42

The Locust Problem
Developments in Intelligence Testing
Tension Reduction and Tension Seeking
Vapour Chromatography
The Electrical Constants of Free Space
Semi-Conductor P-N Junction Devices
Solid Lubricants

2/6

SCIENCE NEWS

EDITED BY ARCHIE AND NAN CLOW

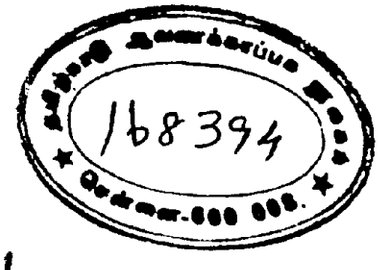
42

PENGUIN BOOKS

Penguin Books Ltd, Harmondsworth, Middlesex
U.S.A.: Penguin Books Inc., 3300 Clipper Mill Road, Baltimore 11, Md
CANADA: Penguin Books (Canada) Ltd, 178 Norseman Street, Toronto 18, Ontario
AUSTRALIA: Penguin Books Pty Ltd, 762 Whitehorse Road, Mitcham, Victoria
SOUTH AFRICA: Penguin Books (S.A.) Pty Ltd, Gibraltar House, Regent Road, Sea Point, Cape Town

—

First published November 1956



Made and printed in Great Britain
by The Campfield Press, St Albans

CONTENTS

18 illustrations are between pages 64 and 65

SOLID LUBRICANTS

E. R. BRAITHWAITE

THE term 'lubricant' brings to mind the seizure of car engines, squeaking door hinges, stubborn lawn mowers, and many other frictional nuisances in our everyday mechanical environment. However, it is well to remember friction is both a nuisance and a blessing, and that life would be rather difficult and peculiar without it. The absence of friction would mean that we should be unable to climb hills or utilize any form of transport. Nature has provided cheap and plentiful lubricants such as air, water, grease, and so on, for an easy-going life, but has not legislated for the high speed unnatural mechanical age in which we find ourselves today.

Friction

The concept of friction and its general behaviour was first introduced in the fifteenth century by Leonardo da Vinci, who found that friction is proportional to load and independent of the area of the sliding surfaces. Later, in the seventeenth century, the laws were rediscovered by Amontons. Coulomb, probably better known for his discovery of the fundamental law of electrostatics, verified the laws experimentally, and distinguished between static friction (the force required to start sliding) and kinetic friction (the force required to maintain sliding). In more recent times the late Sir William Hardy put the study of lubrication on to a firm scientific basis, and still more recently the field has been broadened and soundly established as a new fundamental branch of science, the physics and chemistry of rubbing surfaces, by Dr F. P. Bowden of the University of Cambridge.

In recent years new tools, such as electron diffraction and electron microscopy, have made possible a detailed study of surfaces. Bowden has shown that the friction properties of moving

surfaces depend largely on the contours and profiles of the surfaces. The initial topography of a metal surface does not determine the ultimate contact area because local pressures on the asperities is sufficiently great to cause extensive elastic deformation. Local welding, and hence wear, will therefore occur during the deformation process unless there is adequate protection between the deforming junctions.

Lubrication

At a very early stage in his development man learned to appreciate the value of lubricants, but up to this century their mechanism was not properly understood since it is so intimately connected with the nature of surfaces, and the necessary tools were not available for microsurface studies till quite recently. Under normal running conditions liquid lubricants are effective and we refer to their action as hydrodynamic lubrication. In such conditions the moving surfaces are kept apart by an oil film of finite thickness whose flow is governed by the laws of hydrodynamics. As bearing pressures are increased or the viscosity of the fluid lubricant lowered by an increase in temperature, the lubricant is squeezed from between the surfaces until only a monomolecular film remains. This limiting condition, known as 'boundary lubrication', can be upset by a shock to the running system which may rupture the thin protective film and cause seizure of the surfaces. Under such conditions solid lubricants can be used to advantage as there is no danger of naked surfaces being exposed due to film breakdown.

MECHANISM OF SOLID LUBRICATION

Solid lubricants are materials that will separate two moving metallic surfaces under boundary conditions and decrease the rate of wear. They are usually applied in the dispersed state when their particle size is so small that they pass through filters, and the choice of carrier liquid is almost unlimited. Graphite and molybdenum disulphide, which are the best-known solid lubricants, have also the advantage of being able to form oriented layers on metal surfaces during rubbing which not only

protect the rubbing surfaces if an oil failure occurs, but also lower the surface tension of the metal surface and encourage oil-spread.

Solid lubricants function sacrificially in that they cut down metal wear by wearing away themselves, and their efficiency is governed largely by the presence or absence of contaminants in an analogous fashion to their own action between two metal surfaces (Figure 1). A solid that has to provide a measure of

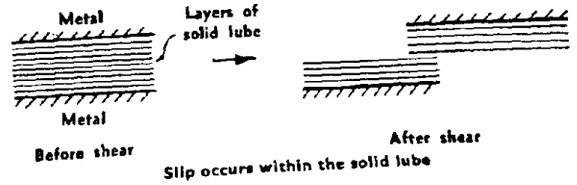


Fig. 1. Diagram illustrating the slip that takes place within a solid lubricant.

lubricity must, therefore, be strongly attached to a metal surface and at the same time possess a regular crystalline structure which can be deformed by a shearing force not greater than the force that binds it to the metal ($F_A \geq F_S$). The lubricating action of graphite is a typical example of such a mechanism. The element carbon exists in three allotropic modifications, amorphous carbon, graphite, and diamond. Graphite is the most stable allotrope and it is made up of parallel platelets of carbon atoms arranged in hexagons. These polyhexagonal platelets, stacked one on top of the other, are held together by the residual valency forces of the carbon atoms (F_S) and these interlayer forces yield easily to the external shear force transmitted by the moving metal surfaces. Such an action can be represented by a simple diagram such as Figure 2 where the moving surfaces are represented as flat. The mechanics of the movement of these platelets in a vacuum is, however, more complicated, and a satisfactory mechanism has not yet been established.

Solids that have been exposed to the air can slide over one

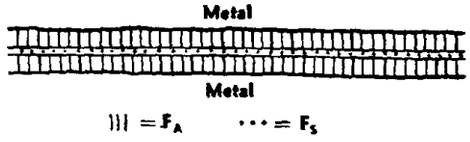


Fig. 2. Forces involved in lubrication.

another owing to the 'ball-bearing' action of the natural impurities adsorbed on to their surfaces, but if the surfaces are chemically clean, i.e., prepared in vacuum, the irregular molecular forces cause adhesion to take place at any points that are in sufficiently good contact. Such surfaces are unable to slip over one another without damage, which is caused by the interlocking of surface asperities; the resistance to easy motion is called friction. It was shown experimentally by Gregory[1] that whilst the welding of surface asperities still takes place in the presence of boundary lubricants, it is not so marked as with unlubricated surfaces. More recently Pratt and Bailey[2] have prepared molecularly flat surfaces on mica and have shown that under those conditions the classical laws of friction do not apply. This is, in fact, good confirmation of the classical laws of friction which are based on the fact that rubbing surfaces are very uneven. The magnitude of the asperities normally present on highly machined surfaces is quite large as will be seen from Table I.

It is not generally appreciated that extremely high temperatures are reached during the rubbing together of metal surfaces.

TABLE 1*

Nature of Surface	*Height of Asperities*
Polished material	0.1 micron
Honed material	1.0 ,,
Milled or turned material	10.0 ,,
Cast material	100.0 ,,

*From Zwikker: *Physical Properties of Solid Materials* (Pergamon Press, 1954, 267).

For example it has been reported that a gramophone needle rubbing against a wax record may reach at the point of contact a temperature of 1,200°C. Bowden and Ridler demonstrated this temperature effect by rubbing together two metals which functioned as thermocouples.[3] They found that mild steel rubbing on constantan under a load of 170 lb per square inch and rotating at 32 ft per second showed a rise in surface temperature of 600° C or more (Inset 9). Various combinations of metals were tried and it was subsequently shown that the surface temperatures were inversely proportional to the square root of the heat conductivities of the metals. Temperatures of this order can obviously have a deleterious effect on mineral oils (which begin to coke at *c.* 200° C) and also have a marked effect on the surface structure of the metals.

Temperature is an important factor in cylinder lubrication whether the engine is driven by petrol, oil, or steam, and at certain parts of the heat cycle conditions in certain engines may exceed the optimum temperature for oils. For example, in a two-stroke engine the coincidence of maximum pressure with maximum expansion temperature makes conventional lubrication difficult and one should therefore use colloidal graphite, which

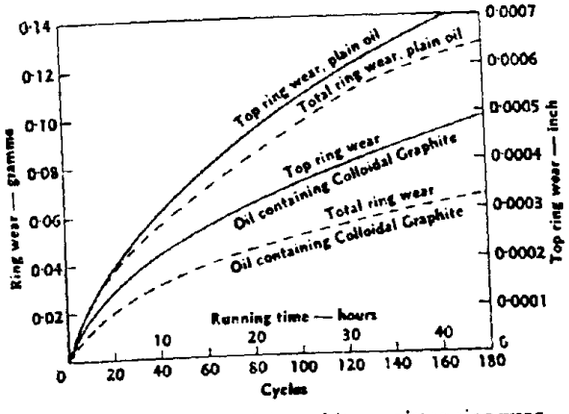


Fig. 3. Effect of colloidal graphite on piston ring wear.

will resist oxidation up to about 500° C and will give adequate lubrication at the top of a cylinder in I.C. engines where boundary conditions predominate. Tests carried out by the Institution of Automobile Engineers[4] have shown that sump oil containing colloidal graphite reduces the cylinder wear by a half on a new engine on which repeated cold starts have been made (Figure 3).

Not only is there a rise in temperature when shearing occurs, but at the same time metal fragments are transferred from one metal surface to the other as for example with engineering tools. Bowden, Williamson, and Lang demonstrated this qualitatively using a radioactive tracer technique. Inset 10 shows an autoradiograph of a sheet of copper which has been hit with a high-speed steel hammer that had been made radioactive. The random orientation of the metal transferred during this operation shows that transfer occurred only at the asperities or high spots of the surface.

Earlier it had been shown that adsorbed films of gas may act as natural lubricants, and this is also true of natural oxides, though the friction of rubbing surfaces contaminated with surface oxide depends largely on the relative hardnesses of the oxide and the metal. For example, with a hard oxide film on a soft metal substrate, the slightest load will deform the metal through the oxide, its behaviour being analogous to an eggshell when cracked. When, on the other hand, the mechanical properties of both metal and oxide are similar, as with copper and copper oxide, the surface film will deform with the underlying metal, thus promoting damage to the metal surfaces. Recently, workers in the United States have studied the effect of the oxides of lead and cadmium on the lubricating qualities of graphite at temperatures up to 1,000° F provided the oxides do not react with the graphite. These tests showed that under particular experimental conditions mixtures of graphite and metallic oxides are effective lubricants even at temperatures above that at which graphite oxidizes.

Adsorbed layers markedly influence the coefficient of friction. For example, the coefficient of friction of platinum varies enormously, depending on whether the surface is neutral or whether it has hydrogen or oxygen adsorbed on it. The hydrogen and oxygen surfaces display a minimum frictional force, while the gas-free surface has a high coefficient of friction. When studying the behaviour of solid lubricants themselves it is not surprising therefore that research workers have given a great deal of thought to the effect of impurities on their lubricity.

The first real attempt to investigate the effect of gaseous surface films on friction properties was made by R. H. Savage using graphite, boron nitride, and molybdenum disulphide.[5] He showed that the friction of clean graphite in vacuum is very high, but that lubrication is effected instantly when any one of a number of polar gases is admitted to the friction apparatus at low pressure. Water vapour is particularly effective in this connection, its function probably being to act as a 'boundary lubricant' between the graphite layers, thus increasing the c–c distance and decreasing the shear. It would appear that many

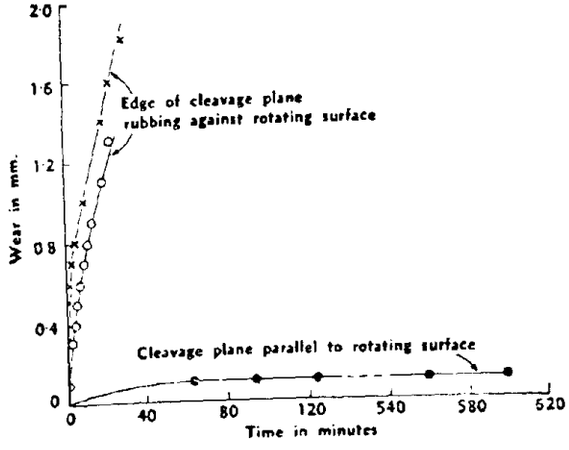


Fig. 4. Graphs showing wear against time in graphite brushes cut from a specimen of highly orientated natural graphite.

people have read Savage's excellent introduction on the effect of contaminant films on graphite and have ignored the later work arising out of it. It is a fact that when a piece of graphite is rubbed in vacuum, the dust particles liberated are abrasive and this is not surprising when we take into account the fact that these fine particles are chemically clean and possess very strongly adhesive forces which cause them to aggregate in a disoriented fashion. Disorientated graphite is not a lubricant. If, however, graphite is first applied to a surface in air, it orientates itself with basal planes parallel to the rubbing surfaces and forms a graphoid layer which allows lubrication even under high vacuum conditions. These orientated layers of graphite are so firmly attached to the rubbing metal surfaces that repeated solvent washes and abrasion with emery paper will not remove them. Their lubricating action is illustrated in Figure 4.*

It appears that molybdenum disulphide acts differently from graphite since its lubricity increases with decreasing vapour pressure, and it has been suggested that this may be due either to the removal of water vapour or to the continuous evaporation of some contaminant such as sulphur which will lubricate the laminae.

The Choice of a Lubricant

Before passing to a consideration of specific solid lubricants and their use in industry it would be well to consider two questions: what are the characteristics of a solid lubricant, and why use a solid lubricant? The answer to this first question is given by the following list of the characteristics of a solid lubricant, from which it can be seen that a solid lubricant must fulfil many conditions:

Hardness/softness; particle size; chemical stability; heat stability; ability to withstand high pressures; physical and chemical nature of surface; presence of adsorbed films on surface; adsorbed internal films; porosity; relation to type of metal which is being lubricated; crystalline form.

The second question is more straightforward and can be divided into four headings.

Why use a solid lubricant?

(*a*) Lubricating oils containing solid lubricants show improved lubricating qualities.

(*b*) Chemical stability of lubricant molecules up to high temperatures is of vital importance. The extremes of temperature met in various running parts of modern aircraft demand an almost impossible degree of elasticity from any liquid lubricant. We have already seen that high local temperatures are produced by rubbing surfaces and, therefore, oils must be stable up to 600° C or more. Silicones and various fluorinated hydrocarbon oils have, of course, been developed to try and cope with this inherent weakness of hydrocarbon oils. These, however, have not yet met with much success for economic reasons, lack of inherent lubricity, and chemical instability on prolonged use. Failing a marked advance in the liquid lubricant field, we must continue to treat solid surfaces in such a way that they will prevent the latent surface forces binding together at the moment of breakdown of boundary films. Not only will a solid lubricant step into the breach when a liquid lubricant film breaks down owing to chemical and temperature instability, but it will also prevent seizure if, for some reason or other, the supply of oil to the system is temporarily or permanently discontinued.

(*c*) Solid lubricants in the form of 'paints' are often desirable

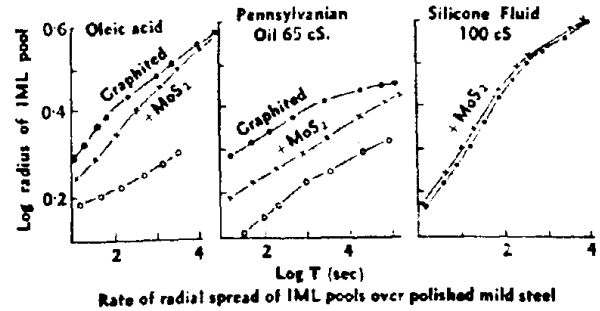


Fig. 5. Rate of radial spread of IML pools over polished mild steel. (*Reproduced by permission of R.A.E., Farnborough.*)

on surfaces where a liquid lubricant cannot be tolerated. Such difficult conditions include extremes of temperature, or systems that do not allow easy entry or flow of liquid or which may be exposed to dusty atmospheres where a liquid lubricant would pick up dust and work it into a grinding paste.

(*d*) Solid lubricants like graphite or molybdenum disulphide, when adhering to a metal surface, display a lower surface tension towards oil than does the metal surface itself. This, of course, enables the oil film to spread more easily and gives added protection if the film is ever in danger of excessive overloading (Figure 5).

TYPES OF SOLID LUBRICANT

Graphite

The first synthetic solid lubricant was made by Dr E. G. Acheson who in 1896 invented electric furnace graphite of very high purity and introduced it into the engineering world as 'Oildag'* (colloidal graphite in oil) and 'Aquadag'* (colloidal graphite in water).[7] Our concept of the structure of graphite has changed little since 1924 when J. D. Bernal showed that it consisted of hexagonal rings of carbon atoms arranged in layers and stacked so that repetition occurs after every two base layers (Figure 6). Interatomic distances are an indication of the bond strength, the shorter the distance the stronger the bond. The short c–c distance of 1.42Å indicates very strong bonding within the layer. The large distance of 3.35Å between layers indicates that the layers are held together loosely and can therefore readily slide over one another. This is one of the reasons why graphite is a good lubricant. Also, these strong binding forces in graphite account for the extremely non-volatile character of graphite. In graphite, each carbon atom within the layer operates three valence electrons so that three σ bonds are present at each carbon. Since the carbon atom possesses two 2s and two 2p electrons, one of the s electrons is promoted and as a result of the subsequent hybridization a bonding π electron is available

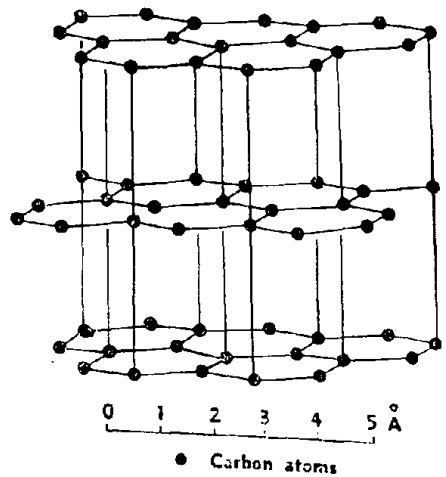


Fig. 6. Structure of graphite.

*Registered trade-mark (Acheson Colloids Limited).

TABLE II

Polymer	*Chemical formula*	*Characteristics*	*Brinell hardness*
Teflon	$[-CF_2-CF_2-]_n$	(I.C.I.)	1–2
Polythene	$[-CH_2-CH_2-]_n$	Alkathene grade 20 (I.C.I.)	1–2
Polystyrene	$\left[\begin{array}{l} C_6H_5 \\ -CH_2-CH- \end{array}\right]_n$	Molecular weight (a) 130,000 (b) 90,000 (c) 66,000	20–25
Perspex	$\left[\begin{array}{l} CH_3 \\ -CH_2-C- \\ COOMe \end{array}\right]_n$	(a) plasticized (b) unplasticized containing titanium oxide filler	25–30

From Bowden and Tabor *Friction and Lubrication of Solids*, (Clarendon Press, Oxford).

which gives graphite its conductivity, metallic lustre, and opacity.

The problem of providing lubrication for jet engine bearings has become a critical factor in the development of the turbojet engine. Liquid lubricants limit the operating temperatures of these engines and the design of efficient coolant systems is often difficult. It has been shown that at such high temperatures graphite is superior, in either air or nitrogen, both to normal hydrocarbon lubricants, and to silicon-diester blends. Graphite also shows superior lubricating qualities in die stamping, forging, and wire drawing. It is a tribute to the efficiency of colloidal graphite as a lubricant that after thirty years all the tungsten filament wire used for electric light bulbs is still drawn through dies lubricated with colloidal graphite. Prof. Thompson[1] has reported favourably on the use of colloidal graphite for wire drawing with other metals. Numerous lubricants were tried for drawing mild steel wire up to 625° C and it was found that 'dag' colloidal graphite in white spirit was the most satisfactory lubricant and that it could be used even at temperatures of about 700° C. Colloidal graphite lubricants are also widely used in the form of 'paints' to give dry films where liquid lubricants cannot be tolerated.

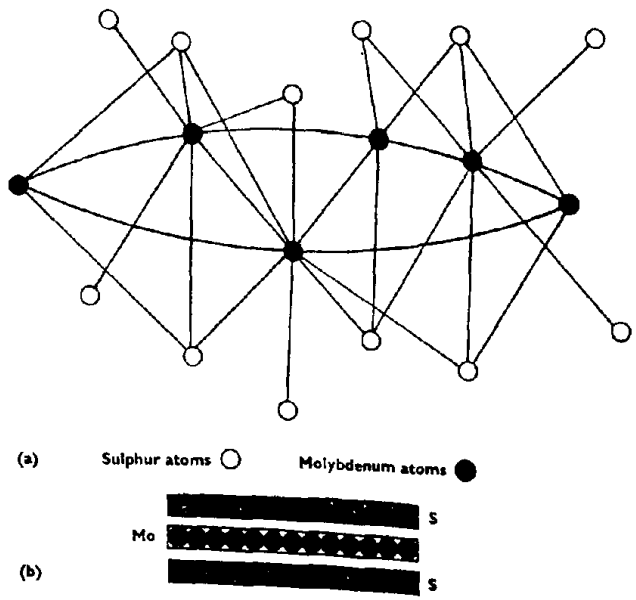


Fig. 7. (*a*) Structure of Molybdenum Sulphide, and (*b*) section showing layer of Mo atoms covered on each plane side by S atoms.

Molybdenum disulphide

Molybdenum disulphide (MoS_2) has a hexagonal lattice similar to graphite with a pronounced cleavage of the (001) planes (Figure 7). It is rather like graphite in appearance and this similarity together with its greasy properties was an early source of confusion between the two materials. Under ordinary conditions it is quite stable, but is subject to oxidation and weathers to a hydrated molybdenum oxide. Apart from surface chemical aspects, its lubricating action is dependent on the structure. Each layer of the compound is composed of a sheet of molybdenum atoms with a sheet of sulphur atoms on each side. The strong Mo–S bonds hold the layer strongly to the metal substrate whilst the weaker S–S bonds allow the laminae to slip easily over one another. The naturally occurring ore (Molybdenite) cannot be used because of the silica contaminant that escapes

TABLE III

Polymer	*Polymer sliding on polymer*	*Polymer sliding on steel*	*Steel sliding on polymer*
Teflon	0.04	0.04	0.10
Polythene	0.1	0.15	0.2
Polystyrene	0.5*	0.3	0.35
Perspex (*a*)	0.8*	0.5*	0.45*

*Intermittent motion.

From Bowden and Tabor, *op. cit.*

the first flotation process. This silica has to be removed by a second flotation, followed by heating to 600° C with either hydrofluoric acid or sodium bifluoride. Like graphite, molybdenum disulphide forms an orientated layer when rubbed on a metal surface either as a dry powder or in the dispersed state.

In some applications such as metal cutting, molybdenum disulphide is most effective as a lubricant when finely dispersed in an oil and it has been found that the stability of the dispersion is affected by the extent of the surface oxidation which causes flocculation of the primary particles, since the oxidized particles are hydrophilic. In all, colloidal molybdenum disulphide provides unusual characteristics under extreme pressure conditions at low or moderate temperatures.

Plastics

The frictional properties of plastics depend markedly on the type of plastic selected as illustrated by Tables 2, 3 and 4. It can be seen that the coefficient of friction for polystyrene and 'Perspex'* are of the same order as for metals, while the value for 'Teflon'* (polytetrafluoroethylene) and polyethylene is much lower. 'Perspex' and polystyrene soften below 80° C, but with Teflon there is little sign of thermal softening even at 200° C and the friction remains low throughout. It has been suggested that the relatively large fluorine atoms in the Teflon shield the smaller carbon atoms. Teflon fibres are stiff and do not interpenetrate, and this allows the long molecules to roll over one another. The fluorine balance in the molecule is important as the substitution of only one of the fluorine atoms by chlorine is sufficient to lower the lubricating efficiency. A dispersion of Teflon will not act as a boundary lubricant, while a Teflon coating on metal suffers from low mechanical strength and low thermal conductivity.

Recently, experiments have been carried out by Bowden and his school on the use of Teflon for ski-runners. The good performance of Teflon on ice is due not only to its low coefficient

*Registered trade-mark (Imperial Chemical Industries Limited).

TABLE IV

Sliding surfaces	20° C	50° C	80° C	100° C	150° C	200° C
Polystyrene on polystyrene	0·5*	0·65*	0·65–0·7*	..	..	..
Perspex on Perspex	0·8*	..	0·85*	..	..	..
Teflon on Teflon	0·04	0·04	0·04	0·04	0·04–0·05	0·04–0·05
Teflon on steel	0·04	0·04	0·04	0·04	0·04	0·04
Steel on Teflon	0·09	0·09	0·10	0·10	0·11	0·14

* Intermittent motion.

From Bowden and Tabor, *op. cit.*

of friction and low thermal conductivity but also to the fact that it is not wetted by water. In the near future an important use of Teflon might be its incorporation into aircraft skis for landing on arctic airfields. So far a mechanical difficulty concerning the fabrication of Teflon on metal has been the difficulty of fixing it to metal firmly without heat, but this problem at last shows signs of being solved since, on a laboratory scale, it can now be welded to metal by grafting with other polymers.

Glass

The idea of using glass as a lubricant might appear curious, for glass is a hard, brittle abrasive substance. However, glass has the right characteristics for the extrusion of steels at temperatures of about 1,000° C, where it softens over a reasonably wide temperature range and retains the required viscosity characteristics: it is also a good thermal insulator. Molten glass is the lubricant used in the Usine Sejournet process for extruding steel under conditions of high pressure and temperature where it facilitates flow on the metal which is being deformed.

* * *

In laboratories many other compounds are being investigated as potential solid lubricants, but they have not been discussed here because as yet they have not (with the possible exception of

indium) achieved industrial importance. Such compounds include the cadmium halides, silver nitrate, boron nitride, mercuric halides, and camphor. In the future, however, we may expect to see some of these being used as solid lubricants.

REFERENCES

1 *Nature* 1946 **157** 443.
2 *Proc. Roy. Soc. A.* 1955 **227** 500.
3 *Proc. Roy. Soc. A.* 1936 **151** 610.
4 Higinbotham : *Proceedings of the General Discussion on Lubrication and Lubricants* (Inst. Mech. Eng. 1937).
5 *J. App. Phys.* 1948 **19** 1.
6 *Trans. A.S.M.E.* 1948 **70** 491.
7 'Graphite and Silicon Carbide: their structure properties, and uses.' Paper 1209 (by the author) published by Inst. of Scottish Engineers and Shipbuilders.
8 *J. Iron and Steel Inst.* 1953 **173** 36.

THE LOCUST PROBLEM

P. T. HASKELL

MANY scientists have subscribed to the idea that a fitting label for the present era would be the 'Age of the Insects'. This is because insects, despite the great disparity between their evolutionary development and that of man, still dispute with him for control of vast areas of the earth. In number of species insects surpass all other animals put together, and in ecological adaptation probably cover a larger area of the world than any other group. For his livelihood and sometimes even for his life man is forced to wage incessant war on insects such as mosquitoes, tsetse flies, and locusts. Of these great plagues the locust has great claims to be considered both the worst and the oldest. Certainly since the beginnings of agriculture, locust plagues have been known to man; the earliest known pictures of locusts are those on an Egyptian tomb of the 12th Dynasty, about 2400 B.C., and many references to locusts and their depredations can be found in ancient Egyptian, Greek, and Chinese literature. The best known reference is probably that in the Bible, where locusts appear as one of the plagues of Egypt. This particular locust was the Desert Locust (*Schistocerca gregaria* Forsk.), and this insect is probably still the greatest menace, and presents the most difficult problem of all the plague locusts.

Locusts are members of the order Orthoptera, which includes crickets and grasshoppers. Locusts and grasshoppers are closely related in one group, the Acrididae, and there is no scientific basis for the use of the general term 'locusts'. But it is a very useful name for the Acrididae which have developed gregarious habits, and which migrate in compact groups or swarms, as opposed to the solitary and sedentary habits of other members of the family. Although locusts are the most harmful of the Orthoptera, it must not be imagined that their lesser relatives,

the grasshoppers, are completely innocuous. Although their menace is less spectacular because they do not form swarms that migrate over large distances, nevertheless they are capable of causing enormous damage to crops. Outbreaks of grasshoppers on a large scale are of common occurrence in America and Australia. In fact these pests can be found in all the five continents, and are only absent from the cold northern forests, the equatorial rain forest, and from high mountainous areas. Figure 1 shows the enormous area of the world subject to attack and invasion by locusts and grasshoppers.

In Europe we tend to think of ourselves as safe from locusts, yet its southern countries, particularly Spain and Italy and the Balkan States, are still subject to locust plagues. Africa has five locust species to deal with. North Africa has the endemic Moroccan locust and it also comes within the vast invasion region of the Desert Locust; then in the valley of the Niger is found the African Migratory Locust (*Locusta migratoria migratorioides*); in the Rift Valley in Tanganyika is the breeding ground of the Red Locust (*Nomadacris septemfasciata*). Finally, in the Karroo region of South Africa one finds the Brown Locust (*Locusta pardalina*). In Australia it is a locust (*Chortoicetes terminifera*) that causes immense damage yearly to crops, but grasshoppers are also present there. In the Americas, the United States and Canada suffer mainly from grasshoppers. Numerous species of the genus *Melanoplus* abound, and do enormous damage; for example, in 1952 in the State of Utah alone, grasshoppers did an estimated million dollars' worth of damage to crops. In Central and South America is found the locust *Schistocerca paranensis*, and the larger countries, such as Mexico and Argentina, suffer great crop losses due to its depredations. In the East, the Oriental Migratory Locust (*Locusta migratoria manilensis*) devastates crops in vast areas of China, Borneo, and the Philippine Islands, while in Pakistan and India we return once again to the easternmost part of the zone of the Desert Locust, which also covers Arabia, and indeed most of the African continent north of the Equator.

In this survey there is only space to mention the most pro-

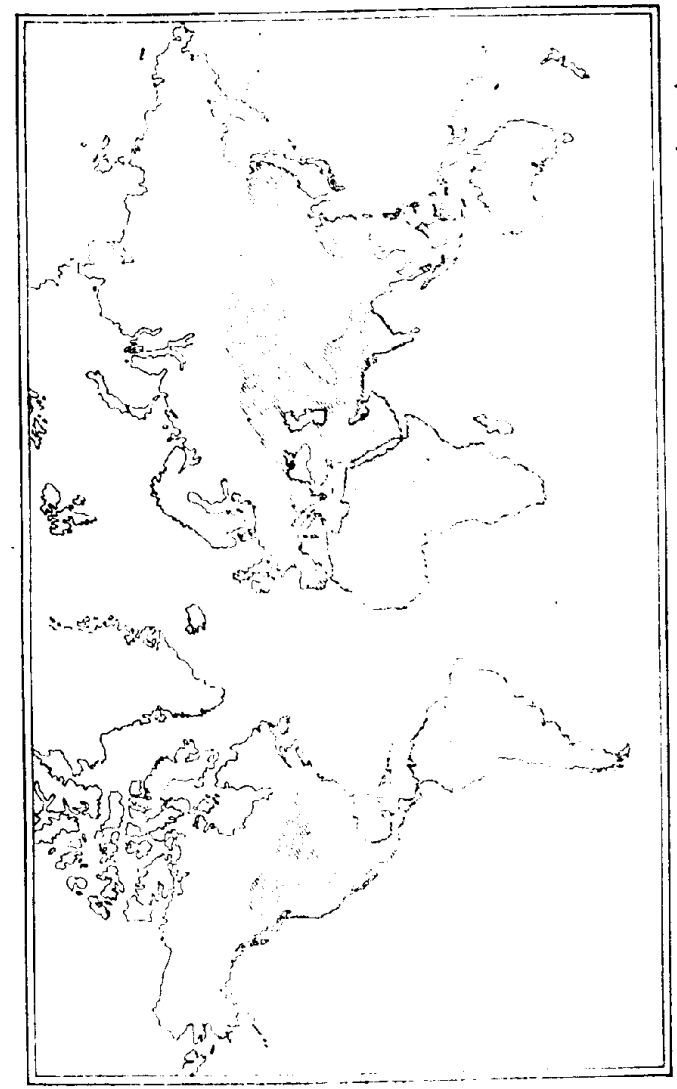


Fig. 1. Map of the world showing, by hatched areas, regions subject to invasion by locusts and grasshoppers.

minent of these pests; there are many other species that cause damage to crops, but their depredations are not on such a vast or dramatic scale as those of the species just given. It is very difficult to collect any reliable estimates of damage done by locusts throughout the world, but statistics gathered by the Anti-Locust Research Centre in London for the years 1925–34 indicated an annual loss of the order of £15 million. Today a conservative estimate would be double this figure; it should be noted that this estimate is an average figure for crop damage over the whole world, and that it does not take into account the extraordinary losses that may occur in particular areas invaded by locust swarms. Such a situation developed in Morocco where in a single season of 1954–5 the Desert Locust caused damage estimated at £4.5 million. The measure of the locust menace may be taken in another way, by noting that nearly every country in which crops are damaged by locusts or grasshoppers has found it necessary to set up an organization for research on and control of the pests. The lead in this respect was given by Great Britain, which in 1929 set up a Locust Sub-Committee of the Committee of Civil Research, and from this beginning has developed the present Anti-Locust Research Centre in London. The principal purpose of the organization is to act as a centre for collecting and exchanging information, and to co-ordinate research on locusts which is undertaken by various regional organizations. It has long been realized that efficient control measures against locusts can only be based on accurate information about the occurrence and biology of the various species, and one of the first results accruing from the setting up of the Centre was that of sending entomologists into the field to explore regions where locusts abounded and to study their biology.

Locust Life History

Locusts and grasshoppers have the same sort of life cycle; the female lays a number of egg-pods in the soil, each one of which may contain up to 100 eggs. According to the climatic conditions of the country, the eggs may lie dormant for the winter, and hatch in the spring; in tropical countries they may lie dormant

in a dry period or hatch in two to three weeks. The Orthoptera belong to the so-called hemimetabolous insects, that is, they do not undergo a metamorphosis as in butterflies or beetles, and the young on hatching resemble the adults, except that they are wingless and of course very much smaller. The nymphs of grasshopper species settle down and live where they hatch, and they move about very little. Although there may be many in a small area, they live independently of one another and show little gregarious behaviour. Recent research in this country by Richards and Waloff indicates that what movement of population occurs is in the main dependent on the changes in vegetation in the habitat.[1] The nymphs grow to adulthood by a series of moults, and when full grown most grasshoppers have two pairs of wings, and are capable of flight. This activity is not often seen in nature however, since most species use flight only as a means of escape from predators, and the distances covered are small, perhaps up to fifty yards. All this behaviour constitutes the main difference between grasshoppers and locusts. The nymphs of the latter, as soon as they hatch, show marked gregariousness and collect together in groups, or hopper bands as they are called. These bands, one of which is shown in Inset 11, may be several square miles in area and the most striking feature about them is their marching behaviour. The relentless march of hoppers, undeterred by natural obstacles such as ravines or water, is one of the fascinating problems of locust research, and more will be said later about this aspect of the work. As the hoppers march on, they undergo growth by moulting just as grasshopper nymphs do, until at last they become fully winged adults. The urge for movement still persists, however, and now the locusts form the vast swarms typical of the species, and migration by flying occurs. The swarms may reach enormous size (up to 200 square miles have been recorded) and contain perhaps a million million locusts, whose collective weight would total many tons. The flights of these swarms are no less amazing, distances of up to 2,000 miles having been recorded, although not, of course, in a single continuous flight.

ANTI-LOCUST RESEARCH CENTRE

It is obvious that these bare facts conceal dozens of biological problems of the first magnitude, and although there are many scientists throughout the world working on locusts, only some of these problems have been tackled. But of basic importance to all locust research is the Phase Theory, enunciated by B. P. Uvarov, now Director of the Anti-Locust Research Centre, as far back as 1921. This theory gave an answer to the question 'What happens to locusts when there are no swarms?' Now it was known that the Orthopteron called *Locusta migratoria* formed swarms; when these plagues died down, or even during them, it was possible to find what was apparently another species of locust, *Locusta danica*, living in the solitary fashion of grasshoppers as described above. These two insects were so different in morphology and behaviour that most specialists agreed that they were different species, as their names indicate. Uvarov, however, suggested that these two species were in fact one, but that this one was capable of existing in two forms, or *phases* as they came to be known. In the 'solitary phase', the insects behave like ordinary grasshoppers, taking no notice of neighbouring locusts and being sedentary in habit; but, if the numbers of these 'solitaria' insects increase in a restricted area, the locusts acquire strong gregarious tendencies, form swarms, and exhibit the marching and flying behaviour mentioned above. Such locusts are then called the 'gregarious phase'. The change from one phase to another is not of course rapid, and intermediate forms occur called the 'transiens' phase. The verification of this theory led to a radical change of policy in connexion with anti-locust measures. Previously, money for research and control would be willingly voted by the various governments concerned during a 'plague period', but during the so-called 'inter-plague years' little money was made available and so continuous observation and research was impossible. When it became evident however that the locusts were always present, and that the control of swarm outbreaks depended on a knowledge of the non-swarming populations present in various areas, permanent research

organizations began to be set up, and as has been said, Great Britain took a leading part in this new approach to the problem.

Cartographic Research

The need for accurate knowledge of the whereabouts and occurrence of locust populations led naturally to the cartographical research that is an important feature of the work of the Anti-Locust Research Centre today. Reports of locust activities reach the Centre daily from some forty different countries. The information refers to all stages of locust activity, ranging from the location and extent of egg-fields and hopper bands to the sizes and direction of movement of swarms. All this information is plotted separately on special monthly maps, and by comparing the maps a picture can be formed of the general situation. This information is used for two purposes. The first is to enable the Centre, by study of the dynamic situation, to prepare monthly forecasts of the changes that are likely to occur as known from past experience, and if necessary to issue warnings to countries or areas likely to be invaded by hopper bands or swarms. This service is, of course, directly related to control measures, and more will be said later about this. The second purpose of mapping is to provide a basis for research. The ultimate aim of all locust research is to investigate the occurrence, distribution, and migration of the harmful species: to have maps covering many years which record this information for many species is obviously a solid basis for investigations.

The maps, however, can only indicate where and when the locusts have moved; they do not reveal the causes of such movement, and it is in the investigation of these causes that a number of specialists enter the field. Taxonomists, morphologists, ecologists, physiologists, physicists, bio-chemists – the services of all these are required in locust research. Taxonomists and morphologists provide the basis for accurate ecological work, and are an essential to all research. Workers must be able to identify the species they find in the field, and a knowledge of their relationship to other similar insects is helpful in making comparative studies of behaviour. The provision of keys to identify the

various species in the field is also an essential requirement of ecology.

Locust Populations

Ecological work on locust populations is of great importance, and is one of the most promising lines of research now being carried out. Its purpose is to investigate the fluctuations in the size of populations and relate this if possible to climatic, topographical, faunistic, and botanical features of the environment. An essential part of ecology is the estimation of the numbers of insects present in a given area; this is a difficult enough task even with small populations of relatively static insects, but with locusts, special difficulties arise in dealing with the marching hopper bands and more especially with flying swarms. With hopper bands one method applicable to small bands which cover only a few thousand square yards is to kill the band with insecticide and then count the corpses. This gives a good estimate of numbers but of course prevents further study of the population. Other methods depend on random sampling combined with survey methods to estimate the size of the band during sampling. Yet another method which has been used with much success, is to capture a number of hoppers from a band, mark them with paint and return them to the band. Later, the band is resampled, and the number of marked locusts recaptured is noted. A statistical method based on this data can then be used to evaluate the size of the total population. The estimation of numbers of locusts in a flying swarm presents a most difficult problem. For one thing, accurate estimates of the area or volume of a swarm are practically impossible; even if these figures could be obtained, the further figure of locust density in the swarm, required to enable an estimate of absolute numbers to be made, is even more difficult to assess with any reasonable accuracy because of the continuous changes in density that occur during flight. The difficulties can be appreciated by looking at the picture in Inset 12 of dense low-flying locusts. One method which gives reasonable estimates is to catch the swarm in the early morning before it begins flying, and to drench a part of it with

insecticide. Sampling of the dead and estimation of the total area covered by the swarm then allows approximate estimates of numbers to be made. These methods may seem crude, but they represent a first quantitative approach to the problem, which is complicated by the often inaccessible nature of the terrain and the migratory habits of the insects. The accuracy desired of population estimates also depends on the use to which the figures are to be put; for control operations, figures relating to orders of magnitude of populations are sufficient, enabling quick practical methods to be used, but for research, more accurate figures are required and investigations into methods of providing these continue.

Another most interesting aspect of ecology is the relation of the fluctuations of population with the vegetation of the area, and this is particularly important where plague locusts are known to have restricted breeding grounds. A good example is the Red Locust, which breeds mainly in restricted areas near Lake Rukwa in Tanganyika and the marsh known as Mweru wa Ntipa in Northern Rhodesia. Investigation of changes in the vegetative cover of the Rukwa, caused by fire and drought, and fluctuations in populations of the Red Locust, show an interesting correlation, and raise hopes that vegetation control may go some way towards helping to control locusts.

The fluctuation of populations, however, is caused by many factors, which range from climatic control, fecundity of the insects, and the degree of parasitism they suffer in the field, to the reduction in numbers due to predators. These aspects of the problem concern the zoologist, physiologist, and the entomologist. A start has been made to the study of the natural enemies of the locust; these include such diverse agents as the micro-organism *Empiroa grylli*, and the fly, *Stomorhina lunata*, which parasitizes the eggs of locusts. Then again, great mortality is caused by birds both in hopper bands and swarms. Swarms of locusts always have their attendant flock of birds, frequently storks, which eat enormous numbers of the insects. It is not unknown for a stork to gorge so many locusts that it becomes too heavy for flight, and has to regurgitate some of its food before

being able to take off. All these manifold factors affect the size of the populations studied, and form an important factor in the control of the pests. But in order to take every advantage of these 'natural' controls, we must know more about how they operate.

The March of the Hoppers

We must now consider the chief characteristic of populations of locusts—their mobility. As mentioned above, both hopper bands and swarms of adults show this pattern of movement by day and resting by night, and of course it is this characteristic that makes the locust an international danger. The relentless marching of hopper bands often gives the observer the impression of high purpose, but closer investigation reveals a number of puzzling features. A reasonable expectation would be that the insects were seeking food, and that they move away from areas of scarcity because of competition for food. Nothing could be further from the truth, and hopper bands have been observed to march from fertile regions into the desert. One of the earliest findings on the marching of hoppers was that it depended largely on temperature, and would not take place when the weather was too hot or too cold. Problems of this sort are best tackled by combining work in the laboratory with work in the field and Dr P. E. Ellis, of the Anti-Locust Research Centre, has for some years been carrying out studies on hoppers on this basis. She early discovered that it was possible to make hoppers 'march' in the laboratory by putting them in a circular arena with a single light source overhead. In this situation, if environmental conditions such as temperature were correctly adjusted, they would march round and round the cage. In this way the various stimuli that initiate and maintain movement in the insects have been studied. It has been found that one of the principal causes of continued movement in hopper bands is due to the reactions of an individual to the sight of one of his companions; the insects tend to move so as to try to keep the images of their neighbours in a stationary position on the retina of the compound eye. This means that if one hopper moves forward, those

on either side of him also move forward to stabilize their retinal images. Other factors may also play a part, such as olfaction and auditory stimuli, but visual stimuli seem to play the most important role in this behaviour.

One may object that this is all very well in explaining why hoppers move continuously, but what starts them marching? Locusts are of course cold blooded, so that during the night when the temperature falls they become immobile; they roost on the tops of bushes without feeding, and only a rise in temperature can stir them to activity, because their muscles can only work within a certain range of temperature. A hypothesis has been put forward by Ellis and Hoyle to explain the initiation of movement in hoppers.[2] During their enforced rest at night the hoppers digest the food in their gut, and excrete a high percentage of the potassium present; this results in a lowering of the potassium level present in the blood which makes the nervous system much more prone to stimulation due to its own spontaneous activity. This in turn, when the right temperature is reached in the morning, leads to muscular spasms and so to small locomotor movement. Once movement begins, it is maintained by the optomotor mechanism referred to above. This ingenious hypothesis has been checked by laboratory tests in which it was found that starved hoppers are much more prone to marching than recently fed ones; it is supported by evidence from the field, where observations show that hoppers forced to stop marching and seek shade by midday heat, often feed, and that thereafter a period of post-prandial resting sets in until the food has been partially digested.

This sort of work emphasizes the great need for the dovetailing of laboratory work and field observation. Another example is work concerning the flight metabolism of the adult locusts. It was observed that swarms of adult locusts can keep flying for hours without rest and this led scientists to wonder what source of energy was available for this tremendous activity. The answer was provided largely by the efforts of the Danish scientists Krogh and Weis-Fogh.[3] By flying locusts in wind tunnels these scientists were able to evaluate the exact amount of

energy output from each locust. Then by bio-chemical tests they were able to show that this energy was derived mainly from the metabolism of fat, which locusts stored in a large organ in the abdomen called the 'fat-body'.

The Maintenance of the Swarm

One of the most puzzling aspects of locust behaviour is that of gregariousness – the ability shown both by hoppers and adults to keep together in compact bands and swarms. This ability is underlined in a striking way when one considers the flight of swarms. The paths of particular swarms have been followed by aircraft for hundreds of miles, and estimates of the area of the swarm taken at intervals during the journey; these measurements show that the swarm remains approximately the same size throughout its journey, a most remarkable fact. What is it that enables millions of locusts in a swarm to keep together in a relatively compact group, landing together at night and taking off together in the morning? One explanation may be that each individual locust is reacting to the series of external stimuli, such as light, temperature, and humidity, in exactly the same way as its neighbour, and thus the whole mass of insects acts like one. Observation, however, encourages us to reject this explanation and substitute one based on a gregarious reaction between locusts; this is to say that one locust does something because other locusts are doing it. This simple statement conceals manifold problems, and in fact represents one of the most interesting and difficult lines of investigation in locust research. The sort of difficulties encountered may be illustrated by describing in greater detail the movement of locusts in swarms. This knowledge was gained by the use of special photographic techniques, involving the taking of photographs of swarms both from the air and from the ground; in this way the fine structure of flying swarms could be studied. A picture of a swarm seen from the air is shown in Inset 13. It was discovered that swarms had what could be described as an 'edge' or boundary, which could be defined as an area outside which the density of locusts dropped sharply. Inside this boundary, the locusts were separated into

numbers of small groups; within each group, the locusts were flying in approximately the same direction, but between groups there was a complete diversity of orientation. A further feature was the forced movement of masses of locusts in the air due to convective turbulence, which is severe in the hot countries where swarms are found; this turbulence leads to the appearance of 'towers' of locusts in a swarm. A moment's reflection will show that such occurrences in a swarm, those of randomly orientated groups and locust movements due to convection, would rapidly lead to the dispersal of a swarm and its ultimate dissolution; yet field observation, as reported above, show that on the contrary swarms exhibit a remarkable cohesion.

These apparently mutually exclusive facts led the scientists who discovered them, Dr R. C. Rainey and Mr J. Sayer of the Desert Locust Survey, and Miss Z. Waloff of the Anti-Locust Research Centre, to produce a hypothesis on locust behaviour that would reconcile them. There was, they suggested, some mechanism operating at the edge of the swarm, which caused individuals or groups of locusts that had crossed the edge to change direction and re-enter the swarm. This inference was suggested in relation to swarm structure of the Desert Locust, and recent observations have indicated the likelihood that such a behaviour mechanism does exist. Of great interest in this connexion is the recent discovery that such a mechanism apparently also exists in hopper bands, since groups of hoppers detached from the band experimentally, so manœuvre as to re-enter it. Such behaviour would obviously have to be mediated by the sense organs of the insects concerned, and the only organs in locusts capable of receiving stimuli over the necessary range to enable such a mechanism to work are the compound eyes and the tympanal or auditory organs. It is possible that the insects react to the sight of their neighbours or to the sight of the swarm as a whole; or, it has been suggested, they may react to the wing noise made by individual locusts or the swarm as a whole. The present author has investigated this latter possibility, and the results showed that probably the wing beat noise played some part in behaviour, but that it was a combination of sensory

factors, including vision, that mediated these complex behaviour patterns. These factors call for laboratory investigation into the sensory physiology of the insects, another indication of the need for the dual approach of field and laboratory work.

Swarm Movement

Leaving now the particular reactions in a swarm, let us consider the problem of the movements of swarms as a whole. Here, call is made on yet another science, that of meteorology, to help to elucidate the field observations. These latter show that, by and large, swarms move with the wind; the larger the swarm, the truer this becomes. A large swarm, say of ten square miles in area and upwards, moves in the direction of, and with the mean speed of, the prevailing wind. Smaller swarms, although moving in the direction of the wind, tend to move at a speed below that of the wind. These facts have been verified by following swarms from the ground, and in the air, and observing their displacement from day to day in relation to the wind speed and direction which prevailed at the heights between which the locusts were flying. The results of these experiments, coupled with data on the general movement of swarms derived from the swarm maps, led R. C. Rainey to advance a most interesting hypothesis relating to the general movement of swarms.[4] He pointed out that since, in general, winds within a few thousand feet of the ground may be considered as blowing from what meteorologists call 'zones of divergence' to what are called 'zones of convergence', then it is to be expected that major displacements of locust swarms will take place towards such areas of convergence. These 'zones of convergence' are, of course, thousands of miles in extent but they are characterized by two factors. One is that winds tend to blow into them, and the second, a corollary of this, is that such areas tend to have a higher rainfall than zones of divergence. In other words, Rainey's hypothesis means that if locusts do drift with the wind then it is likely they will end up in an area with a higher average rainfall than that from which they came. Now it is known from research on the oviposition of locusts that damp soil is favoured by mature females for egg-

laying purposes, and also that it is the best medium for the development of the eggs. Thus an apparently undirected movement of the locusts, that of drifting with the wind, seems to have definite survival value. As Rainey remarks, 'this mechanism may well be of vital importance to the continued existence of species such as the Desert Locust, inhabiting regions of scanty and erratic rainfall with an egg stage almost aquatic in its moisture requirements, and moreover suggests a possible survival value of flight without consistent orientation for other insects'. This hypothesis has the merit of explaining the observed displacements of flying swarms and the general migration cycle as demonstrated by swarm mapping: modifications are naturally to be expected, particularly when it is related to locusts other than the Desert Locust for whose movements it was formulated, but it represents a most important step away from the empiricism that has perforce marked so much locust research up to the present time. Such empiricism was particularly to be expected in the field of locust control, where the main object was to get results in the form of very high percentage kills.

LOCUST CONTROL

The administrative problems of locust control are of course enormous, and it must be remembered in discussing control measures that these do not simply involve the choice of the right method and its application. This comes right at the end of a vast chain of organization, and the functioning of this vitally affects the efficiency of the actual control operations. However, of recent years control development has proceeded so rapidly that research to increase the efficiency of the various methods has become essential.

A good example of this is seen in the application of insecticide dusts against hoppers. In controlling hoppers of the Red Locust by the use of a 6 per cent benzene hexachloride powder, it was found that an application of about 3 lb per acre of the powder gave a very satisfactory kill of the second instar hoppers, but that if the hoppers were larger, about sixth instar, then even amounts between 50 and 70 lb of dust per acre failed to kill

more than 85 per cent of the insects. Since the powder costs about a shilling per lb, it will be seen that the cost of control in this case varies greatly with the instar attacked. This in turn depends on the efficiency of the scouting service, that indispensable adjunct of control, whose job it is to detect and warn control teams of the presence of locusts; if the scouts can early detect small bands of younger hoppers, then control is cheaper and more efficient. Research can help here, and work is in fact being undertaken on the causes of the differential susceptibility of the various instars; this may be owing to increased impermeability of the cuticle, or a more rapid metabolic rate which oxidizes the insecticide rapidly into harmless by-products. But once the reasons are known, then there is a chance of devising a method of overcoming the difficulties, and this is a typical example of how fundamental laboratory research can have direct results in the field.

Use of Aircraft

The most spectacular advance in control methods has been in the use of light aircraft, both for reconnaissance and for the application of insecticides. The advantage of an aircraft for scouting work, especially one equipped with radio communication to the ground control teams, is obvious, and although the high capital cost and need for maintenance are drawbacks, in many instances aircraft have proved not only more efficient but cheaper than large native scouting forces. The application of insecticide by aircraft to locusts on the ground (air-to-ground) or in flight (air-to-air) has been the focus of much research. Inset 14 is a picture of 'air-to-ground' spraying. Development of reliable spray gear was first necessary; then work on the formulation of insecticides, and on the spraying nozzles to give the correct drop spectrum in the spray, were necessary to make the new method workable. And this method brings us full circle back to the ecological problems connected with the estimation of numbers. The success of spraying operations can only be assessed by estimating the locusts killed per gallon of spray applied, and the problem turns out to be the old one of counting

the locusts. However, methods have been devised which give a sufficiently satisfactory answer and enable the operators to assess the relative efficiency of the two methods of aircraft spraying, 'air-to-ground' and 'air-to-air', and to relate the kills obtained by these methods to those produced by other methods of control. Of course aircraft can only be used where airstrips and maintenance facilities can be provided, and these limitations at present restrict their wide-spread use.

One most important factor in the efficiency of control operations of any type is in the deployment of the control teams geographically and temporally, and in this connexion the monthly forecasts of probable locust movements derived from the mapping research at the Anti-Locust Research Centre referred to previously, are invaluable to those planning control measures. This is just another example of the useful dovetailing of fundamental research and practical control. Research into new insecticides and into more efficient ways of using the older insecticides will also have its effect on control methods. It is also possible that the ecological research already referred to, which showed the inter-relation between vegetation pattern and breeding in the Red Locust, may usher in a new type of control, involving land husbandry methods for maintaining a vegetation in which locusts will not produce swarms.

These few considerations have indicated some of the major problems facing those working against locusts. They give no ground for any facile belief that the locust problem will ever be conquered in the sense that these pests will be destroyed, nor even that scientists are at present in a position to indicate the best methods of dealing with any particular locust problem. But, because of the research described – and there is much more besides that could not be mentioned – it is perhaps justifiable to claim that anti-locust work and research is emerging from its empirical, trial-and-error stage into a period of planned experiment, the results of which will one hopes lead to a far greater understanding of the life and behaviour of these insects, and which will lead ultimately to the formulation of the most efficient methods for their destruction and control.

REFERENCES

1 Richards, O. W. & Waloff, N.: Studies on the Biology and Population Dynamics of British Grasshoppers. *Anti-Locust Bulletin No. 17*. (London, 1954.)

2 Ellis, P. E. & Hoyle, G.: A physiological interpretation of the marching of hoppers of the African Migratory Locust (*Locusta migratoria migratorioides* (R. & F.). *J. Exp. Biol.* 1954 **31** 271–9.

3 Krogh, A. & Weis-Fogh, T.: The Respiratory Exchange of the Desert Locust (*Schistocerca gregaria*) before, during, and after flight. *J. Exp. Biol.* 1951 **28** 342–57.

4 Rainey, R. C.: Weather and the Movements of Locust Swarms; a new hypothesis. *Nature* 1951 **168** 1037.

FURTHER READING

Uvarov, B. P.: *Locusts and Grasshoppers*. (London, 1928.)

Uvarov, B. P.: *Locust Research and Control 1929–1950*. (London, 1951.)

Taylor, T. H. C.: Locust Reporting and Anti-Locust Research. *Tropical Agriculture* 1955 **32** 186–92.

Gunn, D. L.: The Red Locust. *J. R. Soc. Arts* 1952 **100** 261–84.

Haskell, P. T.: Locust Research: discussions in London. *Nature* 1955 **176** 590–1.

VAPOUR CHROMATOGRAPHY

A. E. TROTMAN-DICKENSON

THE history of some branches of chemistry may be satisfactorily described in terms of the improvement of analytical techniques, and in the chemical industry analysis is no less important than in the university laboratory. Modern chemical technology is based upon the ability to control complex preparative processes. This control is achieved by the continuous analysis of batches and streams of material. The plant operator then adjusts the equipment so that any deviations in the composition of the material are speedily rectified. In the more modern plants the adjustments are made automatically by electronic equipment. Recent improvements in such well-established fields as electroplating have been based upon improved analysis of the composition of the electrolytic bath. Similarly, the development of new metals and alloys depends upon new methods of determining the composition of melts: very often it is the concentrations of constituents present in only a few parts per thousand or even parts per million that are of overriding importance.

Because analysis comes into all branches of chemistry, the development of a completely new technique is of great importance. Vapour chromatography is an analytical method of very general application since it may readily be used for the analysis of almost all substances that boil at temperatures less than 200° C under atmospheric pressure. It is, therefore, not surprising that vapour chromatography is often described as the most important analytical technique to have been developed in the last thirty years. It is a technique that is particularly well suited to the analysis of hydrocarbons and low boiling organic compounds which are becoming increasingly important industrially owing to the development of petrochemicals, i.e., chemicals made from oil, which include polythene, styrene, and synthetic

rubber. Similar organic compounds are used in modern oil refineries and catalytic cracking plants engaged in the production of high octane petrol. The following example shows the power of vapour chromatography. A typical problem confronting a chemist in a petrochemical plant might be the analysis of a mixture of six organic compounds containing four or five carbon atoms each and boiling in the range $-20°$ to $+30°$ C. There are four types of analytical equipment that can be used for this sort of job. The first is the automatic distillation column. This is a complex piece of equipment costing £4,000 to £5,000. It would require probably 50 cc. of liquid sample and would take about a day to complete the analysis, during which time it would require intermittent attention. Alternatively an infra-red spectrometer, which costs approximately the same, might be used, though a mixture of six compounds is rather too complex. Only about 1 cc. of liquid sample would be required and the machine would need little attention during the hour that it takes to trace out the spectrum. However a skilled worker would probably take several hours to deduce the concentration of components in the mixture from the spectrum. A mixture of this complexity could be analysed on a mass-spectrometer, the basic cost of which is about £15,000. The mass-spectrum of a few cubic centimetres of gas could be recorded in half an hour, during which time the machine would need little attention, though it must be remembered that mass-spectrometers require fairly constant expert attention to keep them 'well-behaved'. Again the snag comes in the interpretation of the spectrum. It would take a skilled worker several hours to work out the concentration of components to an accuracy of plus or minus two per cent, even with the help of a desk calculating machine. It is now possible to buy electronic computers to do this part of the job fairly quickly, but they cost about as much as the spectrometer itself. A commercial vapour chromatography apparatus, such as that shown in Inset 18, costs about £650. It is a straightforward piece of equipment which most people could be taught to use in a few hours. Three cubic centimetres of gas of the six-component mixture could be analysed with it in a matter of twenty minutes, during which time the apparatus could be left unattended. The interpretation of the record would probably take a further twenty minutes. We can see therefore that vapour chromatography is the best method for analysing such a mixture on account of the low cost of the equipment, the very small sample required, the speed of the analysis, and the fact that the operator need not be highly skilled.

Gas-liquid partition chromatography, which is the most important type of vapour chromatography, was developed by A. J. P. Martin and A. T. James, working at the Medical Research Council laboratories at Mill Hill, London. Their work was published in 1952. They developed a suggestion made by Martin and R. L. M. Synge, who shared a Nobel prize, in 1941. Many modifications of the basic technique have been introduced by workers in Europe and America. Already there are seven different commercial apparatuses manufactured in the United States of America.

THE APPARATUS AND ITS OPERATION

The operation of gas phase partition chromatography is straightforward. The key part of the apparatus is a column packed with an absorptive powder, such as kieselguhr, wetted with an involatile liquid. A stream of carrier gas, which is often nitrogen, is passed through the column. The mixture to be analysed is introduced into the stream of carrier gas. As the mixture passes through the column the constituent compounds move at different rates. If the conditions have been well chosen each compound emerges from the column separated and pure. The presence of the separated compound in the stream of the carrier gas may be detected by one of a number of devices; katharometers, which detect changes in the thermal conductivity of the stream of carrier gas, are most commonly used.

The way in which the separation of vapours takes place in the column may be indicated by an analogy. Let us suppose that we have a number of similar brick-shaped blocks made from three kinds of wood of different densities. If these blocks are floated on a pool of water different proportions of the blocks will be

above and below the surface of the water, according to the density of the wood from which they are made. These blocks may be set in a line at right angles to the direction of a steady wind as shown in Figure 1. If they are then released the wind

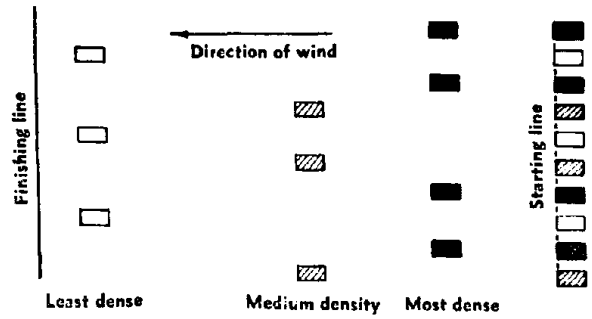


Fig. 1. Diagram showing how a steady wind will separate blocks of wood of different density floating on the surface of a pool of water – an analogy to vapour chromatography.

will blow them across the surface of the water. The blocks made of the least dense wood will travel fastest since they both present the greatest surface to the wind and, because of their shallow draught, pass most readily through the water. The densest blocks sunk deep in the water encounter the greatest resistance to movement while the area of their 'sails' is also least. After, say, ten minutes the blocks will be travelling in groups as shown in Figure 1. It is clear that in eleven or twelve minutes the least dense blocks will cross the finishing line. If the number of blocks crossing the finishing line at this time is counted, the number of light blocks originally present may be determined. The denser and densest blocks will cross the line after about 25 and 40 minutes respectively. Thus a separation has been achieved and a quantitative analysis may simply be made by counting the blocks in each group.

In the chromatographic column, the involatile liquid corresponds to the pool. It is spread on the kieselguhr so that the

greatest possible liquid surface may be presented to the gas. In the analogy the blocks of different density correspond to different types of molecules. A block of wood is found to float with a definite proportion of its volume, which depends on its density, below the surface of the water. Similarly a vapour in the presence of a liquid will partially dissolve in the liquid and partially remain in the gas phase. The proportion of the molecules of the vapour which at equilibrium are in the liquid phase depends upon the nature of the liquid and the temperature. However, it is a rough working rule that at any temperature the higher the boiling point of the vapour the higher the proportion of molecules in the liquid phase. The wind of carrier gas can only move the molecules which are in the vapour so, in general, the most volatile substances move fastest down the column. Readers who are familiar with the operation of distillation will immediately see the similarity of that operation to partition chromatography. They will know that only an exceptional distillation column has an efficiency of 100 theoretical plates, whereas it is a simple matter to construct a chromatographic column with an efficiency corresponding to 2,000 theoretical plates.

A simple form of gas chromatographic apparatus is shown in Figure 2. Commercial apparatus costs many hundreds of pounds, but most of this cost is made up of the expensive electrical and recording equipment. An adequate instrument may be made for a few pounds, in fact, the basic parts could be found in any reasonably well-equipped school laboratory. Accordingly the various parts of the apparatus will be described fairly fully in the hope that some readers may be encouraged to construct their own equipment.

The Carrier Gas

The most convenient carrier gas is nitrogen which can be obtained in cylinders. Its flow may be conveniently regulated with a needle valve in conjunction with the usual type of reducing valve on the cylinder. Hydrogen is also an excellent carrier gas, but hydrogen cylinders are not usually so easy to get. If gas in cylinders is not available solid carbon dioxide may be used as

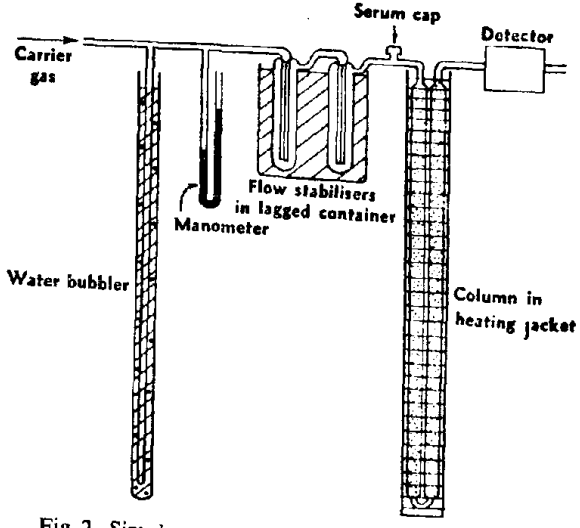


Fig. 2. Simple vapour chromatographic apparatus.

a source of gas. The crushed 'dry ice' should be placed in a corked container and insulated so that it sublimes at a convenient rate.

The Flow Control System

If the apparatus is to function satisfactorily great care must be taken with the system for controlling the flow of carrier gas. Often the troubles that beset home-made equipment can be traced to this system. It is convenient to have a blow-off device after the needle valve to maintain the head of pressure roughly constant. A simple form of blow-off is illustrated in the figure. It is a water bubbler made of glass tubing so arranged that bubbles form when the pressure exceeds about six feet of water. It is also convenient to have a mercury manometer as shown to record the pressure. The carrier gas should then pass through two or more 'flow stabilizers' of about 40 cc. volume made with a fine capillary tube down the middle. The stabilizer should be

placed in a large lagged vessel containing water. These simple devices are very effective in ironing out any surges.

The Injection System

Some device must be incorporated by which the vapour to be analysed can be introduced into the stream of carrier gas. A simple and effective arrangement is to introduce the samples through a medical serum-cap placed, as shown, over a small side tube. The sample, either gas or liquid, should be taken in a hypodermic syringe and injected into the system by pushing the needle through the serum-cap. The cap seals itself when the needle is withdrawn; it remains effective for many analyses.

The Column

The mixture of vapour and carrier gas then enters the column. The column is usually made of glass tubing of quarter-inch bore. It may be straight or coiled, but the hairpin shape shown in the figure is generally most convenient. The column is packed with a mixture of kieselguhr and an involatile liquid such as a dibutyl phthalate, a silicone, or hydrocarbon oil. A readily available packing is the detergent '*Tide*' which is a liquid detergent thinly spread over a powder; probably other detergents of the same type would also prove satisfactory provided they are also thoroughly dried. It is important that the column should be evenly packed. This can best be achieved by tapping the column while the packing is being introduced. For most purposes the column must be heated. This may be done by placing it in a vapour bath, but an evenly wound electric heater is quite satisfactory.

The Detector

Many types of detector have been used and it is in this respect that the various commercial apparatuses differ from each other.

The simplest form of the detector is that due to the Czech worker Janak. It is very suitable for the analysis of low-boiling hydrocarbon mixtures such as 'Calor' gas. Carbon dioxide is used as the carrier. As the stream emerges from the column it

passes through a fine bubbler into a container of caustic soda as shown in Figure 3.

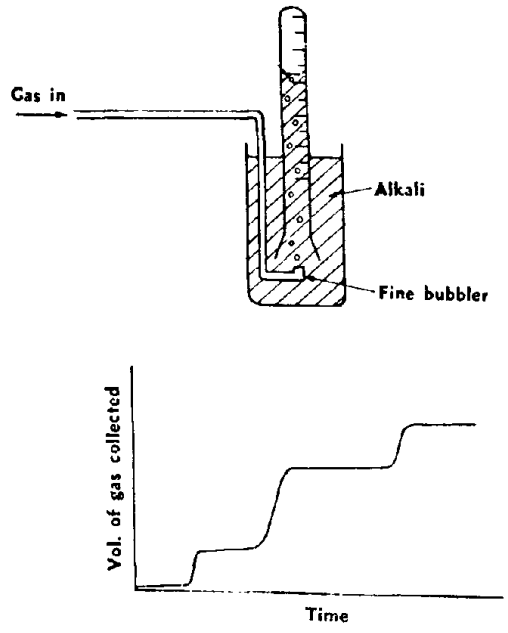


Fig. 3. Simple gas burette detector, and the record produced.

The carbon dioxide is rapidly absorbed by the caustic soda, but when a volatile hydrocarbon emerges from the column through the bubbler it does not dissolve and collects in the simple gas burette. It is then only necessary to note the quantity of gas in the burette at regular intervals to obtain a record as shown in the figure. The time after the vapour has been introduced to the column at which a step occurs in the volume curve is characteristic of the gas emerging. The height of the step (in cubic centimetres) is equal to the volume of the constituent gas.

Another simple type of detector is the titration unit shown in

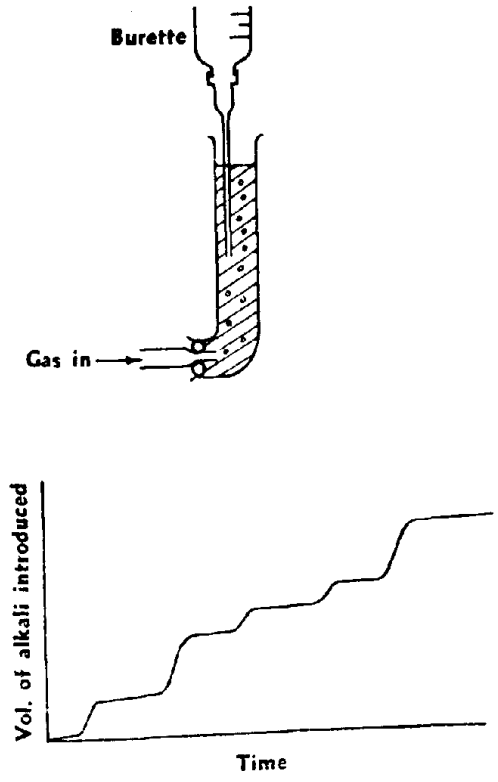


Fig. 4. Titration unit used for detection.

Figure 4, though unfortunately this is only suitable for a limited range of vapours, that is, vapours of acids and bases that are soluble in water. In use the carrier gas emerges from the column and is bubbled through water containing an indicator. When the acid, say, emerges in the stream of carrier gas it dissolves in the water and the colour of the indicator changes. The operator then introduces sufficient alkali to restore the solution to its original colour. He strives to keep the solution at the desired

colour and makes a record of the amount of alkali added after convenient intervals of time. He thus obtains a 'stepped' graph of alkali added versus time as shown in the figure. Again the position of the steps are characteristic of the constituent acids and their heights are proportional to the quantities present. The titration detector has been made fully automatic by using a photocell to detect the change in the colour of the solution. The output of the photocell operates a relay driving the motorized plunger of a microburette. The position of the plunger is continuously recorded by a pen writing on a rotating drum.

Both these detectors yield an 'integrated' chromatogram. The detectors in most commercial instruments do not give this type of trace, but one in which the emergence of a constituent is represented by a peak. This, as we shall see, is not particularly desirable but the detectors have two overwhelming advantages. First, they may be used with an unlimited variety of gases; secondly, they are suitable for use with electrical recording systems. The most widely used detector is the thermal conductivity gauge. A simple version of this instrument is shown in Figure 5. It consists of a heated filament suspended in the stream of carrier gas. The surrounding tubing may be made of glass and immersed in a liquid or it may be drilled out of a metal block. The filament is heated by the application of a constant voltage so that its temperature remains constant. When a vapour emerges from the column the thermal conductivity of the ambient gas rises, because almost all substances have thermal conductivities higher than that of nitrogen. As a result the rate of heat loss from the filament increases and its temperature falls. The filament is made of a metal such as platinum or tungsten whose electrical resistance varies with temperature. Accordingly the fall in temperature is accompanied by a change in the resistance of the filament. This change is recorded by making the filament one arm of a Wheatstone bridge circuit. Then the emergence of a vapour from the column throws the bridge out of balance and produces a potential that may be fed into a sensitive galvanometer or electronic recorder. The usual practice is to place a twin thermal conductivity gauge in an opposing arm of

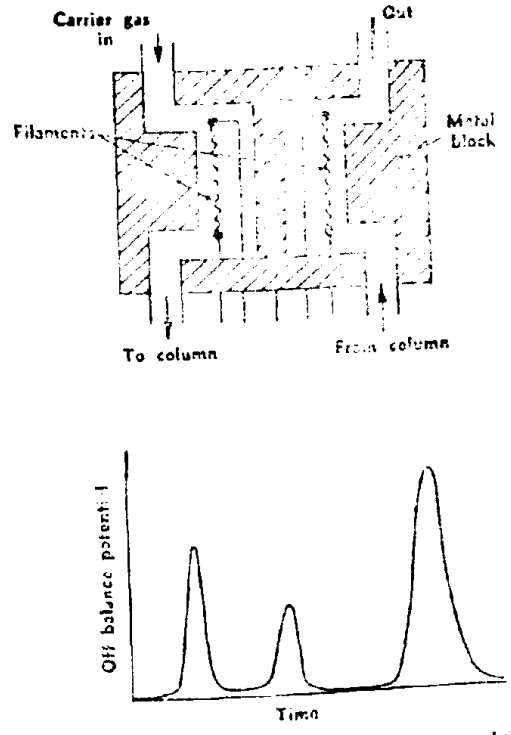


Fig. 5. Thermal conductivity gauge used as a detector and the type of record produced.

the Wheatstone bridge and to flow the carrier gas through it before it enters the column. This arrangement serves to iron out random fluctuations in the flow of the carrier gas, the voltage applied to the filaments, and so on. The type of record that is obtained is a graph of the 'off balance' potential against the time from the introduction of vapour to the column, as shown in the figure. The time at which the peaks occur is characteristic of the nature of the vapour and the quantity of vapour emerging is proportional to the area of the peak. It should be noted that

the proportionality constants differ slightly from gas to gas. The width of the peak for a vapour is nearly independent of the amount of the vapour emerging. Accordingly the peak height is also approximately proportional to the amount of vapour, and use is often made of this approximation since it is much simpler to measure a height than an area. A thermal conductivity detector is used in the commercial apparatus shown in Inset 18, which gives a general idea of the layout and dimensions of gas chromatography apparatus.

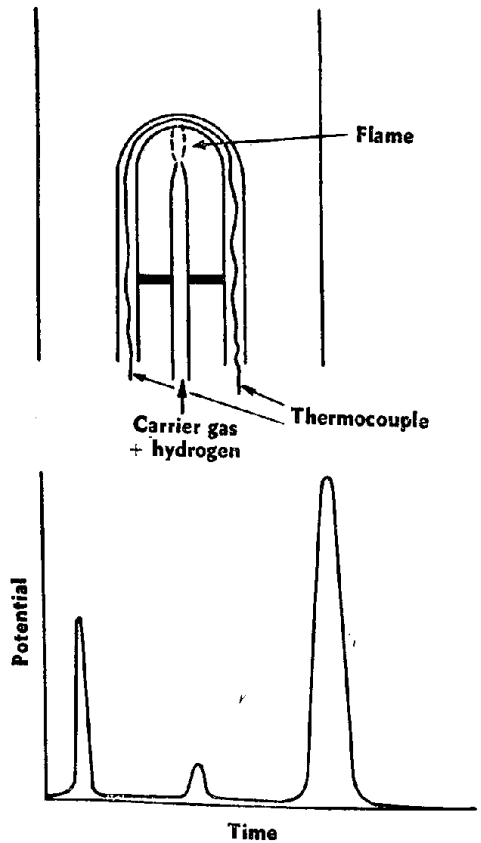


Fig. 6. Microflame detector.

The microflame detector yields a similar sort of record. Its construction is shown in Figure 6. When the carrier gas emerges from the column it is mixed with a steady stream of hydrogen sufficient to sustain a flame about one centimetre long as a tiny jet. The jet is placed in a suitably baffled chimney to reduce the effect of strong draughts. The emergence of an organic vapour in the stream of carrier gas results in a change (usually a rise) in the temperature of a thermocouple placed in the flame. A typical quantity of vapour may cause the thermocouple temperature to change from 600° C to 610° C. As it is not practical to record such a small percentage change directly, a potentiometer is used to balance off the thermo-electric potential when the carrier gas alone is passing. Then the 'off balance' potential consequent upon the temperature change of 10° C may be readily recorded.

The microflame detector has some definite advantages over the thermal conductivity detector which offset the disadvantages of having two streams of gas. In the first place it tends to be less temperamental than most home-made thermal conductivity gauges. In the second, the area of peak obtained with the microflame detector for an organic compound is roughly proportional to the weight of the substance emerging from the column, whereas the area of peak obtained with the conductivity detector is roughly proportional to the number of molecules emerging. It can be seen, therefore, that the microflame detector is more suitable for the analysis of mixtures of compounds of high molecular weight.

SOME APPLICATIONS

The remarkable analytical power and the speed of vapour chromatography is clearly demonstrated by the two typical chromatograms shown in Figures 7 and 8. Anyone acquainted with conventional analytical procedures will immediately recognize that the analysis of such complex mixtures of related compounds could only be accomplished by enormous labour, and

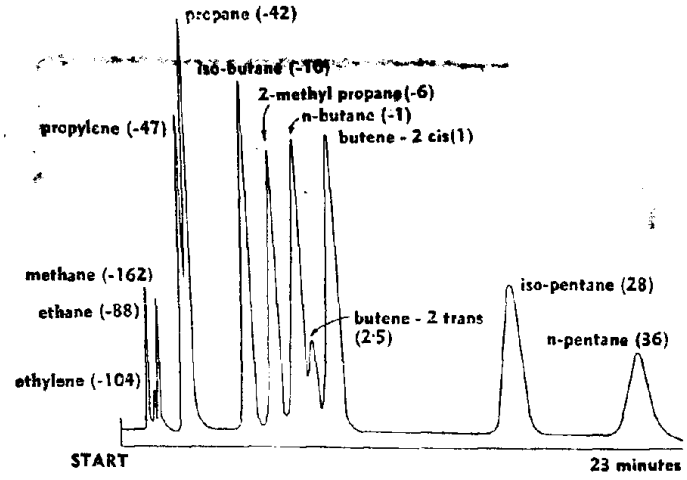


Fig. 7. Typical chromatogram. (*Courtesy Perkin-Elmer Corporation.*)

with large samples, if indeed it could be achieved at all. The most important limitation of vapour chromatography is that hitherto it has not been possible to adapt it to the continuous analysis of a stream vapour. This problem is not often met in the laboratory but it is the principal analytical problem in industrial plants. It would, however, not seem impossible to develop a continuously recording chromatographic apparatus with a packing that is pushed steadily along the column in the direction opposite to the flow of carrier gas.

The development of vapour chromatography is primarily of importance because it provides a quick and simple method of analysis of enormous versatility. In a few years it will be used generally in every sizeable laboratory concerned with work on organic compounds. There are some fields of chemistry in which its impact will probably be particularly great. The chemistry of organic compounds containing large numbers of fluorine atoms has been held up by the difficulties of analysing mixtures of these compounds. It is a peculiarity of fluorocarbons that the

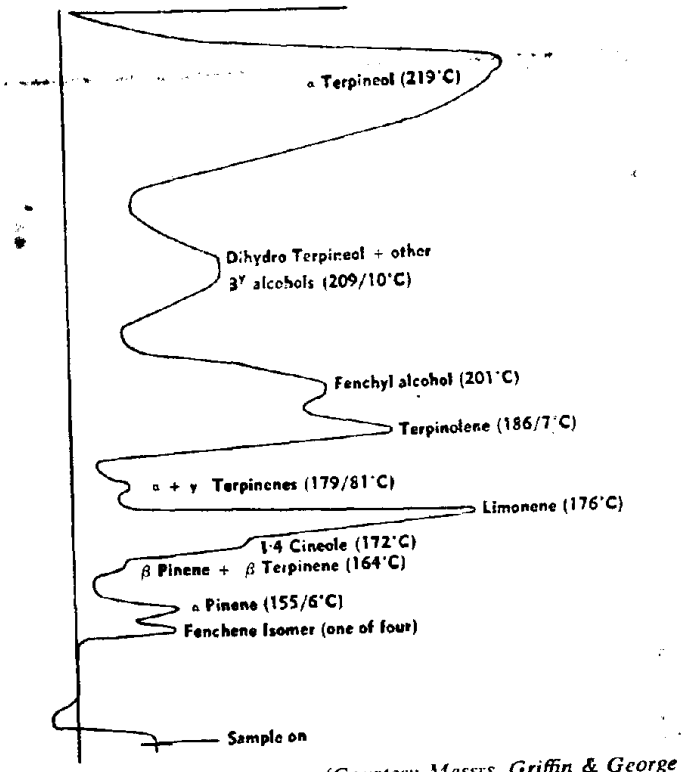


Fig. 8. Typical chromatogram. (*Courtesy Messrs. Griffin & George Ltd.*)

forces between the molecules are often very weak. This is reflected in their boiling points. The result is that distillation processes are often not as convenient as they are when dealing with compounds containing principally carbon and hydrogen atoms. The position is aggravated by the fact that fluorocarbons are often produced by methods which yield mixtures of closely similar compounds. Vapour phase chromatography has proved so successful in such analyses that it has been used in preparative

work as well. It has been found possible, with a column some four inches in diameter, to isolate samples of as much as 20 or 30 grammes of a pure compound.

The rapidity and ease of chromatographic analysis will result in its widespread use in the production of the less common organic compounds. This industry has an importance which is out of all proportion to the volume of its turnover because its products are very widely used in research laboratories. These rarer compounds are usually made by laboratory-type techniques which, though they should yield a satisfactory product, are not subject to the degree of control normal in full-scale production. Now it should be possible to provide at very small cost a reasonably detailed analysis of the majority of the products. This will be a great boon to the research worker and will often save annoyance and frustration.

Martin and James developed vapour phase chromatography to analyse the very complex mixtures of organic acids found in nature. The ability to analyse mixtures of such complexity easily is bound to have very far-reaching effects on the development of biochemistry. It has often proved possible to adjust chemical systems so that the products are not too embarrassingly diverse and thus to make progress. Unfortunately biological systems can rarely be made to behave in such an obliging fashion. Moreover it is often the compounds present in smallest quantities that are of greatest significance. The development of solid-liquid chromatography resulted in a great acceleration of biochemical research. Vapour phase chromatography may well prove to be of equal importance.

FURTHER READING

C. Phillips: *Gas Chromatography* (London, 1956).

WORLD'S FIRST MAJOR NUCLEAR POWER STATION

ARCHIE CLOW

ON the authority of Sir Christopher Hinton it can be said that at the end of the war no one would have dared to predict when a major nuclear energy power station would be in operation; yet little more than a decade later, in October 1956, the basic unit of Great Britain's nuclear power programme, the nuclear power station at Calder Hall, Cumberland, was feeding power to the national grid.

Calder Hall cannot claim to be the world's first nuclear power station; Russia opened hers to western visitors in 1955, but the scale of the two stations is vastly different. In magnitude Calder Hall takes its place alongside the traditional stations being built by the British Central Electricity Authority. In part, the electrical part, it is conventional; the rest is revolutionary in several respects – it tames a new source of energy; it demonstrates the possibility of setting up a power station without the usual elaborate rail-heads or access waterways through which the thousands of tons of fuel have normally to be brought to a steam station. Such is the potency of the new 'fuel' that, instead of thousands of wagons of coal, an occasional lorry-load of uranium canisters is all that will be required.

Public knowledge that Britain was to embark on a programme of nuclear power production dates from April 1953, when the Minister of Supply announced in the House of Commons that the preparation of a site for a nuclear power station, adjacent to the plutonium producing plant at Windscale, was to start immediately. Three months after the announcement in the House the building and civil engineering contracts for the first half of the unit, so-called Calder Hall A, were placed. Now this

unit is in production and the greater part of the civil engineering work associated with the second unit, Calder Hall B, is complete. The programme for Calder Hall A included the following: two nuclear reactors, with each of which is associated four heat exchangers, and four Parsons steam turbo-alternators serviced by the usual condensers and cooling towers. A unique feature of the generating side is the provision of two 'dump' condensers, which can absorb the full energy output of the reactors should, for any reason, the turbines be unable to do so.

The main features of Calder Hall and the adjacent Windscale plutonium factory can be seen from the low hills between them and the mass of Scafell in the Lake District (Inset 1). Calder Hall A, with its cooling towers three hundred feet high and nuclear reactors, lies to the south, the most notable feature of the skyline to the north being the plutonium producing plant at Windscale with the prominent stack rising from the primary separation plant.

A more detailed impression of Calder Hall is given in Insets 2 and 3. The large structures are the 200-feet high reactor buildings, each containing one reactor. External to the buildings are eight heat exchangers (four to each reactor) in which the hot gas from the reactor gives up its heat energy with the production of steam to feed the turbines. It is in the reactors, which take the place of coal-fired boilers, and the heat exchangers that the novelty of Calder Hall lies. From the point at which the steam leaves the heat exchangers the station may be regarded as conventional, with the exception of the 'dump' condenser referred to above. This is necessitated by the power generating properties and characteristics peculiar to a nuclear reactor, which behaves differently from a steam source.

The Nuclear Reactor

The heart of a nuclear power station and the source of its energy is the nuclear reactor in which matter is transformed into energy in terms of the often quoted Einstein relation, $E=mc^2$. The matter in the case of Calder Hall is natural uranium which consists of a mixture of the isotopes, U-235 and

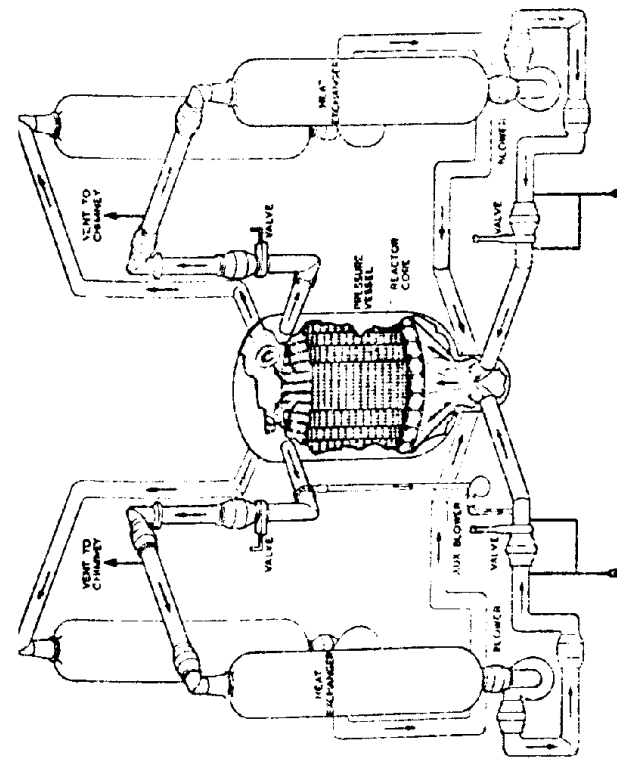


Fig. 1. Calder Hall Reactor: carbon dioxide circulation through reactor and heat exchangers.

U-238, in the ratio of about 1 : 140. In nuclear reactions involving natural uranium and slow neutrons it is the U-235 atoms that are important. When one of these is bombarded with a neutron, ${}_0n^1$, one of a large number of possible reactions takes place, e.g.,

$$ {}_{92}U^{235} + {}_0n^1 \rightarrow {}_{92}U^{236} \rightarrow {}_{57}La^{148} + {}_{35}Br^{85} + 3{}_0n^1 $$

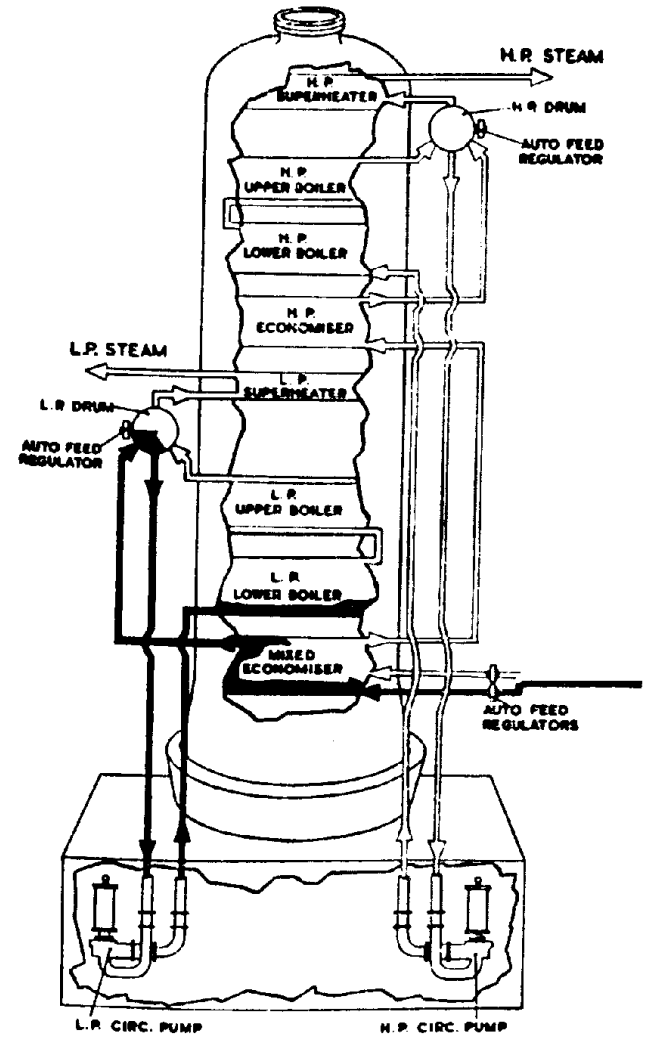
The reaction is accompanied by the disappearance of roughly a two-thousandth of the matter and the release of an equivalent amount of radiation and kinetic energy. If all the atoms in one lb. of U-235 were fissioned (an ideal still far from being attained) the energy released would be equivalent to some 1,500 tons of coal, hence the simplification in raw material supplies. From the above it will be seen that physically one may consider the principles of a nuclear power station as simple; where the difficulties lie is in the technology of the process.

In designing a nuclear power station the problem is how to get the energy released into a form suitable for use in a conventional turbo-alternator. A large number of systems are theoretically possible, and of these, that selected for Calder Hall was to pass carbon dioxide gas, at a pressure of a hundred pounds per square inch, through the reactor where it picks up the heat energy and conveys it to heat exchangers.

The Calder Hall reactors consist of matrices of graphite blocks, the 'moderator', some 20 feet high and 30 feet across, through which pass 1,696 accurately drilled vertical channels to hold the fuel elements. The function of the 'moderator' is to slow down the fast neutrons produced during fission to something like a five-thousandth of their original velocity.

The graphite matrix, which is constructed of graphite free from neutron poisons, sits on flat steel plates set on a steel grid, and the fuel elements are stacked in the vertical channels. Other holes accommodate control rods of boron steel.

The fuel elements are rods of natural uranium over a yard long and just over an inch in diameter. These are contained in magnesium alloy 'cans' specially designed to facilitate the transfer of the heat generated in the uranium to the passing carbon dioxide. Since this gas works more efficiently at increased pres-



g. 2. Calder Hall Heat Exchanger : flow diagram illustrating steam flow. Carbon dioxide at 336° C enters the top of the exchanger and, having given up part of its heat, leaves the foot at 140°.

sure, the graphite matrix is contained in a pressure vessel, a cylinder of two-inch steel plates 37 feet in diameter and 71 feet high, with domed ends, set with its axis vertical. It was fabricated from prepared sections brought to the site ready for welding. Some of the welds took over 24 hours to complete, groups of welders working continuously in shifts.

The carbon dioxide enters the bottom dome through four ducts, flows up through the fuel channels and passes out of the reactor to four separate heat exchangers. From these it is returned to the reactor by a centrifugal blower. When the reactor is working, the carbon dioxide leaves the top at 336° C, heat is absorbed from it in the heat exchanger, and it returns to the pressure vessel at 140° C.

Were the production of heat all that went on in the reactor it would present technical problems enough, but it is, at the same time, the seat of the generation of intense and lethal radiation. This being so it has to be encased in a 'biological shield' of high density concrete (150 lb/cubic foot). The shields at Calder Hall are between seven and eight feet thick, octagonal in shape (for constructional reasons) and 90 feet high. Between the 'biological shield' and the reactor there is also a massive 'thermal shield' built from hundreds of six-inch-thick steel plates.

Such massive structures, with the additional weight of the heat exchangers, presented difficult civil engineering problems which were resolved by placing the whole reactor building on a concrete raft eleven feet thick, with separate rafts for the heat exchangers. The total weight of concrete required in the construction of Calder Hall A and B will run to a quarter of a million tons, reinforced by 6,000 tons of steel bar.

Weight was not however the only problem. Tolerances to within plus or minus a tenth of an inch on both vertical and plan true dimensions were required, and great care had to be taken in placing the concrete in the confined spaces around the large number of charging and other tubes that pass through the 'biological shield'.

Now that everything is in place all that can be seen of the reactor must be viewed from the charge-discharge deck (Inset 5).

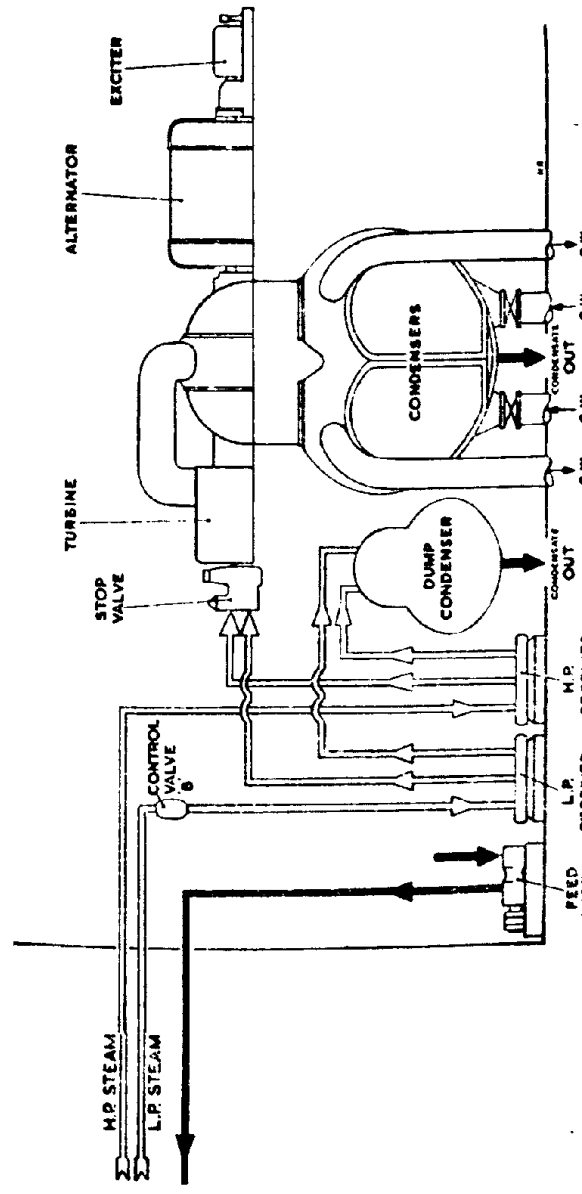


Fig. 3. Calder Hall Turbine Hall: diagrammatic representation of steam flow, illustration, particularly the 'dump' condenser.

Charging the fuel elements and the removal of the used canisters is carried out from machines that run on rails on top of the massive concrete 'biological shield'.

The Heat Exchangers

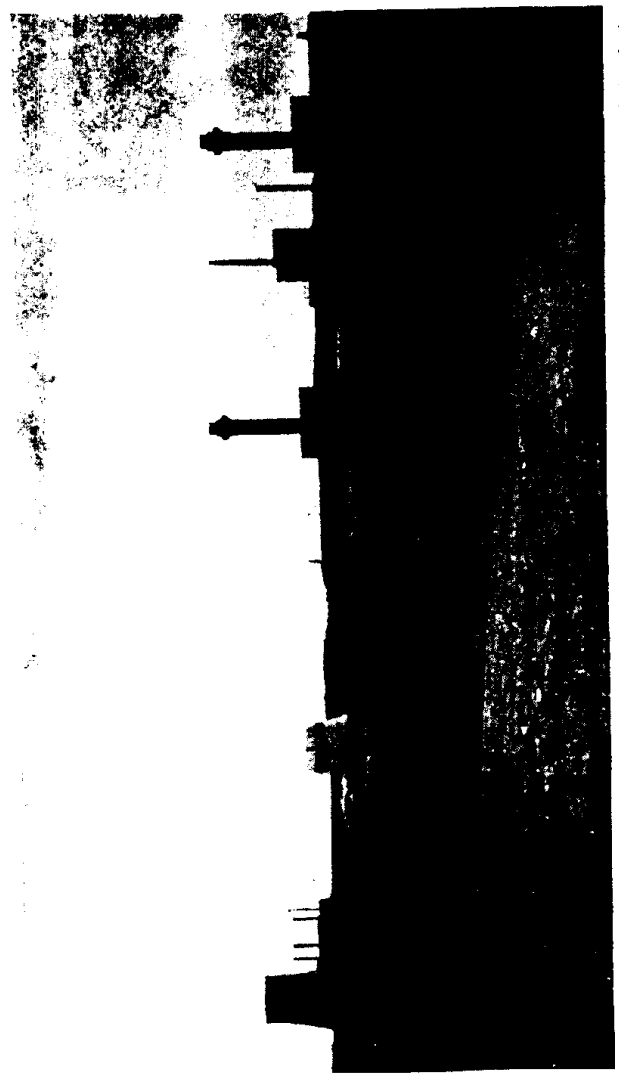
The most easily examined feature of the whole station is the heat exchangers immediately outside the reactor buildings. These intricate pieces of equipment, in which the hot carbon dioxide gives up its heat to a water circuit forming steam, contain high and low pressure evaporators, super-heaters, and economizers, and present a total of nearly 100,000 square feet to the live gas (Inset 4).

The fabrication of the heat exchangers, much of which had to be carried out on site, was one of the largest, if not the largest, operation of its kind ever executed, each vessel being 80 feet high by 18 feet in diameter and weighing 200 tons. Sections brought by road from Scotland were welded together on the site; each weld was X-rayed to detect possible flaws, stresses were relieved, and the finished vessel tested to a pressure considerably higher than that at which it will be required to work. External fittings and internal piping and water bring the total weight of each unit up to 450 tons. But these figures are hardly surprising when one realizes that the eight heat exchangers of Calder Hall A contain between them 250 miles of E.R.W. tube on to which no less than 120 million studs were welded to extend the heating surface.

Electrical Generating Plant

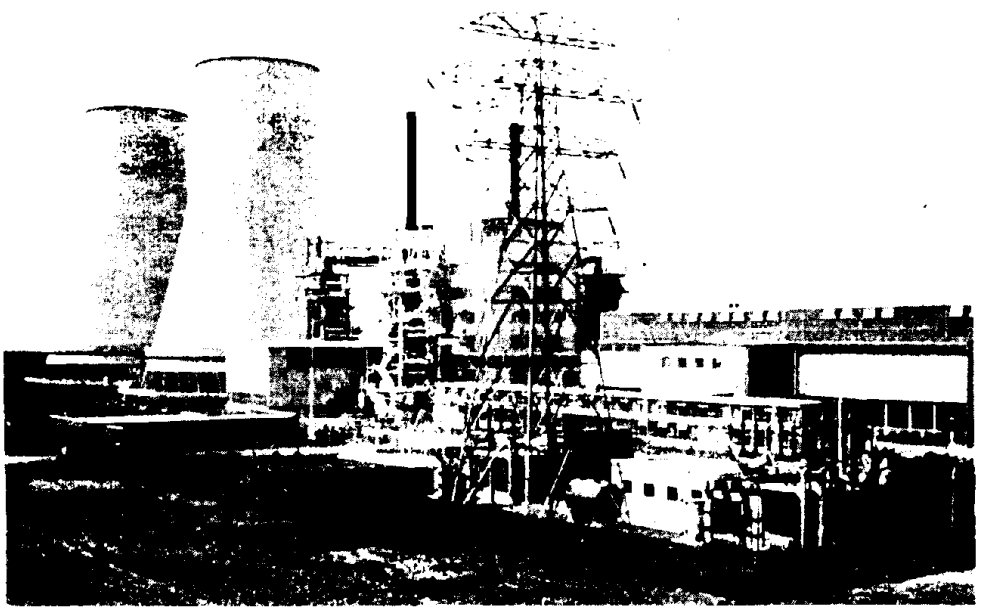
Associated with each reactor there are two turbo-alternators of maximum continuous rating 23,000 KW each (Insets 6 and 7) built on traditional lines, but as a civil engineering operation meticulously planned and executed. Steam at two pressures is supplied, a low pressure supply coming directly from the heat exchangers as well as that derived by expansion from the high pressure blades. The power station is served by two cooling towers – the most prominent feature of the Seascale scene. Each tower is 200 feet in diameter at the base and all but 300 feet high.

FIRST MAJOR NUCLEAR POWER STATION
CALDER HALL, CUMBERLAND

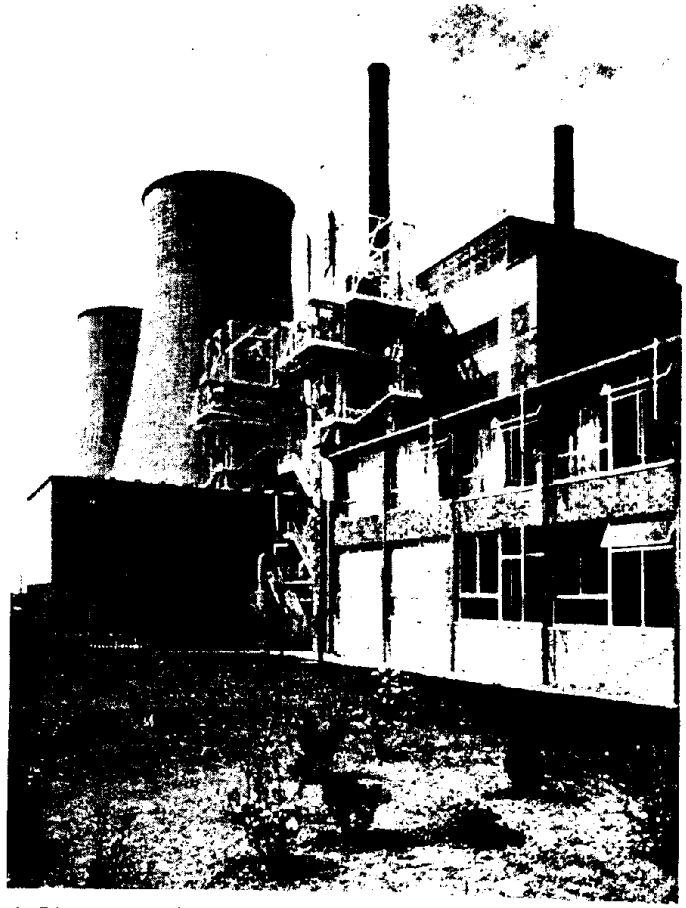


1. Calder Hall Power Station and Windscale Plutonium Works, as seen from the low hills of the Lake District, with the Irish Sea beyond. On the right are the two Windscale plutonium producing piles and on the left the Calder Hall Power Stations A and B under construction.

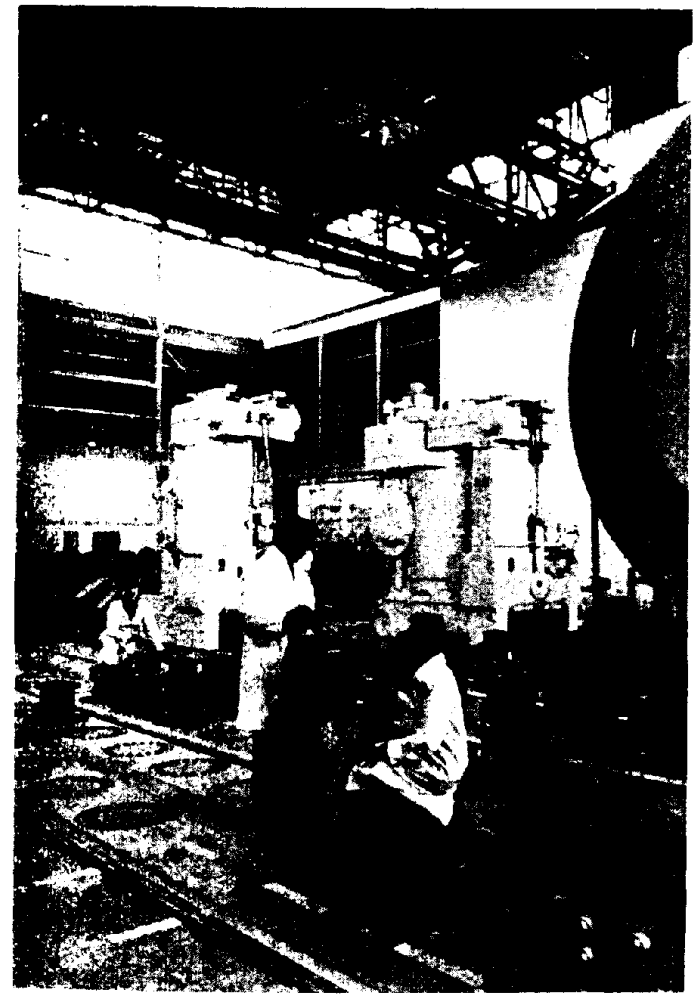
2. Each of the large buildings shown houses a nuclear reactor, the heat energy from which is converted into steam in the four heat exchangers which are built external to the reactor buildings.



3. Part of Calder Hall nearing completion. In the foreground is the Central Electricity Authority's electrical substation through which the energy from the reactors will be fed into the national grid.

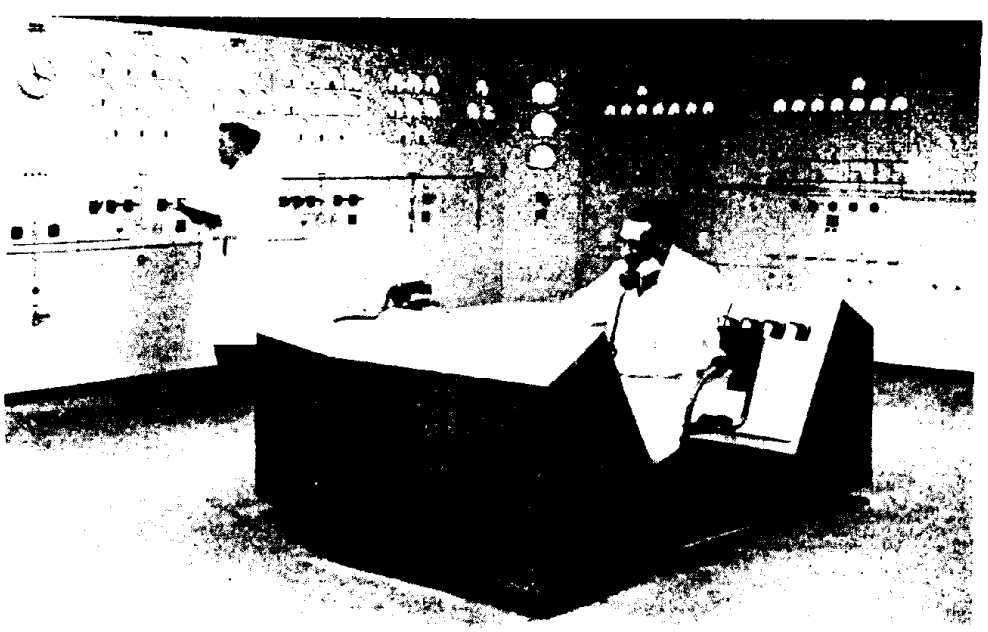


4. The energy released in the reactor is brought out by a stream of high-pressure carbon dioxide. This passes down through the heat exchangers, two of which are seen in the centre of the picture. Each exchanger is some eighty feet high and eighteen feet in diameter, and weighs 200 tons.

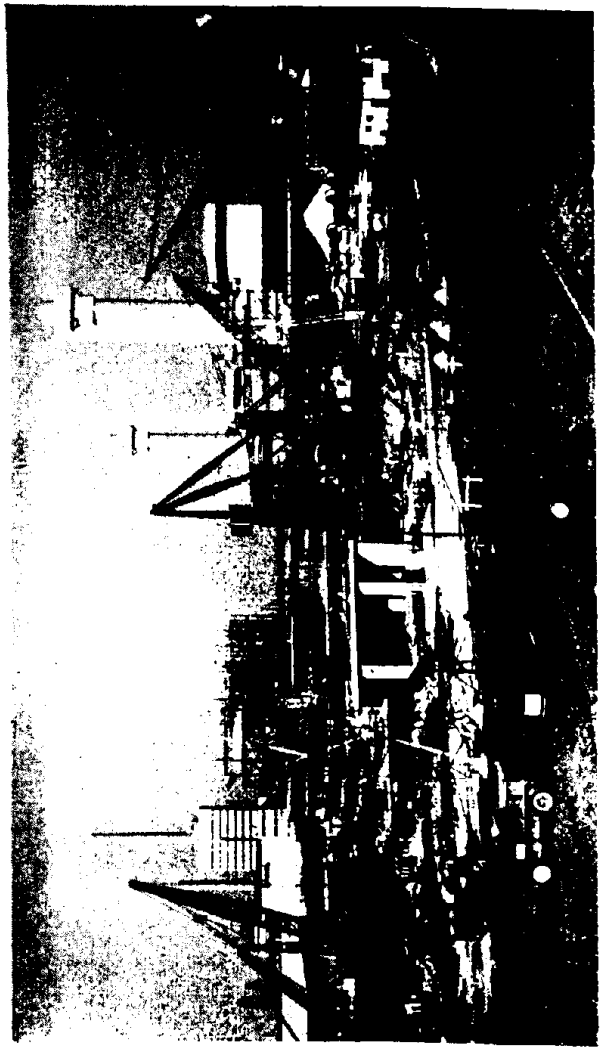


5. The charge-discharge deck of No 1 reactor at Calder Hall Nuclear Power Station. The large structure at the rear-left is a lightly shielded charging machine and at the rear-centre there is a heavily shielded discharging machine. These are used to insert new fuel cartridges and remove highly radio-active spent cartridges from the reactor.

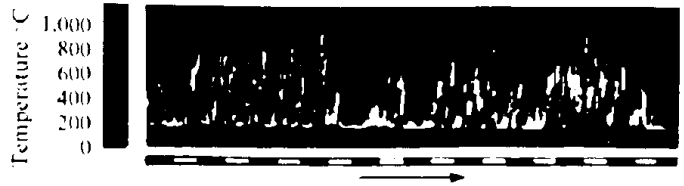
6. The Turbine Hall of Station A at Calder Hall. Each of the four Parsons steam turbo-alternators generates 23,000 kilowatts.



7. The control room for the reactor, showing a control engineer at the main control desk and, in the background, about half the associated instrumentation.



8. With Calder Hall Station A in commission, the second station, Calder B, is already well on the way. Against a background of the neighbouring Windscale Plutonium Works, in this picture can be seen the foundations of the second station being prepared.



9. Cathode ray trace showing the thermal potential difference developed between a constantin slider and a lapped steel surface. (Load 500 g; speed 300 cm. sec.; temperatures exceed 1,000° C; duration 10^{-4} sec.) *Reproduced by permission from* Bowden, F. P. and Tabor, D.: *Friction and Lubrication of Solids* (revised edition) (Clarendon Press, Oxford).

10. Auto-radiograph showing metal transfer during a single hammer blow on a flat surface (× 3.5). (From Bowden and Tabor, *vide supra*.)

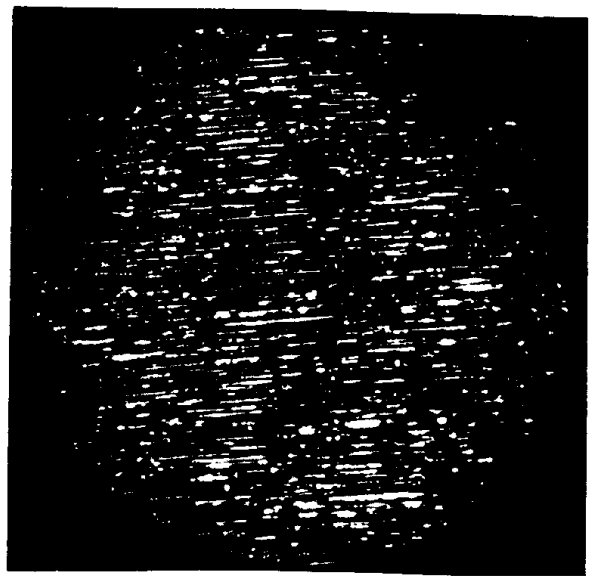
11. A small band of desert locust hoppers marching in the scrub country of the northern frontier district of Kenya. *(By courtesy of Anti-Locust Research Centre.)*

12. Dense low-flying locusts in Tanganyika. *(By courtesy of Anti-Locust Research Centre.)*

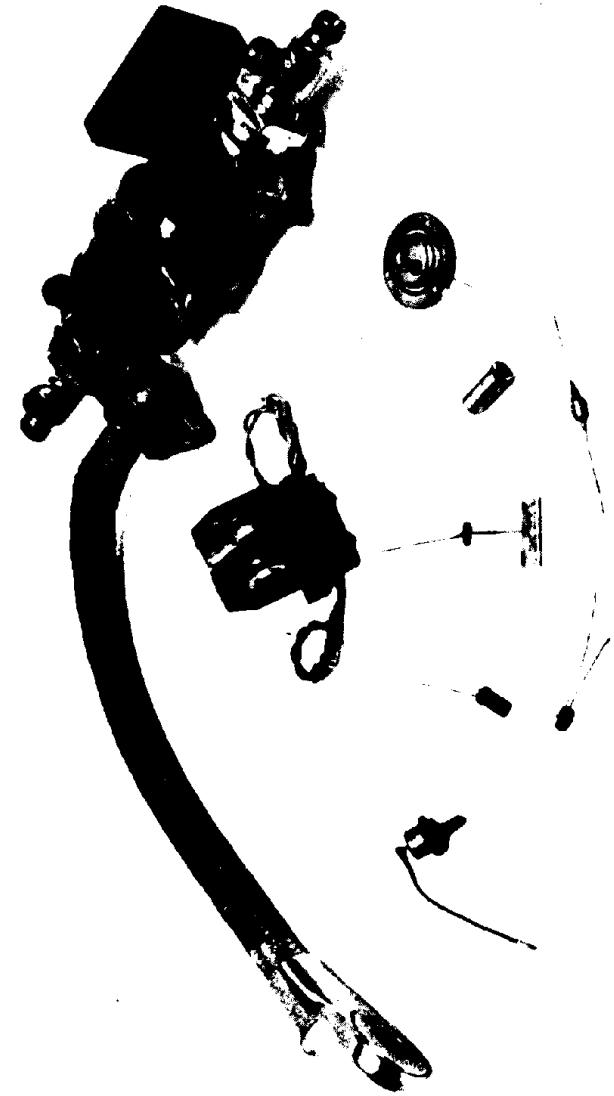
13. Aerial view of locust swarm near Wajir in N. Kenya. *(By courtesy of Desert Locust Survey.)*

14. 'Air-to-ground' spraying of Red Locusts in tall grass in the Kukwa Valley, Tanganyika. The African standing on the jeep is holding a banner to show the pilot the line of flight. *(By courtesy of Anti-Locust Research Centre.)*

15. Cross-section of a germanium p n junction made by the alloy technique using indium. The p-layer is shown as a black line. Actual diameter of junction 4 mm.

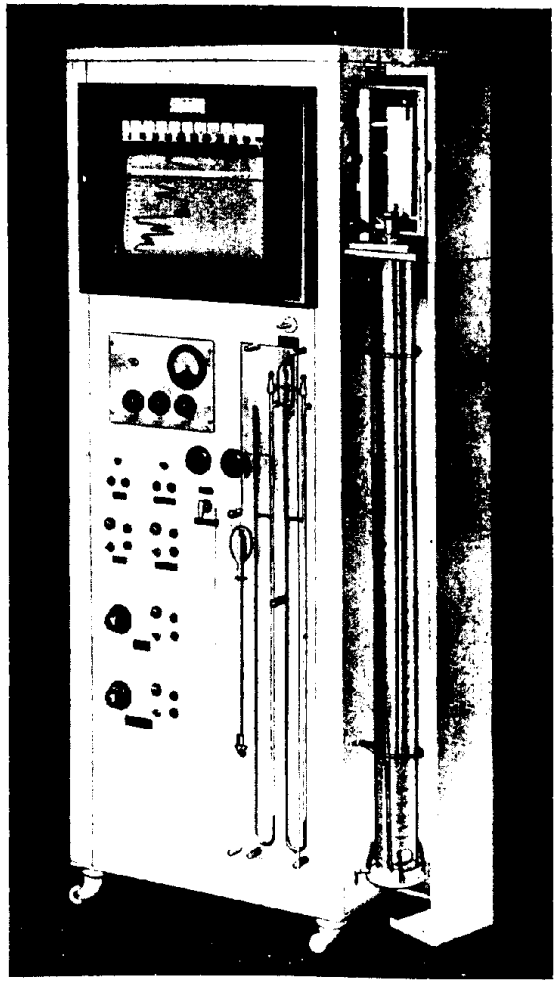


16. Plan view of a germanium p-n junction with the indium bead removed to show the crystal structure of the recrystalized p-layer. (Insets 15, 16, and 17 *by courtesy of The General Electric Company Ltd.*)



17. Semi-conductor devices. In the foreground, various low and medium power junction diodes and transistors; middle, stack of medium power silicon junction diodes with cooling fins; rear water-cooled germanium junction power diode.

VAPOUR CHROMATOGRAPHY



18. Vapour chromatography apparatus. On the left the recorder with the flow control panel beneath, and on the right the column with the serum cap and katharometer above. (*By courtesy of Griffin & George Ltd.*)

By planning their construction carefully a three feet concrete lift per day was reached and maintained. The capacity of each tower is 3 million gallons an hour and five pumps are available to deliver water to the condensers of the two reactors.

With Calder Hall A in commission, the second unit of the scheme, Calder Hall B is already well ahead. Against a background of the neighbouring Windscale plutonium factory Inset 8 shows the foundations of the second station being prepared. These were three-quarters complete when Calder Hall A was opened. Work began on Calder Hall B in July 1955 and with the experience gained in the construction of the sister station, it should be completed in appreciably less time – probably by the middle of 1958.

With the opening of Calder Hall, something that only a few years ago seemed remote and conjectural has become a working reality. But great though the technical achievement may be, it is no more than a promise of greater things to come. Already a vast programme of research and development to make this type of station more efficient and to point the way to the reactors of the future is in hand. Speaking at the British Association Meeting in Sheffield in September, Sir John Cockcroft expressed his belief that power reactors of the Calder Hall type were capable of a great deal of engineering development. One example that he gave was that the use of new alloys based on beryllium to sheath the fuel rods might lead to still higher temperatures to provide a correspondingly greater heat-extraction rate from each ton of uranium fuel.

Under the Government's ten-year 'Programme for Nuclear Energy' private industry is to build a dozen nuclear power stations for the Central Electricity Authority which between them should yield some one to two million kilowatts by 1965. The first half-dozen stations will have graphite-moderated gas-cooled reactors based on Calder Hall experience and they will be designed to produce fissile material for military purposes as well as electricity.

The system selected for Calder Hall is however only one of many and research on alternative reactor systems is being

prosecuted vigorously, and there is every indication that the next stations will show a very substantial advance on Calder Hall. It is expected that they will be able to make more economic use of the various nuclear fuels and so will be able to produce electricity more cheaply than was estimated for the first commercial nuclear power stations. These later stations may be the first in the world to produce electricity from nuclear power on a truly economic basis.

So is Britain responding to the challenge to keep in the forefront of nuclear power development so that she may lead in the application of the power of the atom for the benefit of mankind, just as, in the past, she has led in the fundamental chemistry and physics on which the new technology depends.

DEVELOPMENTS IN INTELLIGENCE TESTING

J. D. NISBET

SIR GODFREY THOMSON, commenting in 1947 on certain conflicting results on the relation of intelligence and family size, suggested optimistically that a period of confusion and controversy often is an indication of an imminent advance in theory. Such in fact seems to have occurred in the field of intelligence testing, because, though the last ten years have produced many conflicting findings, recent modifications of intelligence theory have not only succeeded in resolving the more serious conflicts but have also opened up a new prospect of development in research. A review of some of the recent findings in this field will show why such modifications became necessary.

FOLLOW-UP STUDIES

For various reasons, the post-war years produced a substantial increase in the amount of evidence available on intelligence and intelligence tests. This was partly owing to the post-war expansion of research in the social sciences, and also to the evidence derived from the war-time use of testing and its adoption on a national scale in the education system, particularly in selection for secondary education. Perhaps most important of all, the mere passing of the years has been of special value: because an essential feature of testing is the establishing of the reliability and validity of test scores, and this is a process that cannot be hurried. To know whether a child's test score is a measure of a quality permanent and important in his life, it is necessary to follow up his career through school and preferably into adult life, in order to compare the early score with later test results and with his mature achievement. Since a follow-up may need as

long as twenty years (a recent American study covered a thirty-year interval between two applications of the same test to the same group), it is only recently that results have begun to be available from such long-term studies. The results that have been obtained by this method have been most valuable in their contribution to developments in theory.

One study of this nature was begun in California in 1921 by Terman, and a description of the follow-up after twenty-five years was published in 1947. Terman sifted the school population of California by means of tests and traced in this way some 1,500 children with IQs above 135, that is, scoring in the top one per cent of the population. The careers of these children were followed, information being gathered on their education, physique, personalities, and personal histories. The results provided clear evidence that the ability measured in tests has relevance for success in education and in adult life, though there were exceptions, the group in maturity ranging 'all the way from international eminence to semi-skilled labour'. Over 90 per cent of the group went to college, and nearly 30 per cent of these took post-graduate degrees. By 1947, several were already professorial heads of departments; others were directors of research institutes; one was director of an atomic research plant, another a psychiatrist at the Nuremberg war trials, and others included an artist on Walt Disney's staff, a famous sculptor, a film director, a writer of detective stories, a distinguished composer, and a brigadier-general. Three members of the group were among the specialists invited to attend the atom-bomb tests at Bikini. By 1945, the group as a whole had published some 1,500 journal articles and 90 books, and had taken out patents for over 100 inventions. There were other findings too: their physical development and marriage rate were above average, and their mortality, insanity, delinquency, divorce, and fertility all below average.

Other long-term studies, though perhaps less dramatic in their aim and scope, have produced equally important results. Investigations such as the Harvard Growth Study of Dearborn and Rothney and the Berkeley Growth Study of Jones and Bayley,

have tested and examined children at regular intervals from birth or an early age through to maturity. These studies showed considerable variation in IQ on different occasions of testing, even more than might reasonably be attributed to chance errors or differences in test content. The early literature on testing recognized such unreliability in individual scores; but since retesting over short intervals gives correlations of $+0.9$, there grew up a widely held belief in the constancy of the IQ. Evidence from long-term follow-up was needed to shake this assumption. Thus, the Berkeley Growth Study, which in 1953 completed a twenty-five-year follow-up, reported zero correlations between developmental assessments in the first year of life and tested IQ at age eighteen. By age four, test results showed a correlation of $+0.6$ with IQ at eighteen; and assessments during the period of compulsory schooling agree with each other to a correlation of at least $+0.7$. Such agreement is high, but permits a large fluctuation in individual cases.

EDUCATIONAL GUIDANCE AND SELECTION

The increasing use of intelligence tests in schools for educational guidance has also brought out clearly the considerable proportion of children whose test scores show marked fluctuations, and this has unfortunately contributed to the scepticism of many teachers about the value of the tests. Some of the differences in individual scores, however, are due merely to differences in standardization, and would be eliminated if (as is not the case) all tests were by agreement constructed to give a standard deviation of 15. A further important source of variation is the difference in the test content, the medium of words or pictures or symbols through which the general factor of intelligence is measured: this naturally affects scores, though only to a limited extent.

Although teachers and those concerned with the practical application of tests tend to notice and remember the exceptions, there is substantial evidence to show that tests perform the task of assessment for guidance and selection at least as efficiently as any other known method. The widespread use of tests in

selection for secondary education and with National Service recruits has been accompanied by regular and careful investigations to check their validity for such selection and screening. The extent to which tests are used is now so large that extensive information on their accuracy is available. The instruction manuals of the Moray House Test series, for example, of which two new tests are issued each year, commonly report data on 40,000 or 50,000 results from each test. A number of reports on this topic have been published, and other inquiries, unpublished, are carried out regularly by Education Authorities. The results agree fairly closely in showing that the tests make an important contribution to the selection procedure.

The particular value of the intelligence test in this matter is that it helps to discount the influence of less relevant factors, such as inequalities in the standard of teaching at the primary stage. An able child who comes from a small rural school or who has suffered through absence from school or from poor teaching, is less at a disadvantage on the intelligence test than on examinations of attainment in primary school subjects. Intelligence tests also help to reduce the handicap of poor home background (though the tests are themselves influenced to a certain extent by this). An awkward problem of policy is involved here. Some forms of intelligence test, for example, those relying heavily on verbal problems, are more susceptible to the influence of home environment than others, and tend to favour children from superior homes. Because of this, attempts have been made to develop tests without such 'social bias', as it is called. Generally, these tests employ patterns or abstract symbols rather than words, and in the process of construction, in addition to the usual statistical criteria by which the acceptability of items in a test is judged, problems showing discrimination between children from different social classes are rejected. A number of these 'culture-free' or 'culture-fair' tests have been developed; but though they are of value for psychological research, their practical use in educational selection is limited. For it seems that, the more 'culture-free' a test is, the less efficient it is in identifying those best suited for higher education; the tests with bias

predict subsequent success more accurately. Presumably a good home background aids success not only in the verbal test but also in the course of higher education. Thus the educational administrator may have to choose between equality of opportunity and efficiency of selection.

Mention should perhaps also be made of the recent controversy over the effect of coaching for intelligence tests. Coaching in the technique of solving test problems can artificially boost IQ by five to fifteen points; two or three hours is sufficient and coaching beyond this time has little effect. This is obviously an important administrative problem, but it has little relevance to research on intelligence, since such coaching is highly specific to the test situation, presumably gaining its effect through the saving of time and effort in considering how best to tackle a particular type of problem.

The successful use of intelligence tests in secondary school selection has encouraged experiments to measure their efficiency for selection at other ages. High-grade intelligence tests have been included in university entrance procedures to see if subsequent academic performance can be predicted accurately; and tests have also been used experimentally in selection for higher posts in the Civil Service. But attempts to extend the scope of testing in these ways have generally been disappointing. It is not that the tests are not reasonably efficient with adult groups, for in fact they have proved useful in Army and Air Force selection, as has been mentioned above. But in groups who are receiving advanced education and who are already highly selected for intelligence, testing gives no better estimate than the existing (and publicly more acceptable) measures of academic attainment.

At the opposite extreme, it would be useful if the present crude developmental assessments of young children below the age of three could be improved into accurate estimates of later mental ability. This would assist in the early diagnosis of general mental handicap and specific defects such as deafness, and it would make easier the work of adoption societies which try to gauge the child's probable future capacity in order to match it

with the level of the home into which he is adopted. Scales have in fact been constructed for assessing precocity or backwardness in infants, and such measures show reasonable consistency over short intervals of up to one or even two years. However, as has been shown by the long-term follow-up studies mentioned above, these infant assessments bear disappointingly little relation to test results in adolescence and maturity.

DISTRIBUTION OF INTELLIGENCE

Post-war interest in population trends focused attention on the alleged danger of a decline in the level of national intelligence. The tendency for intelligent parents to have smaller than average families would seem to imply, in so far as intelligence is inherited, a gradual decrease in average intelligence in the nation. This topic was widely reported and discussed when the Scottish 1947 Mental Survey set out to apply to the entire eleven-year-old population in Scotland a group test that had been given to an equivalent population in 1932. The average score in 1947 was in fact slightly higher than the average score in 1932, though the tendency for children from small families to obtain higher average scores than those from large families was confirmed. Subsequent analysis of the results of the 1947 Survey has produced some important information on the distribution of ability, as measured by the test, in the population. Certain sociological data were obtained for a random sample of the eleven-year-olds, and though test score is certainly related to occupational status of the parent, the relation is not close. In a test where the maximum score was 76, 30 per cent of the children whose fathers were professional men scored 60 or over, whereas only 6 per cent of the children of skilled manual workers and 2 per cent of the children of unskilled manual workers reached this high level of score. Yet, because of the relative smallness of the professional group in the population, children from professional homes represented only one-sixth of the total number of children scoring 60 or more. The actual figures may make this point clear. Of the 6,627 children in the sample, 221 were from professional homes, 2,392 were children of skilled manual workers and 1,132

of unskilled manual workers. The numbers scoring 60 or more were: professional, 66; skilled manual, 133; unskilled, 24. Thus the small proportion of skilled manual workers' children who were high scorers actually numbered twice as many as the large proportion of children from professional homes. It should be remembered, however, that this refers to test score at age eleven, and superiority at age eleven is not necessarily maintained into maturity.

An example of how test scores may be influenced after the age of eleven is provided by results reported from a recent experiment in Southampton. The test scores of 865 boys aged fourteen were compared with their scores in a previous test on entry to secondary school: after allowing for regression effects, there was an average gain of five points among the grammar school pupils and an average loss of two points among the secondary modern pupils. A single experiment like this is not enough to establish that such changes occur generally, though a similar result has been reported from Nottingham. Also the result may be due to the heavily verbal type of test used, or to the influence of the pupils' homes, rather than of the schools they attend. Nevertheless, such findings emphasize the need for caution in applying conclusions from the testing of children to the distribution of adult intelligence.

In adolescence, the rate of development slows, and intelligence reaches a maximum in early adulthood. After this peak, evidence suggests that there is a gradual decline in intelligence with increasing age. The war-time use of tests in the Services provided abundant confirmation of earlier investigations on this relation between age and ability to score on an intelligence test. Wechsler (*The Measurement of Adult Intelligence*, 1944) had summarized the results:

> Every human capacity after attaining a maximum begins an immediate decline.... The age at which this maximum is attained varies from ability to ability but seldom occurs beyond 30 and in most cases somewhere in the early 20's. (p. 55)
>
> Certain abilities decline more slowly with age than others. Thus the abilities called for by the General Information and General Com-

prehension Tests hold up much better than do the abilities called for by the Substitution and the Memory Span for Digits Tests. (p. 63)

This conclusion and the supporting army data were based on the testing of 'cross-sections' of the population – equivalent representative samples of different ages all tested about the same time. This method is open to objection mainly on the grounds that it is impossible to be sure of the exact equivalence of the samples, particularly in the later age-groups. An alternative method is to test the same group twice, with a long interval between the tests. Such an experiment is difficult and time-consuming, but the last few years have seen the publication of five such studies, one with an interval of thirty years between tests. In each of these studies, the average score showed a substantial increase with age and not a decline. Certainly some of the tests used were verbal, and vocabulary is known to be strongly resistant to any decline with age, and factors of test sophistication and motivation confuse the picture. Perhaps most important, all five studies dealt with adults of higher education and superior ability; and it may well be that prolonged education in adolescence or stimulating intellectual activity in maturity fosters a resistance to the slow decline in intelligence that seems to occur in the general population.

THEORETICAL IMPLICATIONS

This review of recent work helps to bring out the main paradox to be resolved. The general validity of intelligence testing, its value in educational selection and guidance, as well as the relevance to adult achievement of the ability measured by tests have been confirmed not merely on experimental groups but on very large numbers and in well established practice. Yet evidence accumulates of this and that environmental influence on test score, until the idea that tests measure innate potentiality has become quite untenable. It is not, however, possible to dismiss intelligence test scores as merely the product of learning and environment, for even individual members of a single family usually show a considerable range of intelligence, which would be expected for an inherited trait but could hardly be explained away in terms of unequal opportunity. The correlation between the test scores of identical twins is higher than the correlation between the scores of fraternal twins, and it is difficult to explain this except by reference to a genetical basis for intelligence. Perhaps the bitterness of the controversy over this issue might be reduced if P. E. Vernon's suggestion were adopted, that the present tests should be renamed 'academic aptitude tests'; for verbal tests of intelligence this would be a sounder, though unduly restricted, description of their function. The emphasis in research on intelligence, however, has tended to move away from the old quarrel about the relative importance of heredity and environment towards identifying the precise manner in which environment affects intelligence test score.

The influence of environment on human growth and development is, of course, a notoriously difficult problem, and it is only through a clearer understanding of the nature of development that we can hope to identify and evaluate the environmental influences. The present theory of intelligence owes much to recent developments in learning theory in psychology as well as to borrowings from related sciences. Hebb (*The Organization of Behaviour*, 1949) and Piaget (*The Psychology of Intelligence*, 1950) have developed closely similar theories from quite different angles of approach. Piaget's work is based on the observation of children and a linguistic or almost philosophical method of analysis; he is concerned not with quantitative assessments of the differences between individuals, but with the qualitative analysis of children's methods of thinking. Hebb approached the problem first from the puzzling effects of brain damage on intelligence. Brain injury in adults (outside the speech area) often has surprisingly little effect on their ability to solve problems of the type in the Binet Scale of Intelligence, problems that sample widely different aspects of common everyday experience (e.g., explaining the meanings of words and the differences between words, understanding and reproducing a passage of prose, perceiving analogies and opposites, reproducing digits from memory forwards and backwards, detecting absurdities in statements, etc.). Yet similar injury in infancy or at birth may

cause permanent retardation in intelligence, including inability to score on Binet-type problems. Hebb suggests an explanation by showing the extent to which mature intelligent behaviour is built up through learning, through the establishment of 'phase sequences' in cortical activity.

An analogy is to be found in the development of perception. Bernard Shaw once remarked that he could only assume that he had been born with the gift of language, as he could not remember ever acquiring it. This kind of assumption seems to underlie our beliefs about other so-called inborn abilities. Even the perception of the world around us, apparently so obvious an example of an inborn power, is now recognized as a complex skill developed during the first year of life. When an adult, blind from birth, is given sight in maturity by surgical operation, he has extreme difficulty in learning to distinguish shapes and size and spatial position among the bewildering confusion of visual stimuli, and he sometimes even fails altogether. To state this is not to underrate the importance of maturation: in perception, as in other behaviour, maturation and experience contribute jointly to development.

Hebb's theory is neuro-psychological, and attempts to explain intelligent behaviour in terms of the building up of assemblies or phase sequences in the cortical cells. The important part of his argument for intelligence is his demonstration of the way in which early thinking in infancy develops from sensori-motor experience. From crude eye-movements and touch experience come the first ideas of shape and space; and these in turn, supported by the slow acquisition of vocabulary, lead on to primitive concepts whose rapidly increasing complexity begin to permit, even in early childhood, of the emergence of intelligent problem-solving of a low order. From this, the crucial importance of rich visual and kinaesthetic experience in infancy is clear: fortunately, the ordinary experience of most infants in this country is adequate to provide the means of development at least in this elementary stage of intelligence. Later the acquisition of more abstract concepts, often embodied in words, aids higher development in thinking. Sequences of cortical activity

which are necessary for the initial learning may function in the adult independently of the steps by which they were built up, by-passing the neural connexions that were necessary for their initial establishment. Thus a certain degree of brain injury may be tolerated by an adult in the performance of relatively familiar tasks or the solving of 'common-sense' problems, whereas similar injury in infancy would have prevented the acquiring of the necessary concepts or (Hebb) phase sequences.

An example may help to make this point clear. One item from the Stanford-Binet Scale of Intelligence suggests that since in train accidents the last carriage is often damaged, the last carriage ought to be taken off before the train starts. Of course, the child who has never seen a train or does not know what a carriage is, is unable to recognize the absurdity in this statement. But most children are familiar with trains and carriages by at least age five, yet it is not until the age of eleven that the average child is able to see through the absurdity. The difficulty is in the grasp of the sequence of ideas and particularly in the key word 'last', which involves a relatively complex idea of positional relationship. To the adult who has long ago forgotten that he once had to learn such concepts, there is no problem to solve. It is often difficult for the adult, so familiar with his extensive equipment of adaptive behaviour and verbal shorthand of thought, to appreciate the magnitude of the child's achievement in acquiring such concepts as are compressed into words like 'cause', 'somewhere', 'square', or 'in case'. At an elementary level, it takes time to acquire an understanding of 'more' and 'less', or 'under' or 'inside'.

The mature level of intelligence does not depend merely on this factor of experience. There would also appear to be differences in ability to profit by experience in the sense of ability to build up concepts of increasing complexity, some children being able to do this rapidly under suitable environmental stimulation, and others only slowly. Such differences appear to be genetical. This aspect, more remote from everyday experience but in some respects more fundamental, is what some people mean when they speak of 'intelligence'; and Hebb's solution (now widely

accepted by psychologists) is to speak of two meanings of the word 'intelligence', Intelligence A and Intelligence B. Intelligence A is the capacity for developing methods of thinking and perception, 'an innate potentiality for forming, retaining, and recombining schemata, which is presumably dependent on the genes' (Vernon), or 'the capacity for development, a fully innate property that amounts to the possession of a good brain and a good neural metabolism' (Hebb). Intelligence B is the extent to which this development has in fact occurred, 'the intelligent thinking capacities that have been acquired during infancy and childhood, which also do not develop fully in the absence of suitable environmental stimulation' (Vernon). This is the intelligence that shows itself in behaviour, not Intelligence A, which is hypothetical and cannot be measured directly. Intelligence tests which present a wide range of problems from different aspects of everyday life are sampling Intelligence B, and assessing it with an accuracy proportional to the efficiency and extent of the sampling. Vernon, in fact, suggests that we should also recognize an Intelligence C, the test score, which cannot hope to represent Intelligence B perfectly but is always only an approximation to it or summary of it. However, the important point is that test score is more closely related to Intelligence B than to Intelligence A.

This explains the low correlations between assessments of very young children and later test performance: in infancy children are still in the process of acquiring the basic concepts that underlie the subsequent level of their Intelligence B. Since mental development is cumulative in nature, the more complex behavioural sequences being built on previously mastered simpler ones, assessments improve steadily with age in their accuracy of predicting the mature level of ability. Even by age five, test scores show a fair agreement with scores at age eighteen, presumably because by age five the child has acquired the essential vocabulary structure and the basic concepts of his thinking. After about age seven, the IQ shows relative constancy: considering the differences between tests, in content, in method of administration, and in standardization, it is surprising how

stable test scores are in this period of childhood. This constancy of IQ is partly the result of the relatively standard environment shared by children in this country. Where the environment differs hugely from normal, for example among gipsy or canal-boat children, the assumption of a standard environment breaks down and tests show a steady decline in the ability of these deprived children relative to that of their contemporaries from conventional homes.

Among the main body of children, however, during this period of compulsory schooling the influence of environment appears comparatively slight. The reason for this is that IQ is a measure of relative differences between children at a certain age, and therefore, within any one cultural group where environment does not vary widely, IQ mainly reflects heredity. It would seem that, for the basic development up to the age of ten or eleven, the ordinary home and the normal contact with adults provides a growing child with adequate environmental stimulus. Where differences in home background and cultural training are extreme, as for example in the comparison of different races, of negro and white, or even perhaps in comparing upper and lower socio-economic classes in our own culture, the influence of environment appears greater in producing the observed differences between individuals. Thus it is impossible to separate the influences of nature and nurture: in a group where environment is fairly standard for all, as among the children in a single family, the differences are largely due to differences in Intelligence A, the genetical differences. When canal-boat children or Kentucky mountain children are compared with city school pupils, the contribution of environment appears much larger.

At a later stage, in adolescence and adulthood, environment would seem to play a more important role once again in that the range of differences in environment increases, both in the form of higher education and the mental stimulus (or lack of it) in one's job and leisure interests. In fact, the correlation between intelligence test score and educational attainment rises during adolescence, until in student selection, as has been mentioned above, there is close agreement between them.

The effect of age would appear to be more on Intelligence A than on Intelligence B. The falling off is in ability to master new material or elaborate new concepts, rather than in any loss of well established and regularly practised abilities. The numerous 'phase-sequences' described by Hebb become autonomous or independent of the means by which they were originally built up, so that a decline in the ability to establish new neural bonds, resulting either from age or from brain damage, has little effect on the established level of Intelligence B. Thus performance in vocabulary tests or in every-day problems of the Binet type is less influenced by age than is performance in tests employing unfamiliar symbols or patterns. An analogous observation is the commonly reported experience of mathematical theorists, that their best original ideas occur in their early work as young men, and though later they retain their mathematical abilities apparently to the full, their rich inventive years are past. If this proves indeed to apply also to fields outside mathematics, the current tendency to lengthen courses of instruction must become a matter of concern in the training of research specialists in science and medicine.

RESERVE OF INTELLIGENCE

This view of intelligence and attainment marks a radical departure from the old distinction between inborn ability and acquired information. Such a distinction was implicit in the attempts to measure 'pure' intelligence by 'culture-free' tests, and the present theory explains the relative inefficiency of these tests in predicting school and occupational success. The fact that intelligence tests are not pure measures of inborn ability but involve acquired modes of thinking, is not the weakness it might be thought: for the intelligence we apply to work and studies and everyday problems is not innate potential but depends on the conceptual development that has occurred, and this the intelligence test measures rather well. The really important questions now are: in what way and at what ages do the various environmental factors affect the development of the potential ability determined by Intelligence A, and are these environmental factors under our control or might they be brought under control?

About this extremely little is known. The first five years of life seem to be of crucial importance, and preliminary evidence suggests as influential such general factors as the nature of the home and family environment, the child's relationship with his parents, the harmony of his emotional development, the amount of contact with adults and the extent to which the home stimulates the child's acquisition of vocabulary and his familiarity with the basic concepts embodied in language. These factors, together with formal education, seem to be important also in adolescence, but nothing definite can be said of this stage either. Nor do we know how far we can improve on the present realization of potential ability, or whether existing child-rearing practices already achieve something near the practicable maximum. Hebb goes so far as to suggest that 'the country may be full of potential geniuses, for all we know, and it should be a pressing concern for psychology to discover the conditions that will develop whatever potentialities a child may have'. This sounds rather over-optimistic, and is reminiscent of the exaggerated claims of the early behaviourist, Watson. Nevertheless, it should be remembered that the distribution of ability among eleven-year-old children revealed by the Scottish Mental Survey is rather different from the distribution of ability that we see among adults, and this discrepancy may well be due to the selective influence of environment on adolescent development.

With the present scarcity of trained intelligence, it is increasingly important to be able to assess the size of the untapped reserve of intelligence in the nation. It is already known that early leaving from secondary schools, which deprives the country of much talent, is also linked to home environment. The report on *Early Leaving* (H.M.S.O., 1954) states that 'from the grammar school intake of 1946, in addition to about 10,000 boys and 7,000 girls who took advanced sixth form courses, there were about 5,000 boys and 5,000 girls who had the capacity to do so if they had stayed longer at school. Of these about 2,900 boys and 1,300 girls would have been suited for courses in

mathematics and science' (p. 59). Ten years ago there was controversy about the danger of a decline in national intelligence. Differential fertility may in fact still be causing a genetical loss of high intelligence, and from the very long-term point of view this is not unimportant. But the theory of intelligence that is described here has opened up new lines of thought and experiment. Like the problem of differential fertility, the problem of the full exploitation of intellectual ability in the nation is basically a social problem, and as such not an easy one to solve.

It would be unfortunate if the development of this new approach to intelligence testing interfered in any way with the increasing use of intelligence tests in education. Consequently, it should be emphasized that the earlier evidence has not been disproved and that there is substantial evidence of the value of intelligence tests as instruments of streaming and selection. Perhaps the new approach may help to correct unduly fatalistic ideas about the unchangeableness of the IQ: but owing to the presumably great differences between individuals in Intelligence A, it would be wrong to build up false hopes, for example, of raising severely mentally handicapped children to anything like normal standards.

In research, it seems likely that the use of a single assessment such as IQ is likely to be replaced by a variety of tests each concerned with separate aspects of intelligence. An individual test constructed by Wechsler, which is already widely used, gives ten separate scores which can be combined to provide a single IQ but which also provides information on specific aspects. The difficulty in this is that the reliability of a section of a test is closely related to its length, and from ten short sub-tests at least one of the assessments is likely to be inaccurate. Therefore comparison of scores on the separate sections of a test, as is done in differential aptitude test batteries, has to be made with caution. Lengthening the tests increases the burden on both the tester and his subjects.

The next ten years of research on intelligence thus promise to be fruitful, though the magnitude of the tasks should not be underestimated.

FURTHER READING

Dempster, J. J. B.: *Selection for Secondary Education: a survey* (London, Methuen, 1954).

Hebb, D. O.: *The Organization of Behaviour: a neuropsychological theory* (New York, Wiley; London, Chapman and Hall, 1949).

Heim, A. W.: *The Appraisal of Intelligence* (London, Methuen, 1954).

Peel, E. A.: *The Psychological Basis of Education* (Edinburgh, Oliver and Boyd, 1956).

Wall, W. D.: *Education and Mental Health* (Unesco, 1955). Appendices 1A and 1B, pp. 308–20.

Studies in Education, 7: *The Bearings of Recent Advances in Psychology on Educational Problems* (London, University of London Institute of Education, and Evans, 1955). Concluding chapter by P. E. Vernon on *The Assessment of Children*, pp. 189–215

TENSION REDUCTION AND TENSION SEEKING

J. C. ROWAN

THE idea of negative feedback has become firmly established during the past ten or fifteen years. It is now common knowledge that the governor of a steam engine, the thermostat on an electric iron, and the latest marvels of automation all work by the same means – a closed circuit in which negative feedback operates. And it is further understood, by those who have pursued the matter further, that the mechanism by which the pupil of the eye expands and contracts, and all the other self-regulating mechanisms of the body operate, is exactly similar. All these mechanisms are error-controlled. That is, the measures taken to restore stability are adjusted in accordance with the distance from a set norm, and action persists until the norm is reached and stability restored. Figure 1 illustrates the principle involved.

This frame of reference has been found to be of very wide application. Sluckin quotes the use of this general concept (which he describes as 'the basic concept of cybernetics') in the

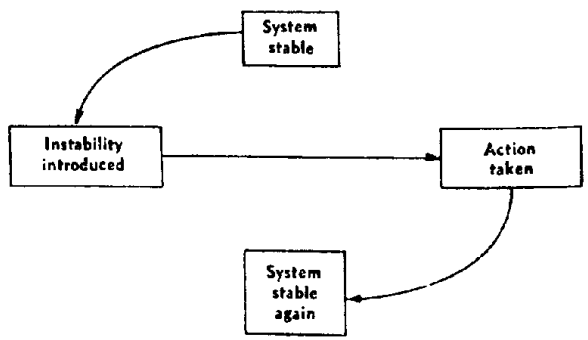


Fig. 1. Diagrammatic representation of error-controlled action.

fields of physics, biology, physiology, servomechanisms, psychology, ecology, and communications.[1] 'The query one may raise', he says, 'is why this way of picturing natural processes is tending to become so popular.' Whatever the reason may be, this concept of *homeostasis*, as it is often called, has become a dominant scientific model. We find men like Sir Geoffrey Vickers applying it even to human social affairs, and asking 'Have we, too, no goal but stability, no activity but adjustment?'[2]

There is a special form of negative feedback called 'tension reduction', which we shall consider here. It is that form of feedback where more than one type of action can be taken to restore stability.

A good example of this type of feedback is given in Jennings' description of the single-celled pond creature *Stentor*.[3] This little animalcule lives in a tube attached to the sides of a pond and spends most of its time eating. If a stream of fine particles is fed into the water currents coming towards it, it takes avoiding action by bending to one side. In most cases this is sufficient to avoid the irritating particles. If this reaction is not successful, it is usually repeated one or more times.

If the first reaction does not work, the second is used. This consisted in reversing the movement that draws nourishment into the mouth of the animal, so that the water current is driven away from it instead of towards it, and the particles are in consequence thrown off. But if the particles continue to come, this reversal is repeated two or three times in rapid succession.

If this too does not work, the third reaction is tried. This is to withdraw completely into the tube. All opportunity to obtain food is lost in this position, however, and the reaction is only maintained for about half a minute. *Stentor* then extends again: its ciliary disc begins to unfold and the cilia to draw water currents in once again. But now the first contact of the irritant particles causes it to retreat into the tube at once. This third reaction may be repeated for ten to fifteen minutes if the grains keep on coming, and after each contraction the animal stays a little longer in the tube than it did before.

At the end of this time, the fourth reaction will take over.

Stentor ceases to extend, but contracts repeatedly and violently while still enclosed in the tube. In this way the attachment of its foot to the object to which it is attached is broken and the animal is free. Now it leaves its tube and swims away, to form a new tube elsewhere.

This is an elegant example of tension reduction, because it is simple and yet illustrates the essential principle very well. As Ashby puts it, 'the animal is maintaining its essential variables within the normal limits'.[4] This is just what a thermostat does, but *Stentor* has four reactions to the thermostat's one. Whereas the thermostat merely has negative feedback. *Stentor* has second-order negative feedback, or tension reduction. Both are goal-directed.

The biologist and zoologist E. S. Russell has analysed a great deal of animal behaviour in terms of tension reduction.[5] He sums up his findings in this way: 'The same goal may be reached in different ways, and from different beginnings; the end-state is more constant than the method of reaching it.' This emphasis on the variety of ways in which action may be taken to restore the stability of an organism makes easy the transition from animal to human behaviour.

Before taking this step, however, it is important to emphasize that both in the example cited and in the general principles just outlined, the organism and its environment form a single system which acts as a whole. If we call behaviour B, the organism O, and the environment E, we may express this fact in the formula $B=f(O, E)$, that is, behaviour is a function of the organism and its environment. This is only rigorously true, however, if we define the environment in this formula as the effective environment: as Ashby puts it 'those variables whose changes affect the organism, and those variables which are changed by the organism's behaviour'. If we take his example of a bird chasing a butterfly, this is clear at once. The bird's environment while the chase continues consists of the air and the butterfly, and the butterfly's of the air and the bird.

Lewin was one of the first to apply this concept in a systematic way to human behaviour.[6] He showed that the formula

$B=f(O, E)$ could be applied to human behaviour, and recognized that the effective environment in this instance would have to be the psychological environment, if it were to satisfy Ashby's criterion. With a human being, it is not so much the real world as it might appear to a perfectly objective observer, but rather his private world, or the world as it appears to him, that enters into his behaviour.

We must restate our basic concept, then, by putting it in this psychological setting, if we want to apply it to human behaviour. This has been done in various ways by workers in this field. A good general formulation has been made in terms of human behaviour in the well-known proposition of Krech and Crutchfield which states that 'Instabilities in the psychological field produce "tensions" whose effects on perception, cognition and action are such as to change the field in the direction of a more stable structure."[7]

It is important to remember that the word 'tension' as used here does not mean nervous tension, or have any neurotic connotation. What we are doing is to use the *construct* of a system in tension to represent a need, or anything that functions as a need, such as a demand or imposition. We can then correlate 'release of tension' with 'satisfaction of need'. Tensions of this kind usually operate by setting up a goal, and the reaching of this goal will release the tension, while action that stops short of reaching the goal will leave it still unresolved. For example, if I smell burning I shall start looking for some explanation, and shall not be able to forget the matter until either I find the answer or the smell disappears.

In tension reduction the need to regain stability is not alterable, but the particular goals that may be set up and may be seen as answers to a specific problem are alterable. *Stentor* has four reactions with which to avoid the fine grains, but it cannot avoid avoidance.

A RIGOROUS TEST

When we are presented with a general concept of this kind, it is necessary to find some way of discovering if it is meaningful.

Sometimes, with general concepts, one finds that something that seems helpful and illuminating is in fact really meaningless and untestable. It is difficult not to feel this about such concepts as 'the instinct of self-preservation', 'the super-ego', 'suggestibility' and many others, for example. One of the best ways of testing the meaningfulness of a general concept is to see whether a machine can be constructed that will exhibit the behaviour postulated. Another good test is to see whether the system can be described in mathematical terms.

Where tension reduction is concerned, both these procedures are somewhat difficult. For example, to construct a machine that exhibits tension reduction involves giving it that 'persistence with varied effort' which characterizes the animal world, and which has often been held to distinguish it absolutely from the world of machines.

Nevertheless, the task has been attempted. Sommerhof, in his *Analytical Biology* (1950) made some progress, but the most successful attempt so far is Ashby's work already mentioned. He has constructed machines, the first of which was the homeostat, which seem to fulfil what is required, and he has also carried out the basic mathematical analysis. This work is of the utmost importance, because if it is successful it means that psychology, the science of behaviour, has a rigorous mathematical basis for further action and research. Nowhere more than in psychology has the lack of such a basis been more damaging, because it left a wide field open for quacks and cranks of many types. If we can on such a basis build a machine that exhibits tension reductive behaviour, using various means to restore stability, retaining those that prove successful, persisting with different methods if success is not reached by the first, and so on, then we can try out on it all those perplexing problems of goal-seeking and mechanism, free-will and determinism, which often prove so intractable when living things are in question.

Let us compare the homeostat with *Stentor*. In both, the unit of behaviour will be the molar unit. This is the unit we have been using all along so far; it is the complete cycle of behaviour as shown in Figure 2.

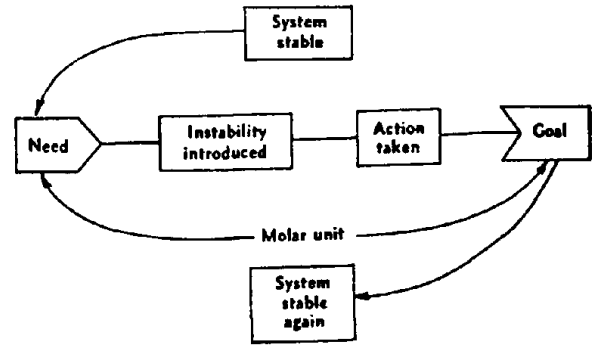


Fig. 2. Diagrammatic representation of molar unit.

The homeostat has four magnetic units, each with a pivoted arm which droops into a curved trough of water. When an arm points to the middle of the trough, all is normal in that unit. A deviation to either side represents some departure from normal, and a deviation of more than 45° causes magnetic action to be taken to restore the arm to its central position.

The feedback circuit passes through three uni-selector switches. These switches have several positions, and each position contains components whose values have been given at random. The whole is so arranged that the uni-selectors will change from one random value to another as long as the magnet deviates from normal; as soon as they select a value that restores the magnetic arm to normal, they stop.

Suppose we start off a cycle of behaviour by changing the input to one of these units. The unit will respond; the arm will be deflected past its critical position of 45° and this will bring about a uni-selector change. If the new value fails to restore the stability of the circuit, a second value is tried. This is a random change, so is no better than the previous one; it is merely different. If stability is still not reached, a third value will be tried, and then a fourth. Let us suppose that the fourth value restores stability, and the magnetic arm returns to its central position. The molar unit is now completed.

Compare this with the description of *Stentor*. In both a disturbance is introduced by the experimenter. In both different responses are used, and in both action ceases with the first response found to be successful. In both the responses are of step-function type, that is, one type of behaviour is changed to a quite different type of behaviour. In both the succeeding trial is not specially related to the preceding, nor is it better than it, but only different. For example, in *Stentor* the first and second reactions may be reversed. And in neither experiment do we know precisely how the successful reaction has been reached, though we do know that action will persist until such a reaction is attained.

This last point is particularly significant. With a homeostat, the experimenter does not know what values will be hit upon by the uni-selectors, though he does know that they will keep on changing until a suitable value is found. When the molar unit is completed, the only way he can find out how the homeostat reached stability is by examining the inside of the machine, much in the manner of a dissection. The homeostat, too, cannot avoid avoidance.

THE BASIC ASSUMPTIONS

Those who have used this approach to the study of machines and organisms have tried to achieve thereby increased rigour in the description of behaviour. In doing so, they have found it possible to be extremely rigorous provided that they held to certain basic assumptions: (*a*) the organism and its effective environment form a single dynamic system; (*b*) this system exhibits tension reduction; and (*c*) in analysing its behaviour, the molar unit must be used.

This approach has led to some of the most interesting and rewarding work being done at present in the field of psychology. The most adequate statement of the approach is probably that found in the work of Krech and Crutchfield already quoted; their account of perception, for example, is one of the best expositions to be found in the literature. If we abandon any of these three assumptions, we shall lose our rigour, and, in particular, any possibility of retaining our mathematical description. For example, if a molar unit has not reached its end-phase, we cannot yet call it a molar unit. It only becomes a unit, and hence a tool of analysis, if it is completed.

Difficulties

'In every case what is happening is essentially an attempt to restore equilibrium to a system of activities which has been disturbed.'[8] That is the position. We have given one or two examples already. But let us look at a rather different example.

Men will often work extremely hard to attain positions of greater responsibility. Such men are obviously moving in the direction of tension increase rather than tension reduction. It may be true that as each new problem presents itself it is dealt with along tension reductive lines, but it seems difficult to regard the seeking after such problems in the same light. And this type of activity, which we may call tension seeking behaviour, is very common. The young man going alone to a dance, the young woman enrolling for a difficult course, the scientist seeking for instances that would disprove his hypothesis, the man setting out to climb a mountain, are all, in their various ways, choosing situations in which their tensions will be increased.

Can we get over this difficulty by making a distinction between different orders of molar units? Can we retain our rigour by postulating a hierarchy of molar units, in which the smaller units find their places within a structure of larger units? Muenzinger thinks that we can,[9] and indeed it does seem that in certain instances this distinction will serve well enough. If we take the first of the examples just given – the man seeking greater responsibility could be described as having the goal of an adequate salary. Such a salary would be an answer to the major tensions now bothering him, and the reaching of this goal would mean to him an overall decrease in tension. In this context the tension seeking behaviour would be understandable and would fall into its place within the more comprehensive unit.

But supposing that he never does reach an adequate salary? Much of human behaviour is dominated by goals that are never

reached, and even some that are in principle unreachable, such as those sought by some saints, scientists, athletes, artists, politicians, and reformers. It seems that introducing larger units is no way out.

More Difficulties

Whether or not we introduce larger units, we are still left with a set of problems by no means peculiar to human beings. These problems are well illustrated by the work of Beach, James, Thompson, Heron, and Melzack on early environment.[10] Scottie dogs were kept in closed boxes from the time they were weaned (4 weeks) to the age of 7–10 months. Each box had two compartments and one dog. The box was opaque and the dog could not see outside. Figure 3 shows the experimental set-up.

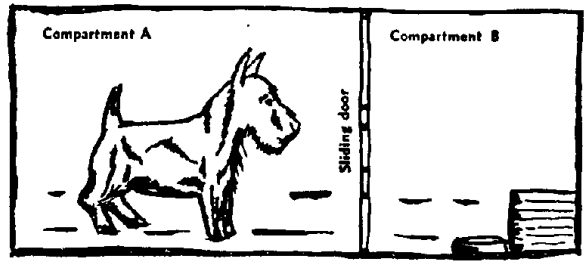


Fig. 3. Representation of arrangement used in rearing dogs.

Each day food was put into the compartment B, the sliding door opened, and the dog went in to eat; the sliding door was closed again, and compartment A cleaned out. Afterwards the sliding door was opened again, the dog came back into its living-room, the sliding door was closed, food-pans removed and all was ready for the next meal-time. Thus for six to nine months the dog never saw or smelled the outside world, or another dog, or its own keepers.

At the end of this time the experimental dogs were let out and given various tests, the results of which were compared with those of another group of dogs (the other halves of the same litters) which had been reared in the ordinary way by Montreal families.

In common with most dogs Scotties are normally fairly sociable animals under certain conditions. To test what difference the experiment had made on this attitude, the experimenters put in a large room two dogs, enclosed in separate chicken-wire pens. They then released into the room one of the normal dogs. It would usually approach one of the pens, stare at the other dog inside, bark and wag its tail. Then the experimenters tried letting in one of the restricted dogs instead. But the restricted animal would usually ignore the pens, and would spend his time exploring the room. This type of behaviour, amounting to a lack of interest in companionship, did not disappear rapidly, but persisted for years after the termination of the experiment.

There was no lack of vitality in the experimental dogs – in fact, when visitors were asked to point out which were the normal and which the experimental dogs, they would usually point to them in reverse, because the formerly restricted dogs were so much more lively and playful.

Yet their behaviour was far from normal. They were more emotional, less intelligent, more curious, less able to avoid pain, more submissive, and generally more immature than normal dogs, and many of these effects were persistent.

Now the main difference between the normal and the experimental dogs was the variety of different tension situations that the dogs experienced. The normal dogs had a rich and stimulating early environment, and experienced a multitude of tensions of various kinds, while the experimental dogs had a bare and uneventful life, with a far more restricted range of tensions. It looks as though animals actively need tensions, if they are to become reasonably normal adults.

Not only is there this general need for tensions, but at a certain stage in the life of an animal a particular source of tension may be vitally necessary, and if it is not provided at that stage its loss may be irreparable. For example, Scott, Fredericson, and Fuller report how a lamb has only to be removed from the

society of other sheep for the ten days immediately following birth, and it will never afterwards follow the flock or become strongly associated with it.[11] Both in the general and in the particular ways mentioned animals need tensions, and higher animals need them more than any others.

TENSION SEEKING

It is by resolving tensions that a man develops his life capacities. Krech and Crutchfield recognize this when they say 'Most blockages can, as a matter of fact, be regarded as beneficial and adaptive, for they energize the individual's efforts toward achievement of his goals and cause him to reorient, learn and grow.' If an individual has various reliable mechanisms for actively seeking out tensions, his abilities will develop so much the faster.

Muenzinger's findings with rats are also relevant here. He made rats choose between two alternative paths to get food, using the ordinary discrimination-box technique. He used hunger as the motive, and noted the number of trials before the correct path was consistently chosen, and the average number of errors. Then he added a further element in that the rats were punished, by a mild electric shock, whenever they made an incorrect choice. And then he made a third series of experiments in which the rats were punished instead each time they made a correct choice.

Now on the basis of the tension-reduction theorem we should expect, it is fair to say, that the rats would do best on the second experiment, less well on the first, and least well on the third. Yet here are the results actually obtained:[9]

Test	*Average Errors*	*Trials*
Hunger-food motivated	30	114
Hunger-food plus shock for wrong choice	11	39
Hunger-food plus shock for correct choice	14	48

It seems here that increased tension, far from creating a block-

age, actually made learning easier, whether it was applied for wrong choices or for right. Once again tensions seem to be playing much more than a merely negative part.

With all we have said so far, it does seem as though tension seeking is as important as tension reduction in accounting for behaviour, particularly in the higher animals, and most of all in man. Yet the scientists who have done so much to clarify our ideas by means of the concept of tension reduction have not always treated this side of the matter adequately. They have tended to treat the tension as given, by some limitation arising internally or externally, and to devote their main energies to describing how it is dealt with.

It is certainly true that Krech and Crutchfield do try not to leave tension seeking out of account when they say that 'the successive states of equilibrium in the psychological field are not static but constantly changing, often towards progressively higher levels'. Snygg and Combs, of the so-called 'phenomenological school', try to escape from the difficulty by introducing the concept of 'self-enhancement',[12] but neither of these approaches seems to be reconcilable with Ashby's rigour. Both suggest an afterthought grafting on something that ought not to be left out of account, rather than treating tension seeking and tension reduction as equal in importance. Yet advance in this field seems to depend on this problem being solved.

The Problem

What approaches are being made towards its solution? We have already mentioned two lines of research that have produced interesting results, in the Canadian work on dogs, and Muenzinger's work with rats; work similar to the latter has also been done with human beings both by Muenzinger himself and by H. F. Wright,[13] who showed that under certain conditions the presence of a barrier to a goal intensified effort.

In the field of cybernetics proper we may remember W. Grey Walter's *Machina speculatrix*, the little tin tortoise that runs about actively exploring its environment by means of a photoelectric cell receptor so arranged as to be constantly scanning

the whole horizon.[14] The receptor is on top of the machine, and rotates continually. When this machine sees a light, it goes straight towards it, but otherwise it explores the environment looking for light. Here at least, then, is a machine that does go out and look for experience and does not simply react to incoming stimuli.

Curiosity is a type of behaviour which is itself worthy of further investigation. The work of Butler and Harlow at the University of Wisconsin and of Welker at the Yerkes Laboratory[15] show that curiosity in rhesus monkeys and in chimpanzees is as reliable a motive as hunger, thirst, or the avoidance of pain. Curiosity in these animals varies somewhat with the object involved. Experiments in discrimination were made using as a reward not food, as with the rats mentioned earlier, but simply the opening of a shutter so that the monkeys could see what was on the other side. The discrimination was learned quickest when another monkey was seen through the shutter, and only slightly less quickly when an electric train running on a curved track was seen. A bowl of fruit was less potent, and an empty room less still, though it was still effective.

In this and other tests, young monkeys and young chimpanzees consistently showed more curiosity than older ones, persistence sometimes being remarkable: in one test where the taking apart of a mechanical puzzle was the only reward for solving it, the experimenters reassembled it every six minutes, and the monkey went on taking it apart again for ten hours 'at which point the experimenters, rather than the subjects, had had enough'. It seems that spontaneous curiosity is a stronger motive than has sometimes been thought.

Tension seeking behaviour is not to be identified with curiosity, of course. Light is thrown on it, too, by other types of behaviour, of which one example is to be found in the water flea, *Daphnia*. J. L. Brooks in experiments in America showed that those species of *Daphnia* which under normal conditions grow large helmets fail to grow them in the laboratory tank unless a random turbulence is imparted to the water.[16] And so we are back to Ashby's suggestion that the concept of random

stimulation and activity has a leading part to play in the understanding of adaptation and development.

CONCLUSION

It is not sufficient merely to collect information about tension reduction and separate information about tension seeking and then try to put the two together. What is needed is a model, or a set of concepts, in which both forms of behaviour can find an honoured place. The insight which the concept of tension reduction has given us is a permanent addition to our equipment, and no one could wish to throw it overboard. What the present writer has tried to do is to exhibit some types of behaviour that will have to be adequately explained by the more comprehensive theory that we are awaiting. The work continues. When it succeeds, or rather in that measure to which it succeeds, we shall be able to build a machine that behaves like a living organism. More important, we shall be able to understand better what makes an individual or a group function in a healthy and well-adjusted way, and we shall be better able to help individuals and groups who are sick or disordered.

REFERENCES

1 Sluckin, W.: *Minds and Machines.* (Pelican, 1954.)
2 Vickers, Sir G.: 'Decisions and Valuations', *The Listener*, 26. Jan. 56.
3 Quoted in 4 and 5.
4 Ashby, W. Ross: *Design for a Brain.* (Chapman & Hall, 1952).
5 Russell, E. S.: *The Directiveness of Organic Activities.* (Cambridge University Press, 1945.)
6 Lewin, K.: *Field Theory in Social Science.* (Tavistock Publications, 1952.)
7 Krech, D. and Crutchfield, R.: *Theory and Problems of Social Psychology.* (McGraw-Hill, 1948.)
8 Ogden, C. K.: *The ABC of Psychology.* (Pelican, 1944.)
9 Muenzinger, K.: *Psychology: The Science of Behaviour* (Harper, 1942.)
10 Quoted in *Scientific American*, January, 1956.
11 Quoted in J. M. Tanner (Ed.): *Prospects in Psychiatric Research.* (Blackwell, 1953.)

12 Snygg, D. and Combs, A. W.: *Individual Behaviour.* (New York, 1949.)
13 Quoted in 7.
14 Walter, W. Grey: *The Living Brain.* (Duckworth, 1953.)
15 Quoted in *Scientific American,* February, 1954.
16 Quoted in Foerster, H. Von. (Ed.): *Cybernetics: Ninth Conference.* (Macy Foundation, 1953).

SEMI-CONDUCTOR P-N JUNCTION DEVICES

R. D. KNOTT AND J. I. MISSEN

SEMI-CONDUCTORS are solids which are insulators at a temperature of absolute zero but which conduct electricity by the passage of electrons at normal temperatures. Germanium and silicon are the two semi-conductors of major practical importance in p-n junction devices.*

ELECTRICAL CONDUCTION BY SEMI-CONDUCTORS

An understanding of electrical conduction in semi-conductors comes from a knowledge of the form of inter-atomic binding in perfect crystals. Both silicon and germanium are tetravalent elements, with four electrons in the outer shell; these valency electrons are the important units in electronic and chemical processes. The remaining inner electrons are bound to the nucleus giving a stable, positively charged atomic 'core'. At absolute zero in the lattice structure of a perfect semi-conductor crystal every atom is bound to its four neighbours, each bond comprising two electrons, one from each of the two atoms that are bound together. All the valency electrons are used up in these stable covalent, or electron-pair bonds; in fact the electrons in them can be considered as the structural units that hold the crystal together. This condition is satisfied in the diamond crystal lattice, in which a typical atom is surrounded by four neighbours placed at the corners of a regular tetrahedron, as shown in Figure 1.

As the temperature is raised, sufficient energy transfer occurs to break some of the covalent bonds and under the influence of thermal agitation the electrons that are ejected from their

* For definition of a p-n junction see p. 101.

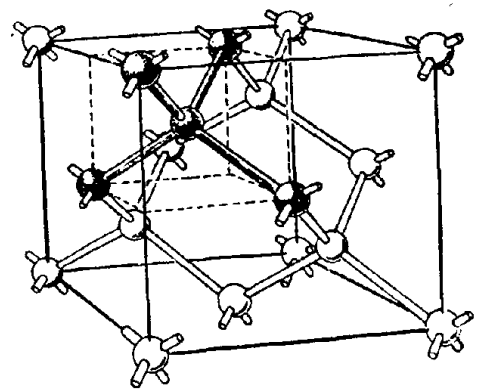


Fig. 1. The crystal structure of a semi-conductor showing how it forms four covalent bonds with its nearest neighbours.

sites in the lattice migrate in a random manner in the crystal. Further, the 'holes' that are left in the bonds by the removal of the electrons behave as positive charges. They also can move about in the crystal as a result of the reciprocal motion of electrons in the valency bonds. The number of 'electron-hole' pairs produced increases exponentially with temperature, their density at any particular temperature depending on the semi-conductor under consideration. It is greater for germanium than for silicon.

When an electric field is applied to a semi-conductor, the ejected electrons drift in opposition to the field and the positive holes in the direction of the field: that is, they add together to give a conventional electric current in the same direction as the field, as shown in Figure 2. On reducing the temperature once more, electrical conductivity decreases. This is brought about by recombination of electrons and holes because, if an electron drops into a hole, both the electron and the hole disappear and the bond is completed. Other processes may produce electron-hole pairs. For example, a light photon falling on the crystal can give up its energy to an electron which is then ejected from its bond.

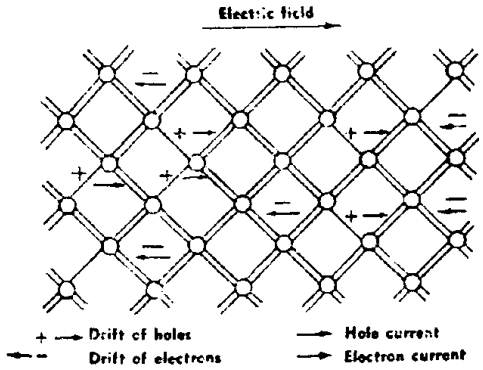


Fig. 2. Drift of holes and excess electrons in an electric field (the figure is a plane diagram of the diamond lattice with the covalent bonds represented by lines).

Conduction by equal numbers of electrons and holes is termed *intrinsic* conduction. In practice this effect is usually outweighed by a second process known as *extrinsic*, or impurity, conduction. The latter can arise as a result of the deliberate introduction of small numbers of either pentavalent, or trivalent, impurity atoms into the crystal lattice of the semi-conductor. In the first case each impurity atom displaces a semi-conductor atom from its site and uses four of its five valency electrons to complete covalent bonds with its neighbours. The fifth electron is free to act as a current carrier at normal temperatures, while the impurity atom remains as an immobile positive charge. Conduction due to pentavalent, or 'donor' impurities, for example, arsenic, is thus due to excess electrons and is termed n-type conduction. In a similar way, a trivalent atom is unable to complete the covalent structure around it. The vacancy, or hole, can be filled by an electron from an adjacent bond to give a 'hole current' at normal temperatures. Conduction in this case is referred to as p-type conduction and the impurity atoms as 'acceptors' [illegible] this type.

N- and p-type conductivity is illustrated in Figure 3 and the various terms already discussed summarized in Table 1.

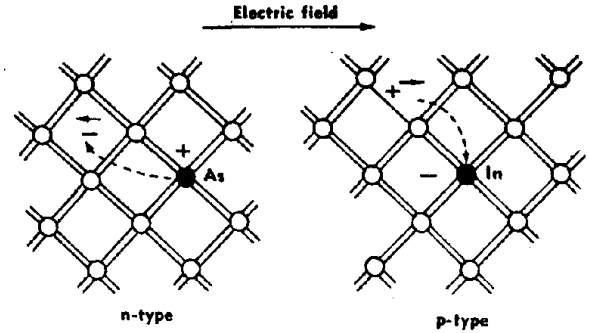


Fig. 3. Extrinsic or impurity conduction in n- and p-type semi-conductor.

THE P-N JUNCTION

An intimate junction between p- and n-type semi-conductors forms a rectifying barrier. Such junctions occur when a piece of semi-conductor has a variable concentration of donor and acceptor centres so that a transition from p- to n-type takes place in a continuous single crystal. Merely placing two pieces of semi-conductor of opposite conductivity together does not constitute a p-n junction since layers of surface oxide, roughness, and other factors prevent its formation.

Figure 4 shows a p-n junction under thermal equilibrium and bias conditions. Since the 'hole density' is greater in the p- than in the n-region, holes tend to diffuse to the n-region; a tendency that is prevented by an electrostatic field so arranged as to make the potential more positive in the n-region. Some holes will acquire sufficient energy to climb this potential hill and diffuse into the n-region where they recombine with electrons. This hole current is exactly balanced by the rate at which thermally generated holes diffuse from the n- to the p-region. A similar situation holds for electrons. When a negative potential is applied to the p-region, the height of the

potential barrier increases so that a negligible number of holes surmount it. Hence, the hole current into the n-region is practically zero. The hole current from the n-region is, however, substantially the same as in the unbiased state, depending only on the thermal rate of generation of electron-hole pairs. Hence the reverse hole current reaches a small limiting value, that is, it saturates. Precisely similar comments apply to the electron flow under reverse bias.

TABLE I

Conductivity type:	n-type	p-type
Conduction by:	Excess electrons	Holes
Energy band in which carrier moves:	Conduction band	Full or valency band
Sign of carrier:	Negative	Positive
Valency of impurity atom:	5	3
Name of impurity atom:	Donor	Acceptor
Typical impurities:	Phosphorus P Arsenic As Antimony Sb	Boron B Aluminium Al Indium In Gallium Ga

In the forward biased condition, positive potential is applied to the p-region, the hill is diminished in height, so a large hole current flows over it into the n-region, and a large electron current into the p-region.

The current-voltage characteristic of a typical p-n junction is shown in Figure 4. To complete the discussion of its form, mention must be made of the 'breakdown' that takes place at a critical reverse bias. This occurs when charge carriers reach sufficiently high energies in the barrier region to eject further electrons from the valency bonds. The result is an avalanche effect similar to ionization in a gas discharge tube.

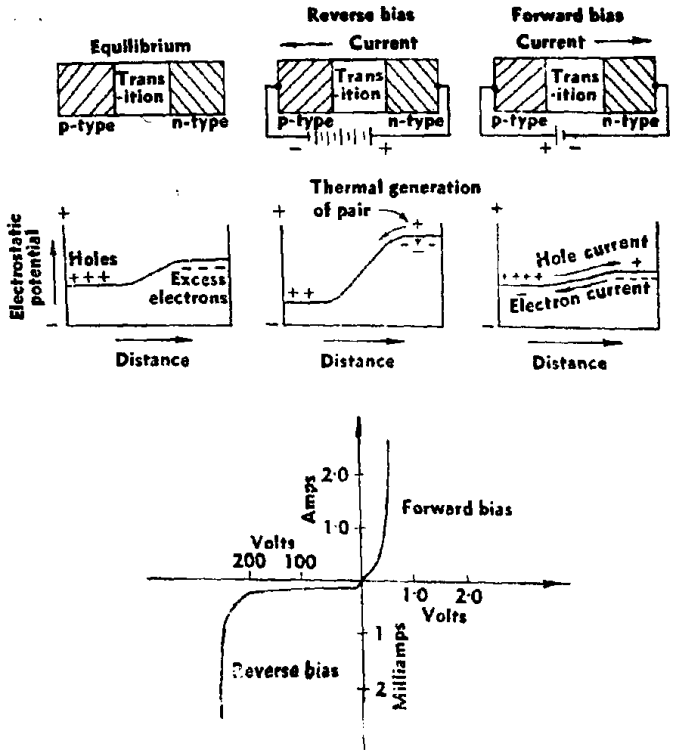


Fig. 4. The p-n junction under equilibrium and bias conditions, and a typical characteristic.

THE JUNCTION TRANSISTOR

In the junction transistor two p-n junctions, the 'emitter' and the 'collector', are formed in a piece of semi-conductor. The centre region between the two junctions, or 'base' region, may be of either conductivity type, giving rise to a p-n-p and n-p-n transistor structure depending on whether the base is n- or p-type. We shall discuss the p-n-p unit, which is shown diagrammatically in Figure 5, together with the bias supplies which

activate it as an amplifying device. Bearing in mind the duality of electrons and holes, a similar treatment can be applied to the n-p-n transistor.

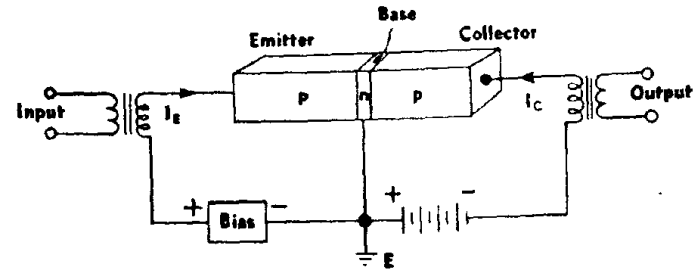


Fig. 5. The p-n-p transistor biased to act as an amplifying device.

When the collector is negatively biased it has a high impedance so that only a small electron current flows to the base. If the emitter is now biased slightly positive it injects holes into the base region. Under the influence of the collector field these migrate towards the collector and, provided that the base region is suitably narrow, substantially all of these holes will reach, and be swept through, the field, to be collected as an increment of controlled collector current.

The variation of collector current in response to a change in emitter current at a fixed collector bias is called α the current gain. In the junction transistor α approaches unity but, generally, does not exceed it. Because of the finite time taken by the charge carriers to travel from the emitter to the collector, the value of α decreases as the frequency of the signal being amplified is increased. It is customary to use the term alpha cut-off frequency, or f_α, for the frequency at which α has decreased to 0.7 of its value at low frequency.

As an amplifying device, the transistor may be connected in the three basic configurations shown in Figure 6. These are designated the common base, common emitter, and common collector configuration, and correspond approximately to

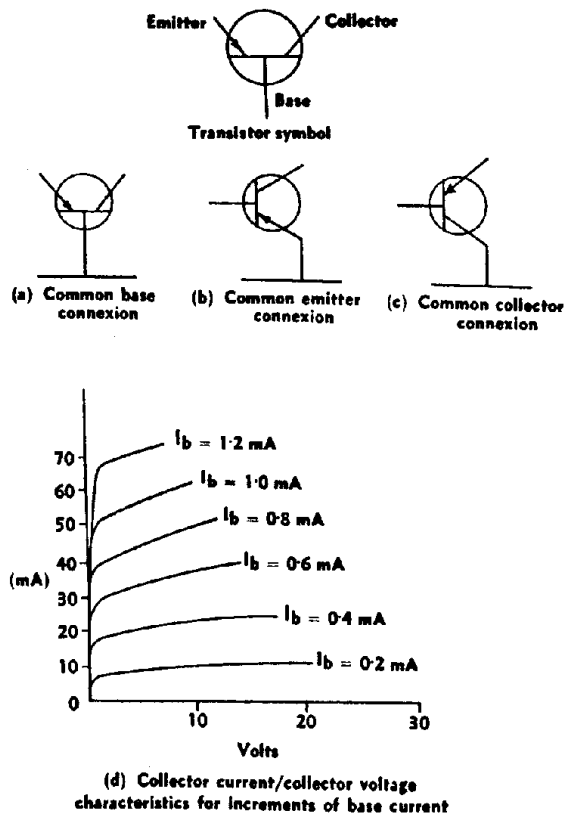


Fig. 6. Transistor circuit configurations and characteristics.

grounded grid, the conventional grounded cathode, and cathode follower circuits in thermionic valve terminology.

In discussing transistor action, the common base circuit configuration was described. If now the common emitter circuit is considered, the signal current (i_b) into the base can be regarded as the difference between the emitter current (i_e) and the collector current (i_c). Then the ratio of collector to emitter current

$$i_c/i_e=\alpha \text{ and } i_b=(i_e-i_c)$$

$$\therefore i_c/i_b=\alpha/(1-\alpha)=\alpha_{cb}$$

Typically α may have a value 0.98 whence $\alpha_{cb}=49$. Thus, if a transistor is operated with emitter earthed, and a signal is applied between base and earth, a current amplification of nearly 50 is possible for a typical α of 0.98.

The input resistance of a transistor in common base and common emitter connection is low, while the output resistance is relatively high. Power gain occurs in the transistor due to the effective transfer of current from the input to the output. Hence we get a considerable power gain even in the common base amplifier where α is approximately one. Owing to the much higher current gain of the common emitter amplifier the overall power gain is greater than that of the common base amplifier. In the common collector amplifier the voltage gain is unity or less although a current gain of approximately α_{cb} can be obtained so that considerable power gain is possible. In common with the cathode follower, it presents a relatively high input impedance and low output impedance. At low frequencies the common emitter arrangement is used when maximum power amplification is required. At higher frequencies the position is not so clear because α_{cb} varies more rapidly with frequency than does α. The seriousness of this depends on the particular circuit application and the extent to which such circuit techniques as negative feedback are applied. However at frequencies higher than about 0.3 f_α, it is often found that greater and more reliable amplification can be obtained by using the common base arrangement.

In Figure 6(d), transistor characteristics are shown giving collector current/collector voltage characteristics for increments of base current. A similarity to pentode characteristics may be noted. Transistor characteristics have the advantage of high linearity and of permitting operation to low collector voltages. Transistors can be used at supply voltages of as little as 0.15 volts although it is general practice to use a voltage in the range 1.5 to 24 volts, depending on the particular application.

PRACTICAL P-N JUNCTION DEVICES

In order to produce a p-n junction device, the piece of semi-conductor that is used must be part of a high quality single crystal which has an accurately controlled impurity content. It cannot be perfect in the sense described earlier, but it must be free from major defects of the type that are found in many natural crystals. Conductivities of from 0.02 to 2 reciprocal ohms per centimetre are used in devices and these correspond, in n-type germanium for example, to a ratio of impurity element to germanium atoms ranging from about one in 1,000,000,000 to one in 10,000,000.

The key process used in achieving these very high states of purity is zone refining. In this process, a narrow molten zone is made to traverse the length of the ingot as a result of which the impurities are segregated at one end. This contaminated end portion can then be cut away leaving an ingot of high purity, polycrystalline material. This is re-melted in a crystal growing unit and the correct quantity of p- or n-type impurity added. A seed crystal is then dipped into the melt and withdrawn slowly so that a large single crystal grows from it. In this way crystals weighing more than a kilogramme have been made.

The three most important methods that have been developed for producing p-n junctions and transistors are as follows: (*a*) changing the impurity type during the growth of the single crystal, which may then be sawn up into units of the desired size; (*b*) the alloy method, in which a layer of opposite conductivity type is redeposited on a wafer of semi-conductor from an alloy of the semi-conductor and impurity element produced *in situ*; and (*c*) by solid state diffusion techniques, in which the semi-conductor is exposed to the vapour of an impurity of the opposite conductivity type.

An example of a junction made on n-type germanium by the alloy method, using indium as the p-type impurity, is shown in Insets 15 and 16. To form this junction the germanium wafer and indium were heated in a hydrogen atmosphere to 500° C and then cooled slowly. As the temperature increased the indium

melted and formed a liquid indium-germanium alloy which, on cooling, redeposited a thin layer, containing sufficient indium to make it p-type, as a continuation of the original crystal.

To complete a p-n junction rectifier, 'ohmic', that is non-rectifying, contacts are made to the p- and n-regions and the unit is hermetically sealed in a container. Hermetic sealing is essential with all crystal devices because their electrical properties are impaired even by minute traces of impurities, such as water vapour, on the semi-conductor surface.

A range of semi-conductor rectifiers is now available with junctions of up to one inch in diameter. A range of devices is shown in Inset 17. At the rear a water-cooled rectifier capable of passing currents of some 100 amperes is shown. In the foreground is shown a selection of smaller silicon and germanium rectifiers, and transistors capable of output powers of hundreds of milliwatts to ten watts in push-pull connection.

In the transistor field there are three interdependent design criteria, namely: maximum power rating, upper frequency limit, and maximum operating temperature. Power rating depends primarily on the rate of flow of heat away from the collector junction. In power transistors that have to dissipate heat equivalent to several watts, the collector is usually coupled to a heat sink by way of the container that holds the device. Heat flow then depends on the area of the collector junction. Smaller transistors rely for cooling on conduction of heat along the leads connected to the three electrodes.

With increasing frequency the value of the current gain decreases, thus giving a frequency limit for the transistor. This frequency limit is affected by the area of the collector junction, the base width, and the mobility of the carriers in the base region. To achieve a high cut-off frequency, small junction area, narrow base width, and high carrier mobility must be combined. The last is a property of the semi-conductor, germanium being preferred to silicon in this respect. As an example of the other two variables, germanium p-n-p transistors with a collector diameter of 0.005 inch and a base width of 0.001 inch can be used up to about 1 Mc/s. To achieve 500 Mc/s, the highest yet

claimed, a collector diameter of 0.0015 inch and a base width of 0.00006 inch are necessary. This is outside the scope of all but the solid state diffusion techniques.

With regard to maximum operating temperature, silicon has a marked advantage over germanium, 150° C being possible compared with 85° C for germanium.

APPLICATION OF SEMI-CONDUCTOR P-N JUNCTION DEVICES

A tree showing some applications of p-n junction devices is shown in Figure 7. Point contact devices have not been included since they were the subject of an earlier article in *Science News*.

The p-n Junction Diode

The junction diode forms a very large branch of the semiconductor tree. It possesses a number of very desirable features. Reference to the characteristic of a germanium junction diode in Figure 4 shows that it approaches closely to the ideal rectifier in behaving like a switch with very low resistance in the forward direction, and very high resistance in the reverse direction, and that it is capable of withstanding a high reverse voltage. Compared with germanium, a typical silicon junction diode would exhibit a slightly higher forward resistance, but would possess a reverse resistance of many megohms and would withstand 400 volts or more in the reverse direction. It would also be capable of operation at higher temperatures. As a result of the low forward voltage drop, particularly in germanium, p-n junction diodes are very efficient as rectifiers for converting alternating into direct current. This is particularly marked in low voltage high current supplies, where in the past, using selenium or mercury-arc rectifiers, the forward voltage drop across the rectifier circuit was an appreciable fraction of the D.C. voltage across the load. However, for rectifying at higher voltages, p-n diodes show considerable advantages over selenium rectifiers, since they will withstand higher reverse voltages, so that fewer are required in series, thus permitting higher efficiency.

Finally, one very desirable feature of p-n junction diodes is

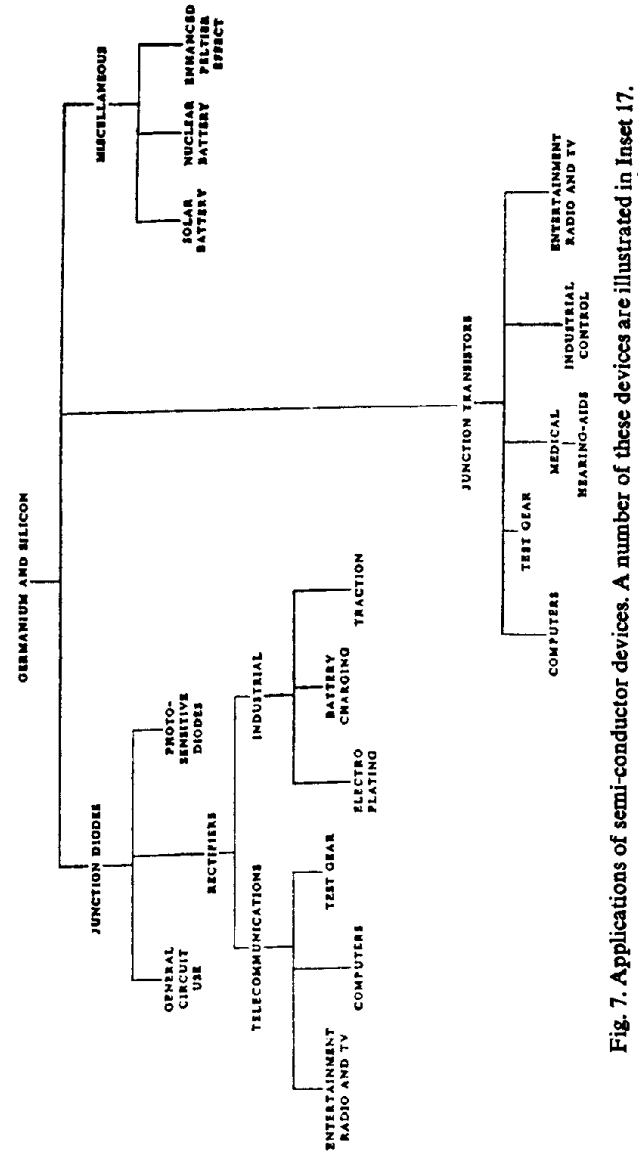


Fig. 7. Applications of semi-conductor devices. A number of these devices are illustrated in Inset 17.

the fact that they do not 'age'. Ageing normally presents itself as an increase in forward resistance with time, which causes efficiency to drop and ultimately results in diode failure due to overheating. Hence p-n junction diodes should operate for many thousands of hours without failure.

Such devices will in future be used in a large number of applications as replacements for thermionic and selenium diodes. Thus the average domestic TV receiver uses approximately ten diodes for general circuit use and as power supply rectifiers. Up to the present, these diodes have been small thermionic types with point-contact germanium diodes as detectors, and selenium rectifiers for the H.T. supply. It is expected that replacement of the thermionic diodes by p-n diodes will ultimately reduce assembly costs, and result in greater reliability. Replacement of the selenium rectifier should provide greater efficiency and result in fewer faults, since the p-n junction rectifier is not subject to ageing. More complicated telecommunication equipment and computers require p-n junction diodes in large numbers.

High current applications for rectifiers are possible in such processes as plating, battery charging, and various electrochemical techniques. However a number of problems in the use of semi-conductor rectifiers still present themselves, such as their operation in series and parallel and the protection of the diodes against short circuit and overloading. Protection against excessive currents is more difficult in circuits using p-n junction diodes for two reasons. First, owing to their lower forward resistance, less current limitation is provided and a higher current can flow under short circuit conditions. Secondly, as a result of their smaller junction area, and hence smaller size, their thermal capacity is very much less than comparable selenium rectifiers. Hence if the output from a p-n junction rectifier is short-circuited, the rectifiers may be destroyed due to excessive temperature rise in a time of the order of twenty milliseconds or less. This contigency can be guarded against by the use of suitable fuses and other circuit breakers and, possibly, also by redesign of the supply transformer.

Before leaving the topic of junction diodes, some of their other properties may be mentioned. One of the most important subsidiary properties of the p-n junction is its photo-electric properties. This was briefly mentioned earlier; it results in the junction acting as a source of e.m.f. when light falls on it.

The phenomenon may be used as the basis for a light sensitive device where the resulting current operates a relay or indicating instrument. If the device is suitably designed and of large area, it may be utilized as a 'solar battery' which provides a source of power when sunlight shines on it. In sunny climates such devices offer great possibilities and 'solar batteries' have been used in the U.S.A. as H.T. supply for telephone repeaters, particularly in isolated locations where power costs are considerable. They can produce 30 watts per square metre of junction area, at an efficiency of over ten per cent. The nuclear battery utilizes the e.m.f. produced in a p-n junction when radio-active particles bombard the junction. These particles produce an effect similar to photons in that they cause holes and electrons to surmount the potential barrier. Potentially a battery with a very long life should be possible.

Furthermore, by-products of nuclear fission, such as strontium-90, may be used as a source of the radio-active particles so that costs could be kept small. At the present time, however, such a battery is not a practical proposition since the bombardment of the junction causes dislocations in the crystal lattice.

The Transistor

The other important branch of the semi-conductor tree is that of the junction transistor. Some of its properties have already been considered. Other advantages of transistors are small size, ruggedness, and instant operation on switching on. Probably their most attractive feature, however, is their high reliability. At the present time the main limitations of transistors are their inability to operate at very high frequencies, and the rather limited amount of output power they can handle. There are, however, a large number of important situations in which satisfactory performance can be obtained. At high ambient tempera-

tures the life of transistors is reduced, but the safe maximum temperature for continuous operation is being increased gradually. Although there is some variation of electrical characteristics with temperature, this can be overcome by fairly simple circuit techniques. In order to overcome the limitation on output power, considerable interest has been taken in the design of Class B push-pull amplifiers. These are more efficient than Class A amplifiers and enable larger output powers to be handled; the reason for this is that power limitation is really set by the amount of power actually dissipated in the transistor. If the present-day difficulties of frequency and temperature are considered, germanium transistors are limited in application to lower and medium frequencies in reasonably cool locations. They are already finding considerable application in hearing-aid amplifiers and broadcast receivers. It is believed that all present-day hearing-aids produced in the U.S.A. use transistors. In the field of radio they are likely to be used initially in portable receivers, and in automobile receivers, where the 12-volt accumulator may be used as a direct H.T. supply.

Another big field of application for transistors is in electronic computers, and for industrial control where advantage is gained from their reliability and circuit manœvrability. Also, their low power consumption is an advantage when a large number of devices have to be operated. The photo-effect in semi-conductors may also offer possibilities in the shape of a photo-transistor. This is a variation on the photo-diode and makes use of the change in current of a p-n diode, which is biased in the reverse direction, when light is incident on it.

The advent of the transistor has provided science with an amplifying device that has the mechanical simplicity and ruggedness of a resistor or capacitor, and without the complication of auxiliary power supplies such as are required in the magnetic amplifier and the thermionic valve. As a consequence a number of new applications may be found for it, where previously other devices were considered unsuitable on account of their bulk, expense, or inadequate reliability.

THE PREDICTION AND DISCOVERY OF THE NEUTRINO

A. W. HASLETT

AFTER retaining for 25 years the status of a hypothetical, but theoretically useful, particle, the neutrino seems at last to have been detected. The first suggestion of its existence was made by Wolfgang Pauli in 1931 in order to explain the observed facts of one form of radioactivity, that in which electrons are emitted. In the interval the neutrino has become one of the essential entities of modern physics. Yet its detection as a free particle would have remained impossible but for the development of atomic energy; even to devise and carry out an experiment which would yield evidence not dependent on wishful thinking has taken American physicists three years. Those responsible are F. Reines, C. L. Cowan, and others of the Los Alamos Scientific Laboratory, in the United States.

The interval between suggestion and confirmation can be sufficiently explained by the largely nebulous qualities with which the theoretical neutrino was endowed. Not only was it, like the neutron, supposed to have no electric charge – the property which enables the tracks of most nuclear particles to be followed – but also, and here unlike the neutron, it was supposed to have a negligible mass when at rest. No other properties were assigned to it but those needed to balance the equations written by physicists – in the first place in terms of energy, but later also in terms of spin, i.e., angular momentum.

The difficulty presented by radioactivity and referred to above was that, whereas alpha particles and beta rays, when emitted from a nucleus, are emitted with a definite energy characteristic of the emitting nucleus, beta particles (electrons) are emitted with all energies less than a certain maximum, which

maximum alone is a characteristic of the nucleus. The most direct explanation would have been that, either in the nucleus that breaks down, or in the daughter nucleus formed from it, there was a continuous range of possible energy levels, so that the balance of energy released would then also have a continuous range. This, however, would have been at variance, not only with the basic idea of quantum mechanics, but also with all other evidence of nuclear properties. Hence the introduction of the neutrino as a way out: because, if a neutrino were formed in the disintegration as well as an electron, the neutrino could be supposed to carry away whatever energy the electron did not. At the time when Pauli's suggestion was made, the neutron had not been discovered by Chadwick; and, because it had no electric charge, Pauli's hypothetical particle was described at first as a 'neutron'. Three years later, when both the neutron and positive electron had been discovered, and the invention of further new particles seemed therefore more plausible, Pauli's ideas were elaborated into a quantitative theory by Enrico Fermi; and it was Fermi, the Italian-speaking émigré, who gave 'il neutrino' its name, the 'little neutral one'.

There was one earlier piece of evidence which bore on the problem. In 1927, C. D. Ellis and W. A. Wooster, working in the Cavendish Laboratory, Cambridge, had measured the heat given out by beta particles in the predominant mode of breakdown of radium-E. They discovered that the heat given out corresponded not with the maximum but with the average energy of the beta particles. This showed that, if neutrinos were emitted, as was later suggested, they must have had so slight an interaction with matter that they had escaped from the apparatus with virtually no loss of energy – in other words it showed that they were highly penetrating particles.

During the next twenty years the theoretical status of the neutrino gained progressively in strength. Except for beta particles of the lowest energies, the neutrino theory accounted, as did no other, for most observations made of beta activity; and the more accurate the measurements, the better, on the whole, was the agreement. When mesons were discovered the neutrino had again to be invoked to account for the breakdown of these particles – in particular for the breakdown of the μ- and the π-meson, the two first kinds to be discovered. And the neutrino was needed again to explain the breakdown of the neutron. But although the neutrino could be invoked to carry away energy as required, and to conserve angular momentum, the trouble remained that it 'had no act' – quoting W. F. G. Swann – in which it could be caught and detected.

'No act' is an exaggeration, although only a small one. The postulated neutrino has a very slight probability of interaction with protons, and so with matter in general. If the neutrino is regarded for simplicity as a point, the theoretical prediction is that it would have, in round figures, one chance in 10^{16} of interaction with any proton which it might encounter. The result of the interaction, when it occurs, is the formation of a positive electron and a neutron. The positive electron will, after slowing down, encounter a negative electron, and yield gamma radiation by annihilation. The neutron, also after slowing down, can be absorbed by a suitable nucleus, in practice cadmium, and again will yield gamma radiation. There are thus two primary requirements for the detection of neutrinos: first, that there should be the largest possible source of beta activity to produce them; and, secondly, that there should be a low level of background radiation, including that due to neutrons.

The solution is to use a complete, high-power, nuclear reactor as the source of neutrinos. The fission products formed in its working provide the necessary high level of beta activity; and, since these are formed continuously and many of them are extremely short-lived, there is a gain from using the whole reactor as the source, as opposed to fuel elements taken from it. In meeting the second requirement – that of low background radiation – the very slight interaction between neutrinos and matter appears, surprisingly, as an advantage. The normal shielding of the reactor, designed to absorb neutrons and gamma-radiation, is penetrated readily by neutrinos; in fact, extra shielding is necessary.

A first experiment was carried out in 1953 by Reines and

Cowan at the Hanford plant of the United States Atomic Energy Commission. Although a high background was experienced, 'it was felt that an identification of the free neutrino had probably been made'. To get a more definite result, a new experiment was designed with further colleagues, and the equipment for it was taken to the Savannah River plant. Here, in the reactor building, there is an underground room which gives 'excellent shielding from both the reactor neutrons and gamma rays'. For shielding from cosmic radiation, the whole mass of the reactor building is available, and for good measure the detecting apparatus was enclosed in a paraffin and lead shield.

The detecting apparatus was in the form of a 5-layer sandwich. The outer and middle layers were liquid scintillation detectors in steel tanks 2 ft thick, 6 ft 3 in. long, and 4 ft 6 in. wide. Really enormous detectors were thus used. The target layers between them were 3 in. thick, and consisted of polyethylene boxes containing a solution of cadmium chloride in water. The scintillation detectors were 'viewed' by several dozen photomultiplier tubes, connected in parallel, and the outputs from these were fed through pre-amplifier units to an electronic analyser. This was in a trailer parked outside the building.

The problem of analysis was to pick out pairs of associated gamma-ray pulses – first, one from the annihilation of the positive electron, then one from the absorption of the neutron – and to do this with good enough discrimination in time and energy to distinguish the wanted signals from background effects. Many different counts under different conditions were therefore made. The two main results were clear-cut and impressive. First, the frequency of pairs of gamma-rays attributed to the interaction of neutrinos with protons was within five per cent of the frequency calculated from theory, the uncertainty of the theoretical prediction being in fact greater than that of the measurements. Secondly, the frequency of gamma-rays associated with neutrinos was twenty times greater than that of gamma-rays attributed to background effects. Errors in identification were thus unlikely to be important, but for all that a number of checks were also applied. One was to vary the power output of the reactor, to the extent of shutting it down altogether. Because of the high contribution made to total radioactivity by short-lived fission products, this should – and did – affect the flux of neutrinos. Another check was to reduce the concentration of protons in the target boxes by diluting the ordinary water in them with heavy water. This, clearly, should halve the frequency of interactions; and this was the effect found. Yet another factor which could be varied was the concentration of neutron-absorbing cadmium ions present in the water. Within the limits of concentration used, the effect should have been – and was – to alter the average time-lag between the annihilation pulse and the neutron-absorption pulse, but not the frequency of associated pairs of pulses. Finally, the original shielding already mentioned was increased by stacking the detector shield with bags of sawdust saturated with water. When this was done the frequency of pulses associated with neutrinos was not affected significantly; so, if the pulses originated in particles coming from outside the tank, these particles were highly penetrating.[1]

It is in the nature of such experiments that it is impossible to meet all possible criticisms in an entirely watertight way. A possible criterion is to assume for the sake of argument that the theory of the neutrino had been elaborated, as it had been, and its interaction with the proton predicted, but to forget that it was also a theoretically necessary particle. The evidence for the neutrino then appears as good or better than that for many recent discoveries in nuclear physics. With the theoretical usefulness of the neutrino thrown into the scales, the identification appears not only probable, but close to conclusive. The finding of contrary evidence would have been highly disturbing to physicists. The evidence found is in fact a tribute, not only to the ingenuity of the experimenters, but the insight behind what (in 1931) was one of the bolder speculations in physics. The neutrino was, and remains, a most unlikely kind of particle.

The question arises whether there may exist also a second kind of neutrino still awaiting detection. Just as the neutrinos already detected are emitted in those forms of nuclear breakdown in which negative electrons are emitted, so also, the emis-

sion of positive electrons should be accompanied by the emission of neutrinos, not necessarily of the same kind. The most likely source of the sun's energy is a nuclear cycle which begins with the interaction of two protons, and leads at this stage to the formation of positive electrons and neutrinos.[2] In spite of its great distance, the sun should still, on the earth, provide the most plentiful source of these neutrinos, and the possibility of detecting them, and of distinguishing them (if different) from 'ordinary' neutrinos, has been discussed lately by Professor O. R. Frisch in a broadcast talk.[3]

[1] Cowan, C. L., Reines, F., Harrison, F. B., Kruse, H. W., and McGuire, A. D.: *Science*, 1956, **124**, 103.
[2] See, for example, the article by C. M. Bondi and H. Bondi in *Science News* 1954 **34** 18 *et seq.*
[3] *Science Survey*, 18 Oct. 1956.

CORRESPONDENCE

Bernard Mayo's articles on 'Theoretical Entities' in volumes 32 and 39 have stimulated a number of writers to express their views. One of these letters is printed: it is hoped to print others later.

BERNARD MAYO'S article on theoretical entities (*Science News 39*) seems to perpetrate some elementary misconceptions on the subject. As an unabashed supporter of the analogy of the concepts of the chair and the electron, I should like to point out the following weaknesses in his argument.

Firstly, he speaks of a theoretical entity as something 'unobserved and unobservable', and cites the gene and the super-ego as similar concepts, logically, to that of the electron. And the unobservability of the electron 'follows from Heisenberg's principle'. What this principle actually states is: That it is impossible simultaneously to determine the position and velocity of an electron so that the uncertainties in each determination multiplied together are less in magnitude than Planck's constant, h. This principle applies to chairs as well as electrons; and it does not postulate the unobservability of either.

Likewise, the theoretical entity, the gene, is only cytologically 'unobservable' while it is part of a purely mathematical heredity theory. In the study of cytogenetics it is part of the equipment of a cell, situated upon a chromosome, and probably made of desoxyribonucleic acid. Its role in heredity theory does not preclude its physical existence.

Secondly, he wishes to speak of 'non-empirical concepts' in the theories of physics, and includes 'the electron' and 'gravity'. Surely this is the essential confusion, in which he fails to distinguish the electron of Schrödinger's equations from the electrons that comprise matter, gravity of gravitation theory from individual gravity fields, and would also confuse the conceptual chair of the carpentry text-book with the oaken one owned by Mr Jones.

Of course, gravitational fields are empirical. One can demonstrate that of the earth by releasing a heavy object in mid-air as readily as one can demonstrate an oaken chair by sitting upon it.

He claims that 'an electron has precisely those properties which are postulated of it, and no others'. Indeed it has such properties, in order to be considered an electron; so have chairs such properties, or they are not considered chairs. He then admits that electrons have two variable properties (position and velocity), whose particular values cannot be a part of the theory; and does not realize that a similar state of affairs occurs with all physical objects. Of course the variable properties of most of them number more than two; the electron is essentially the simplest.

To the objection that we 'see' an electron hitting a fluorescent screen, he replies with the phenomenological argument that 'what we see is a flash of light'. One can only assume by this that his criterion of physical observability is the reflection of light, unfortunately not a property of electrons, or gravitational fields. A similar argument is advanced against cloud-chamber tracks; 'the interpretation as the track of an α-particle depends upon a certain theory'. Assuming that Mr Mayo accepts this theory, one must suppose his explanation would be: 'This is the track of an alpha-particle (but of course alpha-particles are only parts of a theory about tracks on photographic plates).'

Finally, he contends that flour is not 'made of' atoms. A long time ago the Greeks wondered whether matter was infinitely divisible or not. Modern science has decided, I feel conclusively, that it is not. In other words, matter is made of atoms; which statement is an empirical fact. And the fact that atoms form part of Dalton's theory of chemical action, or that they are not small objects reflecting light, has nothing whatever to do with it.

P. NICHOLLS

St John's College
Cambridge
22 April 1956

SOME BOOKS RECEIVED

INTRODUCTION TO ASTRONOMY by Cecilia Payne-Gaposchkin (London: Eyre and Spottiswoode, 1956), pp. 508, 50s.

The author of this attractive volume was born and educated in England. For the past twenty-five years she has taught at Harvard, where she is Phillips Astronomer and Chairman of the Department of Astronomy.

This new 'introduction' to astronomy is no child's guide to the heavens, it is a well balanced account of current outlook valuable to anyone who considers it worth while to make himself familiar with the mid-twentieth century view of the universe. It admirably reflects its author's liberal outlook and sense of the unity of science, history, and the arts. Astronomy, it cannot be denied, plays an important part in the cultural development of the human race. Not only was this true in classical and mediaeval times, it is equally true today when man has turned his giant instruments to search the frontiers of space far beyond the confines of the solar system. This advance in outlook, and change in emphasis, is reflected in Dr Payne-Gaposchkin's book. Thus, while the first half of it keeps within the terms of reference of the title, the second part breaks through into stellar astronomy. This latter part is perhaps the more useful: it includes a general outline of the stellar universe, an introduction to astrophysics, and a detailed description of the stellar systems. For those who wish to ponder on the philosophy of the subject there is a final chapter on the origin of the universe.

The publishers claim that this is the 'best and most up-to-date book available both for the student and for the general reader'. Without seeing all the others one cannot judge of the truth of this statement, but it certainly makes a considerable number of publications in the same field out of date. The illustrations are particularly pleasing and there is an adequate index.

GUIDE TO THE HEAVENS by H. P. Wilkins (London: Frederick Muller, 1956), pp. 107, 12s 6d.

Writers of terrestrial guides may always have their works rendered obsolete by fire, tempest, flood, the Queen's enemies, or simply by changes that pass for human progress. Writers of guides to the stars are on surer ground: they look for a relative degree of permanence

in the objects they describe: and Dr Wilkins displays a touching faith in the longevity of his readers (and the usefulness of his guide) for he includes diagrams of the Eclipses of the Sun visible in Great Britain up to A.D. 2056 and the positions of the Planets up to A.D. 2000.

One hopes however that by that date there will have been a demand for a second edition which will enable some improvements in the arrangements to be made. The star maps for the first three months of the year are first given without any identification of the stars, only then is there an index map for one month. This is workable no doubt, but the tiro may have lost interest by that time. There is no index to the Southern Circumpolar map, though one might have been more useful than the views of Mars through a 6-inch telescope, equipment of the kind not possessed by the readers for whom the book is intended. Nevertheless this is an interesting short compilation, though unfortunately the price is rather high.

GUIDE TO MARS by Patrick Moore (London: Frederick Muller, 1956), pp. 124, 10s 6d.

The author of this little book, having already produced guides to the Moon and to the Planets, narrows his field to the red planet, which, at the present time, is particularly well placed for observation. Very few amateur astronomers will however have equipment that will enable them to take advantage of the present approach of Mars to the Earth. In spite of this limitation Mr Moore's guide supplies a background of information concerning the problem of Mars against which any new information coming from the great telescopes can be placed.

And for those who need a little fantasy to go with the more serious science, there is a concluding chapter on voyages (future not past, presumably) to Mars.

THE LIFE OF LUDWIG MOND by J. M. Cohen (London: Methuen, 1956), pp. 295, 22s 6d.

The avid devourers of biography will almost certainly find the technical detail in this book heavy going, while the scientific and technical historians will be impatient with the goings on at 'The Poplars' and at the 'Palazzo Zuccari'. This is not to say that it falls between two stools. The rise of the British chemical industry, in which Ludwig Mond played a leading role, is an important part of the social history of the 19th century. Unhappily it has hardly begun

to be written; witness this, the first biography of a key figure, which has only appeared half a century after his death.

The most serious omission is the neglect of the economic background of Mond's life. Ludwig Mond, a German Jew, if not penniless certainly not well off, comes to Victorian England. With a process for manufacturing soda made practicable by the Belgian Ernest Solvay he decides to enter a field in which British entrepreneurs were the pioneers and in which several fortunes had already been made. He scours the country in a pony and trap looking for a suitable site for a factory, raises capital with difficulty, and before success comes his way, passes through a very uphill period which is certainly inadequately described in this biography. Mummified in the text-books the equations of the process look deceptively simple: it is a pity we know so little of the trials and anguish that accompanied the development of the technology that is based on them.

Almost unnoticed by the reader there is a transition to affluence; a house in London, a palace in Rome, French chefs, a collection of Old Masters, and so on. All this and never a mention, so far as the reviewer could see, that it must have been achieved by selling washing soda at a fraction of a penny a pound cheaper than his established competitors. This is not to depreciate Mr Cohen's effort; the *Life of Ludwig Mond* can be read with great pleasure and interest; it is merely to call for more and fuller social, scientific, and economic histories of the period of Britain's great industrial expansion.

SCIENCE AND CIVILISATION IN CHINA by Joseph Needham assisted by Wang Ling, Volume II (Cambridge: at the University Press, 1956), pp. 696, 80s.

The first volume of Dr Needham's vast projected work was described as a 'profoundly original and disturbing contribution to human knowledge', and in volume II, pursuing his great scheme of historical synthesis, he continues to tear the scales from our excessively westernized eyes. Volume II deals with the history of scientific thought from ancient times to the middle ages and adds to the disturbance caused by its predecessor.

One can but speculate how profound will be the effect of such a pioneer bridge between distant cultures, particularly when one realizes that we can expect similar studies to come within the foreseeable future from the southern and eastern republics of the U.S.S.R.

ENGINEERING IN HISTORY by R. S. Kirby, S. Withington, A. B. Darling, and F. G. Kilgour (New York, Toronto, London, McGraw-Hill Book Co. Inc., 1956), pp. 530, 64s.

The title of this excellent volume is significant. It is not just a history of engineering: that is included, naturally, though the reader is not overburdened with an excess of detail, but the main purport of the authors is to discuss the interplay between the development of engineering and the society that has been its milieu throughout history. To quote the publishers: 'The authors of this book present the history of engineering as one of many human activities and show how various factors in the physical and cultural environment conditioned engineering advance and how, in turn, engineering has had its impact on the rest of society.' And the authors have had considerable success in achieving their objective.

Emphasis is, as might be expected, on civil, mechanical, and electrical developments, but there is just enough correlative metallurgy to satisfy those who see much of the advance of engineering as the utilization of materials (cheap steel, concrete, glass) made available by other technologists. Chemical engineering is not included although, as the supplier of materials used both for power production and for construction, it can be argued that it might with advantage have been discussed. To do so might well have made the volume too large, and with this one successfully launched the authors might well explore the apparently narrower but very profound subject of the influence of available materials on social and technological developments.

Many of the illustrations, particularly the earlier ones, can be commended, and there are useful general and special bibliographies, and an index.

Professor Kirby and his fellow authors may not explain Agincourt solely in terms of economic and technological history but they go some way towards doing so.

ABOUT OUR CONTRIBUTORS

R. R. BRAITHWAITE graduated in Chemistry at Manchester University in 1942 and later received the M.Sc. degree of that University. During the war he was commissioned in the Instructor branch of the Royal Navy and was on the staff of the Royal Naval College, Devonport. On demobilization he was elected Fellow of the Royal Institute of Chemistry and became Chief Research Chemist of Acheson Colloids Limited. His main interest, other than surface chemistry, is philosophy, and he was awarded a Fellowship of the Philosophical Society of Great Britain for his thesis on 'Retrospect'.

P. T. HASKELL took a degree in Zoology at the Imperial College of Science and Technology, and on graduation in 1951 was appointed Assistant Lecturer in Entomology in that Department. He was made Lecturer in Entomology in 1953 and gained a Ph.D. in 1955 for research on insect sounds. The same year he was invited by the Anti-Locust Research Centre in London to undertake field work in Africa connected with sound and behaviour in locust swarms. He is now in charge of physiological research at the Centre.

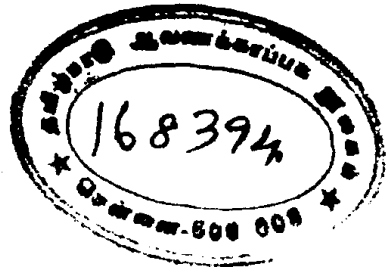
R. D. KNOTT graduated from the University of London 1948, followed by M.Sc. (Mathematics) 1951. Is a physicist working on the development of semi-conductor devices at the Research Laboratories of The General Electric Company Limited, Wembley.

J. I. MISSEN graduated from Imperial College 1949; M.Sc. (Physics) 1955. Since 1949 has been working on the application of semiconductor devices at the Research Laboratories of The General Electric Company Limited, Wembley.

JOHN D. NISBET, M.A., B.ED., PH.D., is Lecturer in Education in the University of Aberdeen. He is a Fellow of the British Psychological Society and author of *Family Environment*. He first encountered intelligence tests at the age of eleven; subsequently studied at Moray House, home of the well-known test series; and for the last eight years has done research on the topic in Aberdeen. His wife, a protozoologist, is among contributors to Penguin *New Biology*.

J. C. ROWAN was born at Old Sarum in 1925, and received an arts education, only becoming really interested in science during the war when he was on telecommunications work in the Royal Signals. His main interests are philosophy, psychology, and cybernetics, and his hobby talking about them: other interests include painting and poetry, and he has published articles, stories, and poems. Says that his dearest wish is that scientists would stop being barbarians, and that artists would stop being scientific illiterates!

A. F. TROTMAN-DICKENSON was born in 1926 and educated at Winchester and Balliol College, Oxford, where he represented the university as a long-distance runner. He has worked at the National Research Laboratories, Ottawa, Manchester University, and with E.I. du Pont de Nemours & Co. At present he is a lecturer at Edinburgh University. He has published some thirty papers on chemical kinetics and a book *Gas Kinetics*. He is now writing one on *Free Radicals*. He is married to a Polish economist and has one son.



2

(1)	(2)	(1)	(2)

SCIENCE NEWS

Outstanding articles from recent issues include :

33

Instinct and Learning: *S. A. Barnett*
High-Energy Nuclear Disintegrations: *P. E. Hodgson*
Weather, Climate, and River Erosion in the Ice Age: *G. H. Dury*

34

The Energy of the Stars: *C. M. Bondi and H. Bondi*
Uses of Mass Spectrometers: *M. P. Barnett*
Biological Individuality: *G. M. Wyburn*

35

How We Form Concepts: *W. Mays*
Meson Theory of Nuclear Forces: *G. Feldman*
Adaptations in Diving Mammals: *R. J. Harrison*

36

Growth, Regeneration, and Tumour Growth: *John O. Laws*
The Scintillation Counter: *E. S. Williams*
The Size of Atomic Nuclei: *L. R. B. Elton*

37

Fifty Years of Relativity: *W. B. Bonnor*
Nature Conservation and Ecology: *J. B. Cragg*
The Automatic Factory: *H. D. Turner*

38

Report from Geneva: *A. W. Haslett*
More Fluorine or Less?: *Joan Banus*
The Physiology of Survival at Sea: *G. R. Hervey*

39

Cosmic Ages: *P. Reasbeck*
Zone Melting: *C. H. L. Goodman*
Mnemonic Systems and Devices: *I. M. L. Hunter*

40

Soviet Embryology: *R. D. Newth*
Image Intensifiers: *L. Mandel*
Acclimatization in Insects: *K. Mellanby*

41

From Mauve to Macromolecules: *Peter Sykes*
Science and Technology of Metal Adhesives: *N. A. de Bruyne*
The Parasite's Point of View: *Geoffrey Lapage*