amplitude, and consequently of detail.

In its brief life of some fifteen years electroencephalography has attracted the attentions of the overenthusiastic and the sensation-monger (how often has one said, 'No, it is not a lie-detector!'), and in the eyes of some has been discredited thereby. Nevertheless, it has a solid record of achievement and has produced a body of new knowledge obtainable in no other way. Although it has established itself as a useful clinical aid its brightest future probably lies in the help which it can give in elucidating that most fundamental of all man's problems, the workings of his own brain.

REFERENCES

No comprehensive text-book of this subject exists, but one is in the press and should be published in the Spring of 1950.

Electroencephalography, edited by D. Hill. Macdonald & Co., London.

Less complete and now nine years old is:

The Atlas of Electroencephalography, F. A. and E. L. Gibbs., Boston, 1941.

The first number of an international quarterly journal was published in February, 1949:

Electroencephalography and Clinical Neurophysiology. Montreal.

Reference to a representative selection of the very large number of individual papers is impossible but mention may be made of the first two papers of Adrian, which are classic:

Adrian, E. D. and Matthews, B. H. C. The Berger rhythm; potential changes from the occipital lobes in man. *Brain* 57, 355 (1934).

Adrian, E. D. and Yamagiwa, K. The origin of the Berger rhythm. *Brain* 58, 323 (1935).

And for technical methods to:

Parr, G. and Walter, W. G. Amplifying and Recording Technique in Electro-biology. *J. Inst. Elect. Eng.* Feb. 1943.

Synthetic Fibres

R. W. MONCRIEFF

THE description of a fibre as 'synthetic' is applied only to those in which the polymeric molecules which compose the fibre are synthesised from simple chemical substances. Fibres which are regenerated from natural polymers such as cellulose or casein, by dissolving them and extruding the solution into a precipitating medium, are described as 'man-made' or 'artificial' but not as 'synthetic.' Such fibres as viscose, cuprammonium, cellulose acetate, alginate, casein, soya bean, and zein all depend for their fibrous properties on the long polymeric molecules which occur naturally in wood, cotton, seaweed, milk, soya and corn; they are not truly synthetic. But within the last twelve years a considerable number of fibres have appeared which are built up from simple chemicals; in these the polymer is made synthetically: nylon, Vinyon, Velon, Terylene, Pe Ce and Orlon.

Fibrous Molecules

Before a substance can be synthesised, it must be analysed; it would, for example, have been quite useless to attempt to synthesise penicillin until analysis had revealed its constitution. So it was with fibres. Twenty-five years ago fibres were thought to be complex and irregular bodies with no simple design. The first real insight into the structure of solids came from the X-ray diffraction experiments carried out by Laue and W. H. and W. L. Bragg; these experiments were made on simple crystalline bodies and X-ray diagrams revealed the dimensions of the molecule and by inference the arrangement of the atoms in the molecules.

This method of analysis was extended by W. T. Astbury

to fibres, and the outstanding discovery emerged that the molecules that made up wool, cotton, and other fibrous materials were themselves very long and very narrow; in other words they were 'fibrous.' It is a characteristic of natural fibres that they are all long in relation to their breadth, and, in the cases of cotton, wool, flax and hemp, the fibres are on the average 1,000 to 2,000 times as long as they are broad. Just as it needs reasonably long fibres to spin into a good yarn (cotton with a fibre length of less than $\frac{1}{4}$ inch is almost impossible to spin), so a prerequisite for a good fibre is that the constituent molecules must themselves be very long. The long narrow shape of the natural fibres is a reflection of the long narrow shape of their molecules.

Synthesis of Long Molecules

When it was appreciated that molecules must be very long for fibrous properties to be manifest, chemists envisaged the possibility of building up long molecules and so making new fibres, - synthetic fibres. They remembered how Emil Fischer, attempting to build up proteins, had combined nineteen molecules of glycine together to give a polypeptide. Could this work be extended to give very much longer molecules? Fischer had found it by no means easy to build up his relatively short polypeptides and thirty or forty years later, despite the tremendous advances in knowledge of chemical structure and synthetic methods that had been made, the problem was still very difficult. One reason for this was that the chemistry of very big molecules, of polymers, had been neglected; if an experiment produced an insoluble, infusible, non-crystalline and unreactive material it was thought to have failed and the product was thrown into the waste bin.

Carothers' Superpolymers

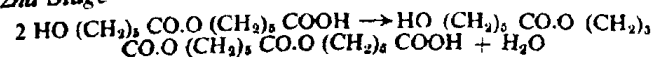
W. H. Carothers, given a free hand by his employers, Du Pont de Nemours & Co., realised that these 'failures' were more closely allied to such everyday materials as wood,

rubber and fibres than were the crystalline, soluble, and fusible contents of the bottles on the chemists' shelves; and he deliberately set out to make polymeric substances. He reasoned that if he took a substance which had a reactive group at either end of its molecules, two such molecules would combine to give a molecule twice as long, and still with a reactive group at either end. Then two of these would combine with each other, and by such a continued process of doubling, very long molecules would soon result. Carothers showed that when such a substance as ϵ -hydroxycaproic acid was suitably heated the molecules joined up in the following way:

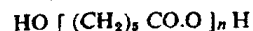
1st Stage



2nd Stage



Eventually very long molecules with the formula



where n was a relatively large number, of the order of 50 or 60, would be formed.

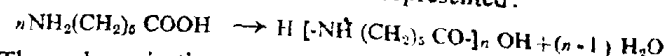
Note that, because relatively long molecules may sometimes unite with still relatively short molecules, n is not necessarily an integral power of 2; it would be so only if reaction took place exclusively between molecules of similar chain length. The molecules have linked up to give long linear polymers, and when these polymers exhibited fibre-forming properties Carothers termed them 'super-polymers', as an indication that they were very long indeed. If the polymers were made from a hydroxy-acid they were super-polyesters, as they were linked up through ester groupings.

The first success in synthetic fibre formation came when a chemist on Carothers' staff poked a glass rod into a molten superpolymer and observed, doubtless to his astonishment,

that when he withdrew it a long filament came out with it and rapidly set solid; and, still more surprising, even when it was cold and solid it could be stretched just like rubber to about four times its original length, but, unlike rubber, it did not retract to its original length when the tension was released. Before it was stretched the filament was dull and rather weak, but afterwards it was strong and lustrous. These changes were brought about by orienting the molecules parallel to the fibre length, the orientation being the molecular accompaniment of the physical stretching of the fibre. The logic behind the work was vindicated; if sufficiently long molecules could be built up, the products would be fibre-forming. The superpolyesters gave fibres, but the fibres suffered from the disadvantages that they were not very stable to chemical attack and that they had such low melting points that trouble would inevitably have been experienced in ironing fabrics made from them.

Superpolyamides

It was found that if instead of using a hydroxy-acid as the starting material, an amino-acid was used, the product – now known as a superpolyamide – was free from these defects. If, for example, ϵ -amino caproic acid was used, the ultimate polymerisation reaction could be represented:



The polymerisation, or condensation, is carried out at a temperature of about 280°C . either in a good vacuum or in nitrogen atmosphere. The final product melts at about 220°C .

Nylon 66

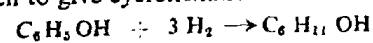
Amino acids are relatively difficult and expensive to make, and, probably largely for this reason, the device was adopted of making polymers from diamines and dibasic acids, both of which are more readily accessible. The particular compounds most suitable in this respect are hexamethylene

diamine and adipic acid; each of these compounds contains six carbon atoms and can be made from phenol, a cheap and plentiful raw material. For this reason the polymer was known as '66.' The name 'nylon' was coined for it; it is euphonious and the 'on' ending has a textile flavour (*cf* cotton, rayon).

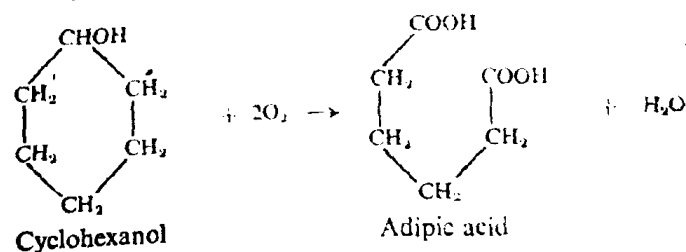
If sebacic acid $\text{COOH}(\text{CH}_2)_8\text{COOH}$ is condensed with hexamethylene diamine the product is 'nylon 610.' Clearly, by varying the chain lengths of the intermediates a large number of different nylons can be made; very many have been made, but 66, the one first made, has still the best all-round properties; though for bristles the greater rigidity and lower moisture-imbibing power of the 610 product are preferable.

The Manufacture of Nylon

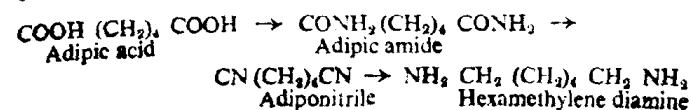
Phenol is the starting material; it is catalytically reduced with hydrogen to give cyclohexanol



The cyclohexanol is oxidised with nitric acid, which breaks the cyclohexane ring and yields adipic acid:



The adipic acid is converted to the amide, thence to the nitrile which is reduced with hydrogen in the presence of a cobalt catalyst to hexamethylene diamine.



The hexamethylene diamine and the adipic acid are dissolved separately in methanol, and, when the solutions are mixed, the salt hexamethylene diammonium adipate is precipitated in a pure form. It is essential to use pure reactants in polymerisation reactions as small quantities of impurities may sometimes block the functional end groups and interfere with the course of the reaction. The nylon salt is further purified if necessary. It is then heated with water under pressure at about 220° C.; the pressure is used to prevent loss of volatile diamine, the water to secure a uniform melt; once the early stages of polymerisation are effected, the pressure is released and the temperature raised to 280° C., heating being carried out under an inert atmosphere, to avoid the decomposition which occurs if nylon is heated in the presence of air. As it is required to have a product with a molecular weight of about 12,000, a stabilizer, a trace of 'impurity', is first added to prevent the reaction going too far; acetic acid is the stabilizer used. If a dull yarn, matt or pearl, is eventually required, 1 or 2 per cent of titanium dioxide, a white powder with excellent covering power, may be added at the polymerisation stage.

When polymerisation is complete, the molten polymer, which in appearance is like molten glass, is extruded through a slot in ribbon form on to a conveyor belt; it is immediately quenched with cold water sprays to reduce the crystal size. The ribbon is broken into chips and these are fed into a spinning machine, in which they are melted and extruded through very fine orifices. The streams of molten nylon solidify on emergence from the orifices into the air and form nylon filaments which are wound, stretched or 'cold-drawn' to about four times their original length to give lustre and strength, and then are ready for normal textile processing.

The Properties and Uses of Nylon

The uses to which a textile fibre can be put are determined by the properties of the fibre, so it will be convenient to consider the two together.

Nylon is strong; its tenacity is about 5.5 grams per denier; a length of about 49 Km. would just break under its own weight. It has a high elongation (22 per cent) at break, and recovers well from strain, i.e., it is highly elastic. Even if it is knotted it retains most of its strength, a most unusual property in a strong textile fibre. These properties have made it particularly suitable for use in parachute fabric and harness, and for tow ropes. It can also, because it is so strong, be used in very fine denier (gauge) and this, combined with its excellent elasticity, makes it admirable for women's sheer hose.

The absence of hydroxyl groups in its structure makes it much less hygroscopic than cellulosic fibres and while for some purposes, such as plastic waterproofs, this is advantageous, low moisture absorption is not in general helpful in its use for apparel, particularly underwear. It has excellent resistance to abrasion and in the staple form will probably be mixed with wool; the admixture of 10 per cent nylon should increase the length of life of woollen socks by 50 per cent. It is immune from attack by moths and mildew.

The chemical resistance of nylon is excellent; even hot concentrated solutions of caustic soda have little effect on it, and it is stable to dilute acids. Its chemical stability has made it useful for such purposes as filter cloths, chemical workers' clothing and fellmongers' brushes. These last are used to paint the insides of the skins of slaughtered sheep with alkali to loosen the wool and make it easier to pull off; formerly bristle brushes which were badly attacked by the alkali were used and lasted only a few weeks, but nylon brushes last almost indefinitely. Nylon is a versatile fibre and there can be no doubt that it will invade most textile fields; a great advantage is that modification of its physical properties is possible by suitably modifying the intermediates from which it is made.

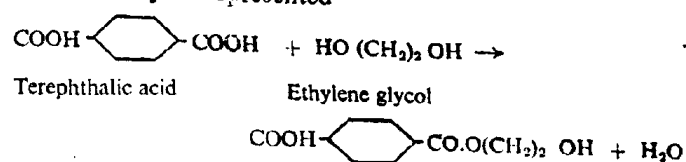
Elastic Nylon

An example of this modification is to be found in elastic nylon. This resembles rubber in its stretching and con-

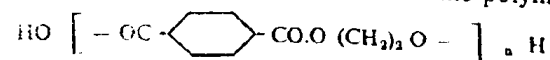
tracting powers. It is made from sebacic acid, hexamethylene diamine, and N-isobutyl hexamethylene diamine. The presence of the isobutyl groups prevents the long linear molecules lying so close together and reduces the cohesive forces between the polar groups of the nylon molecules. When the elastic nylon is 'cold drawn' the molecules cannot arrange themselves so compactly and, owing to the distortion that takes place in the side-chains (the isobutyl groups), no stable molecular arrangement can be taken up, so that when the tension is released the elastic nylon reverts to its original shape; it contracts as rubber would.

Terylene

Although Carothers found that his polyesters melted too low to make satisfactory fibres, more recent work carried out by chemists of the C.P.A. in this country has shown that a satisfactory product can be obtained by condensing terephthalic acid with ethylene glycol. Because methyl terephthalate is more easily purified than terephthalic acid, the ester is preferred for use in the polymerisation instead of the free acid, but the final polymer is the same. The reaction may be represented



and ultimately on further condensation to the polymer



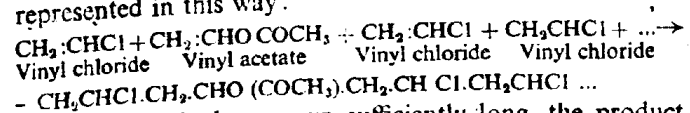
which is Terylene. This fibre has a melting point of 245° C., not quite so high as that of nylon 66. It is melt spun and cold drawn. It has a high strength, of the order of 7 grams per denier, but is more resistant to water than nylon and has so far proved more difficult to dye. The inclusion of the benzene rings in the structure will make this fibre more

inflammable than nylon, but in other respects it is generally similar to nylon.

Vinyon

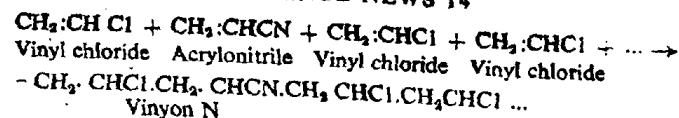
Whereas nylon and Terylene are formed by 'condensation polymerisation', in which water or some other small molecular substance is split off, Vinyon and some other fibrous polymers are made by 'addition polymerisation', in which no chemical is split off. Although Vinyon fibres are very different chemically from nylon, they are similar in respect of their long chain molecular structure; this is the factor that determines whether a material will be fibre-forming or not.

Vinyon appeared very shortly after nylon. It was made by co-polymerising 88 - 90 per cent vinyl chloride with 10 - 12 per cent vinyl acetate. The polymeric chain is built up by heating these two substances together in the presence of a catalyst such as aluminium chloride. The reaction may be represented in this way:



When the chain has grown sufficiently long, the product will give fibres; at first these are only weak, brittle fibres, but when the chains correspond to a molecular weight of about 20,000 the fibres are strong and are spun into Vinyon. The fibre is spun not from the melt, but like cellulose acetate from an acetone solution. After spinning, it is stretched to confer a high strength on the fibre. The outstanding features of Vinyon are its extreme chemical and bacterial resistance, which make it suitable for filter cloths, protective clothing, insect nets, etc. Its low melting point and poor dimensional stability in hot water have militated against its general use for traditional textile purposes.

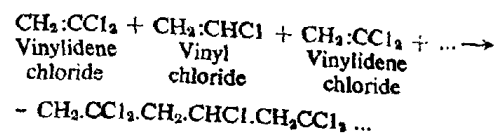
An improved form, known as Vinyon N, has more recently been made by co-polymerising vinyl chloride and vinyl cyanide (also known as acrylonitrile). The reaction is:



Vinyon N has a higher melting point and much better dimensional stability than the earlier Vinyon, as well as a high strength of 4 grams per denier, and will probably find a wider range of outlets. Because of their extreme resistance to moisture, these fibres are as strong wet as dry, but are very difficult to dye. Perhaps the most satisfactory method of colouring them is to incorporate a pigment in the polymerisation mixture. Vinyon polymer is made by Carbide and Carbon Chemicals Corporation and spun by the American Viscose Corporation.

Saran

Saran is a co-polymer made by the Dow Chemical Co. from vinylidene chloride and vinyl chloride. Both these intermediates can be made from cheap sources - sea water and petroleum. The reaction is again one of addition polymerisation:



When the chain has grown so that its molecular weight is 20,000 the product is saran. It can be spun from the melt into a yarn known as Velon, which is characterised by great resistance to chemical and biological attack and which, on account of its high chlorine content, is practically non-inflammable. At present, Velon is slightly yellow and on exposure to sunlight darkens slightly. It has found considerable use for upholstery, particularly for motor-cars. Like Vinyon, it is extremely resistant to water, difficult to dye by ordinary methods and probably best coloured during polymerisation.

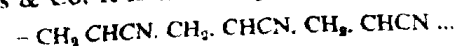
SYNTHETIC FIBRES

Pe Ce

It will have been noted that saran has a very high chlorine content. A product that in this respect was similar was made by the Germans. It was called Pe Ce and was apparently used (in staple form) for the manufacture of non-flam uniforms for the Luftwaffe flying personnel. The product was made by dissolving polyvinyl chloride in tetrachlorethane and chlorinating it until the chlorine content rose to 64 per cent. This increased chlorine content renders the fibre not only non-inflammable but also acetone-soluble, and it was spun from acetone solution.

Orlon

This new fibre has recently been announced by Du Pont de Nemours & Co. It is essentially a polyacrylonitrile



and is spun from solution. Apparently its manufacture was delayed, pending the discovery of a suitable solvent from which it could be spun. Solvents for the polymer are now stated to include dimethyl formamide and *m*- and *p*-nitrophenols. It has been claimed that in continuous filament form it is silk-like and in staple fibre it is wool-like. Its dry strength is 4.5 grams per denier, wet strength 4.3 grams per denier, and looped strength 3.8 grams per denier; its elongation is about 20 per cent; its elastic recovery is 85 per cent after a 4 per cent elongation. It 'sticks' when ironed at 190-200° C., which is rather low. It is a light fibre with a specific gravity of 1.18. It is suitable for outdoor purposes as even after one and a half years' outdoor exposure it retains 77 per cent of its strength. It has fair chemical resistance; not so good in this respect as nylon, it is also inferior in resistance to abrasion. Uses to which it is expected to be put include: window curtains and shades, umbrella fabric, rainwear, outdoor jackets and sports clothing, dress shirts, blankets, ties, woven lingerie and tricot fabrics, filter fabrics, tents, tarpaulins and awnings,

car tops, outdoor furniture, nets and cordage, and mixed with wool in socks. This is an ambitious list and probably time will show that it has peculiar advantages for only some of these purposes.

It will be seen that although it is only eleven years since nylon – the first synthetic fibre – was introduced, rapid development has been made in this field. Several other fibres are already important from the industrial and commercial standpoints; thousands of different synthetic fibres have been made experimentally, and it is safe to predict that there will be more new fibres produced on an industrial scale within the next few years. It seems possible that specified physical properties can be obtained by suitably selecting the intermediate products from which the polymer is to be made. One fibre will be best for one purpose, another for a different purpose, and so on. For any particular use it should soon be possible to make the ideal fibre.

The Calendar

W. A. PRICE

SIMPLIFICATION of the Calendar has often been proposed, and the proposal is being pressed at the present time. A note on the history and construction of Calendars may be of use.

Sidereal time depends on observations of the passage of stars across the field of a transit instrument. These form a perfect regulator for such clocks as those which drive the equatorial telescopes, and less directly for other time-pieces. The interval between two passages of any one star is a sidereal day, divided into hours, minutes and seconds: and other periods are comparable with one another by their measures in sidereal time, an absolute scale for the purpose.

The Calendar is concerned with the Sun's movements and depends on two periods, the Solar Year and the Solar Day, both of which can be determined in sidereal time. The solar year is the period in which the Sun describes its whole path in the Ecliptic, and returns to the same point as seen against the celestial map. It is constant from year to year, and the equinoxes occur at fixed sidereal times. The solar day is the period between two successive passages of the Sun across the Transit instrument. It varies from day to day, indeed no two solar days are exactly equal; but an average period called the Mean Solar Day (M.S.D.) can be calculated and is constant. Our clocks and watches show 24 hours to each M.S.D., and the Sun is always due south of any place at about 12 o'clock, midday, noon of local time. The solar year is 365.2422 mean solar days, i.e., 365 days, 5 hours, 48 minutes, 46 seconds of our clocks. The Calendar is a numbered list of mean solar days and with a number

denoting the year provides a date for any event past, present or future.

Among such events are the solar equinoxes; and the dates of these should be, as nearly as may be, the same in the calendars of the several years. Seeing that the lengths of the solar year and the mean solar day are incommensurable, to effect this was a problem that needed solution. The Greek astronomer Sosigenes gave the answer to C. Julius Caesar. Since 365 days is a little less than 1 solar year, and 366 days a little more, two kinds of year must be used, viz. one of 365 days (a common year) and the other 366 days (a leap year) respectively. Some arrangement of these should reduce the annual displacement of the dates of the Equinoxes to a small amount. The following systems have been used or proposed, the average displacement of the equinox in each case being under the heading. Mean annual error.

				Mean annual error
(1)	1 leap year in every	4 years	..	11 min. 14 sec.
(2)	7	29	..	1 min. 8.3 sec.
(3)	8	33	..	19.4 sec.
(4)	31	128	..	1.08 sec.
(5)	97	400	..	25.9 sec.
(6)	132	545	..	Imperceptible.

Arrangement (1), the Julian Calendar, arranged by Sosigenes, was started on January 1, 46 B.C.; and subject to correction at two subsequent dates has continued to be followed in Western Europe to the present day.

The accumulated error was corrected in 325 A.D. by Constantine the Great at the First Council of Nicaea, three days being omitted from the calendar.

Errors were again corrected in 1582 by Pope Gregory XIII (a step in the counter-Reformation) who directed that 10 days were to be omitted from the calendar, and that after the year 1600 three of the leap years of the Julian system in each period of 400 years should be common years, viz. the years 1700, 1800 and 1900; 2000 being a leap

THE CALENDAR

year as before. This system, (5) is the Gregorian Calendar. Instituted by the Roman Curia in 1582, it was not adopted in England until 1752, when 11 days were omitted from the English Calendar. Old Style still operates in some ways. The common fields of the manor used to be opened for grazing on Lammas Day, August 1. Under New Style this became August 12, on which date the moors are still available for grouse shooting.

Under Old Style the financial year began and ended on March 25. New Style made this April 5 as it is now.

Though systems (3) and (4) have advantages over the Gregorian system (5), the last serves all practical purposes, and change would be troublesome. System (6), under which the annual error practically vanishes, has no useful significance. Systems (2) and (3) were invented by a Persian, Omar Khayyam, the famous mathematician, astronomer, philosopher and poet. He was appointed to be State astrologer by the Seljuk conqueror Alp Arslan, and the Persian calendar of his construction was introduced by the Sultan of Khorassan about 1078. In this system three epochs of system (3) were followed by one epoch of No. (2); the whole forming the remarkably exact system (4). By what means Omar, having neither telescopic sights nor modern clocks, was able to observe so accurately is not known. The Transit Circle was his one instrument of precision.

BOOKS RECEIVED

PHYSICS TELLS WHY by Overton Luhr (Allen and Unwin, 1947) 16/-.

Most books on Physics for the layman concentrate on the probings of the atomic nucleus, or on the philosophical implications of Indeterminacy, or of an expanding Universe. Here is one which instead sets out to explain the way things work – telescopes, polaroid filters, rockets, radio, musical instruments, motor-bikes, thunderstorms, and so on; and does it excellently. A great deal of it covers ground already trodden by a General Science course in schools, and if its price makes it too expensive as a textbook, it should at least find a place in the library.

J.L.C.

ATMOSPHERIC TURBULENCE by O. G. Sutton (Methuen's Monographs on Physical Subjects, 1949) 16/-.

Professor Sutton is already well known to regular readers of *Science News*, and this book covers in more detail and with the precision aid of mathematics the subject matter of his recent article *The Restless Wind*. It is likely to be of interest to meteorologists and physicists, rather than to scientists in general, but should be within the reach of senior science students in sixth forms, if of rather specialised appeal.

J.L.C.

About Our Contributors

W. A. Cobb qualified medically in 1936 and trained as an anaesthetist. During the war he worked in a head-injuries hospital, and, not having enough to do, began the study of E.E.G. in 1942. He is now working in the Department of Applied Electro-Physiology of the National Hospital for Nervous Diseases, Queen Square, London.

R. Howard Cricks, son of a pioneer of cinematography, is a cinematograph engineer and consultant, and has long been keenly interested in the relations between television and the film. He is a Fellow of the Royal Photographic Society and of the British Kinematograph Society, to which society he is technical consultant, and editor of its journal, *British Kinematography*. He is technical editor of the *Ideal Kinema*; technical correspondent to the *Kinematograph Weekly*; and author of a standard textbook, *The Complete Projectionist*.

Felix d'Herelle was born in Montreal in 1873, educated in Paris and trained as a doctor at Montreal University. He was Professor of Bacteriology in Guatemala (1901-6), bacteriologist to the Mexican Government (1907), and then joined the Pasteur Institute in Paris. Professor at Leyden (1921-3), and at Yale, U.S.A. (1926-33), he retired to Paris, where he continued working until his death in February, 1949.

Sir Harold Spencer Jones, F.R.S., has been Astronomer Royal since 1933. He was educated at Latymer Upper School, and Jesus College, Cambridge, where he took both the Mathematical and the Natural Science Triposes, and was then elected Isaac Newton Student and Smith's Prizeman, and made a Fellow of his College. From 1913-23 he was Chief Assistant at the Royal Observatory, Greenwich; and from 1923 H.M. Astronomer at the Cape of Good Hope. He is the author of many scientific papers; and of *General Astronomy*, 1922; *Worlds Without End*, 1935; *Life on Other Worlds*, 1940.

R. W. Moncrieff was born in 1903 at Preston, and graduated at Manchester. He was employed from 1928 to 1945 as a research chemist by British Celanese Ltd., in whose service he invented Fortisan; from 1945-48 he was with Paton & Baldwins, Ltd. He has carried out original research on the synthesis of new fibres, and on the relations between the chemical constitution and physical properties of fibres. A book on *Mathproofing* is in the press.

R. D. Preston is a physicist who has worked in a botany department on the molecular structure of plants since 1929. In 1932-3 he was 1851 Exhibition Fellow and in the following year a Rockefeller Foundation Fellow at Cornell University, U.S.A. At present Reader in Plant Biophysics in the University of Leeds. He is old.

Fredrick William Twort, F.R.S., was, from 1909 till the destruction of the laboratories by a bomb in 1944, Superintendent of the Brown Institution and Professor of Bacteriology in the University of London. He is the author of numerous scientific papers, the discoverer in 1910 of the first vitamin (K) known to be needed by bacteria, and discoverer (in 1915) of the bacteriophage.

G. P. West is a veterinary surgeon and an associate member of the Guild of Agricultural Journalists. After six years in the Army and 18 months in the Foreign Office, he is now on the staff of the National Veterinary Medical Association.

R. W. G. Wyckoff was one of the pioneers of X-ray crystallography, and made early contributions to the theory of the subject; he has also compiled a systematic index of all known crystal structures. He worked at one time in the Carnegie Institute at Washington and at the Rockefeller Institute; at present he is at the National Institutes of Health at Bethesda, Maryland. His recent work has been in a field quite different from his earlier crystallographic studies, and in contrast to their severely theoretical nature, has involved most elegant experimental techniques, some of them developed by himself; it consists in the study of the structure of very small living organisms, such as bacteria, bacteriophage, and viruses, by means of the electron microscope.

J. L. Crammer is editor of Science News and a doctor and until recently was a Squadron Leader in the R.A.F. Medical Service. He was educated at Cambridge and University College Hospital and specialised in biochemistry.

GLOSSARY

ANISOTROPY: The possession of different properties in different directions in space; for example the property of expanding more in one dimension than another, asymmetrically, when evenly heated.

DAMPING: Rapid decrease in the amplitude (magnitude) of a vibration.

DENIER: A textile unit, the weight in grams of a 9000 metres length of a yarn or filament. For example, a 150 denier rayon yarn weighs 150 gm. per 9000 metres, or 60 metres of it weighs 1 gm. If the same yarn is composed of 30 filaments (usually written 150/30) each filament will be 5 denier. If a 150 denier yarn just breaks under a load of 300 gms., it is said to have a tenacity of 2 gms. per denier.

LYSIS: The fact of dissolving; thus *bacteriolysis*, the dissolution of bacteria, *lysed*, dissolved; *lytic*, having the property of lysis.

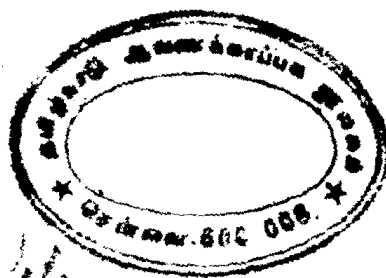
MILLISECOND: one thousandth of a second.

MUTANT: A new variety of living thing, with one or more new inheritable characteristics not possessed by its parents or forbears. The new feature appears suddenly, spontaneously, as a special change or mutation, or 'sport'.

POLYSACCHARIDES: A great class of complex chemical substances produced by living matter, characterised by the presence within each molecule of a number of combined sugar molecules, or their close chemical relatives, as can be shown by their isolation in pure crystalline form after hydrolysis (digestion, splitting up with hot acids, or with enzymes) of the given polysaccharide. Thus starch (which yields glucose), cellulose (glucose and other sugars), cartilage (amino-sugars), are three examples of the group.

PUSTULE: A tiny spot in the skin, a small pus-containing swelling.

TRACHEID: The elongated fibre produced by a plant cell of a higher plant. If it enlarges and becomes hollow, it forms a part of the water-conducting system of the plant; if it becomes greatly thickened until it is almost solid all the way through, it forms part of the hard woody skeleton of the plant.



2.01.14/6

01492/4

INDEX

Alpha rhythm of brain	122 <i>et seq.</i>	1
American television	14, 28, 30	
Amplifiers for E.E.G.	125, 136	
Amplitude modulation	10	
Bacteriophage, appearance of	62	
genetics and	64	
recent work	61	
reproduction of, nature	66, 68	
variability of	50	
Berger, Hans	121	
beta rhythm of brain	124	
blindness and E.E.G.	122	
brainwaves, technique for		
recording	127	
Cambium (new wood), nature of	104, 109, 110	
camera, mechanism of television	18 <i>et seq.</i>	
Carothers, W. H.	142, 148	
cellulose, structure of	95, 97	
cholera and bacteriophage	52, 54	
Distemper (canine)	71	
E.E.G. (electro-encephalogram)		
and blindness	122	
difficulties in	128	
and epilepsy	131 <i>et seq.</i>	
frequency analysis of	138	
records	123, 129, 132, 134, 135	
and twins	131	
Fits, epileptic, and flashing		
lights	137	
foot-and-mouth disease	69	
fowl typhoid and		
bacteriophage	53	
Hard pad disease of dogs	71	
Kant	75	
Laplace	76	
life, nature of	48	
Modulation, amplitude		
moon and the weather	116-11	
'Neurotropic', meaning	71	
nucleic acids	6	
Nylon	144-149	
Orlon	151	
Pad (dog's) disease of	71	
Pe Ce (synthetic fibre)	151	
proteins, structure of	95, 96, 141	
Research, nature of scientific	114	
Saran (textile fibre)	151	
scanning, sequence of events	17	
in	24	
Schmidt optical system	24	
Scophony television process	24	
solar system, views on,		
of Alfvén	8	
of Hoyle	8	
of Jeans and Jeffreys	70	
of Kant and Laplace	75	
of Lyttelton	87	
of von Weizsäcker	88	
streaming in plant cells	107	
super-polymers	142	
Swedenborg	75	
Television camera	18, 20	
colour in	3	
films and	2	
large screen	24	
limitations of view	1	
nature of radio signal in	11	
scanning in	17	
Terylene		
Theta rhythm of brain		
Valonia seaweed	9	
variability of bacteriophages		
Vinyon (synthetic fibre)	1	
Vitamin K, discovery of	36,	
Wave guides	1	
weather, the moon and the	11	

[illegible]

Pelican Books

NEW TITLES AND REPRINTS

NOW AVAILABLE

DIGGING UP THE PAST - <i>Sir Leonard Woolley</i>	A4
RELIGION AND THE RISE OF CAPITALISM - <i>R. H. Tawney</i>	A23
THINKING TO SOME PURPOSE - <i>Susan Stebbing</i>	A44
BALLET - <i>Arnold Haskell</i>	A122
OPERA - <i>Edward J. Dent</i>	A150
LOCAL GOVERNMENT IN ENGLAND AND WALES - <i>W. E. Jackson</i>	A162
THE PYRAMIDS - <i>J. E. S. Edwards</i>	A168
THE WORLD'S WEALTH - <i>W. G. Moore</i>	A173
THE INVENTOR AND HIS WORLD - <i>H. Stafford Hatfield</i>	A178
MAN THE UNKNOWN - <i>Alexis Carrel</i>	A181
BRITISH HERBS - <i>Florence Ranson</i>	A183
MUSIC 10! - <i>Constant Lambert</i>	A195
THE ROMANS - <i>R. H. Barrow</i>	A196
BEFORE PHILOSOPHY - <i>Henri Frankfort, Mrs Henri Frankfort, John A. Wilson and Thorkild Jacobsen</i>	A198
THE ARCHAEOLOGY OF PALESTINE - <i>W. F. Albright</i>	A199
POTTERY AND CERAMICS - <i>Dr Ernst Rosenthal</i>	A201
HORSE-RACING - <i>Dennis Craig</i>	A203
THE SYMPHONY - <i>Ed. Ralph Hill</i>	A204
PRINCIPLES OF SHAKESPEARIAN PRODUCTION - <i>C. Wilson Knight</i>	A210
THE MEANING OF ART - <i>Herbert Read</i>	A213
SAILING - <i>Peter Heaton</i>	A214