SCIENCE NEWS

EDITED BY A. W. HASLETT

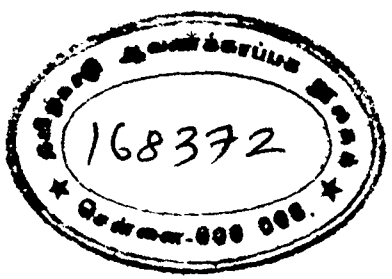
20

PENGUIN BOOKS

HARMONDSWORTH · MIDDLESEX

First published May 1951

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



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

CONTENTS

26 illustrations between pages 64 and 65

EDITORIAL

PROGRESS in science depends as much on methods and techniques as on ideas to guide their use. The classical example is the part played by the vacuum-pump in the sequence of discovery which began with 'the little indivisible particles' of Sir William Crookes, and which led through x rays and radioactivity to nuclear physics. The last few years have seemed more prolific than most of such innovations, whether wholly new or so far expanded and given precision as to be new effectively. There are examples enough in our past issues: the development of the photographic plate as a self-recording laboratory for the study of nuclear changes and particles; the use of filter-paper, with water and a second solvent, as a means for chemical separation and identification; the growing applications of the 'tracer' method, both in the biological sciences and in chemical research, with possibly a more critical attitude in interpretation; and the use of radio-methods in astronomy, from which further results are recorded elsewhere in this issue. In some of these cases, there is an obvious indebtedness to technical effort, undertaken for different purposes in the first instance; in others, the new method is so simple in essentials that, given the idea and incentive, it could have been developed at any time within some considerable period, so that it may be a problem for the future historian to explain the delay in application.

Further examples of new or recently developed methods are to be found in each of the first two articles. R. S. Nyholm writes on *The Shape of Inorganic Molecules*. Here there has been no sudden innovation, but rather the development during ten or twenty years of a variety of physical methods, both experimental and theoretical, with effects 'not yet ... fully appreciated'. In the following article, William H. Marshall discusses the *Colours of the Stars*, including the information which can be obtained from them. He ends with the advances in stellar classification which are beginning to result from the use of photoelectric methods. This also has been a slow development, but with a

sudden acceleration from new equipment. It is likely to provide some new problems for astrophysicists: for the specifications of stars, to which theoretical models must conform, will have been given greater precision. Finally, in the illustrations which follow p. 64, there is visual proof of the value of aerial photography to the geologist.

The two following articles come naturally together, and both have topical relevance. Kenneth Oakley attempts *A Definition of Man:* an exercise which has become the more necessary with the continued discovery in South Africa of fossil man-like apes, or ape-like men, of which the status remains uncertain. He is followed by Sir Richard Paget, the author of a theory of the origin of language which, after nearly a quarter of a century's interval, has found recent support from the work of Professor Jóhannesson on the roots of the Indo-European languages primarily, and more lately on a sample drawn from six 'unrelated' families of languages.

Regular readers will have learnt with regret from *Science News* 19 that J. L. Crammer has found it no longer possible to continue as editor. The character acquired by *Science News* under his editorship has been as much his creation as that of the authors who have written for him; and his combination of flair and critical judgement have been reflected in all the nineteen issues for which he has been responsible. Few journals can have been better served by their first editors; and his successor can have no better hope than to maintain the high standards which he has set and achieved.

One minor change has been foreshadowed, in part, through the inclusion of a cumulative index in *Science News* 15. With the guidance already given by the listed *Contents*, it has seemed a natural compromise to abandon the shorter indices which, so far, have accompanied individual issues, and to provide instead a fuller annual index. On the first occasion, this will cover *Science News* 16–22, so bridging the gap from the last cumulative index.

THE SHAPE OF INORGANIC MOLECULES

R. S. NYHOLM

THE last few years have seen a re-kindling of interest in inorganic chemistry, too long neglected in relation to its importance. Two factors in particular have helped to direct attention again to this subject. The more obvious is the development of atomic energy, leading to a number of new elements made for the first time in nuclear reactors and giving rise to new problems in the separation both of these new elements, and of the old, under conditions of intense radioactivity. The second has come more slowly and with effects which have not yet been fully appreciated. Over the past two or three decades a variety of new physical methods, both experimental and theoretical, have been developed which are suitable for the study of the structure of molecules, and the application of these to inorganic molecules has yielded results of special interest. In this article we are going to consider the shapes which occur in simple inorganic molecules, how they arise, and some of the modern methods which are used to determine them.

Unlike his colleague the organic chemist, whose interest is confined to the element carbon and, to a lesser extent, the elements oxygen, hydrogen, nitrogen and sulphur, the inorganic chemist considers the whole Periodic Table of some 98 elements as his domain. Some of the rules for the building up of these chemical elements, from the simpler to the more complex, are given later in this article – those of them, at any rate, which concern the chemist. The nuclei of their atoms he is content, for the most part, to leave to his colleague, the physicist. The interest of the chemist is with the number of electrons surrounding the nucleus and the way in which these are arranged; they lead to as much diversity in geometry and resulting properties as the inorganic chemist could desire.

The organic chemist is limited in the number of molecular shapes which he encounters to the tetrahedral and triangular arrangements of bonds to carbon (Figs. 1 and 2), the two V-shaped bonds to oxygen (Fig. 3) and the pyramidal bonds to nitrogen (Fig. 12). However, in inorganic chemistry the existence of at least a dozen different shapes has already been established and undoubtedly more will be discovered. The experimental research worker has been helped considerably over the past two decades by the theoretical results of the investigators in quantum mechanics. These workers, whilst still limited to purely qualitative predictions in most cases, have nevertheless established the circumstances under which the various shapes may be expected to occur and have stimulated the experimental chemists in their search for new types of compounds.

TYPES OF BONDS BETWEEN ATOMS

Broadly speaking, the chemist recognizes two main types of bonds between atoms, 'ionic' and 'covalent' bonds. In an ionic bond the two atoms are held together by electrical attraction between a negatively charged atom (a 'negative ion') and a positively charged atom (a 'positive ion'). Thus in sodium chloride (common salt), NaCl, the sodium ions (Na^+) and the chloride ions (Cl^-) are arranged in a three-dimensional cubic lattice (Fig. 4) and the cohesion of the framework depends almost entirely upon the electrostatic forces of attraction between these charged ions. The covalent bond is of a different type and results from the sharing of a pair of electrons between two atoms. Thus in methane (CH_4), for example, each of the four hydrogen atoms is held to the carbon atom by a covalent bond consisting of a pair of electrons one of which came originally from each of the atoms (Fig. 1). In this figure the origins of the electrons are distinguished by representing one by a dot (·) and the other by a cross (×). It must be emphasized that once the bond is formed the two electrons are then *indistinguishable*. One most important property of a covalent bond is the fact that it is directed in space around the central atom and this direction depends upon the particular electron being

used. The importance of this distinction between an ionic bond, owing its origin to the attraction of electrically charged atoms, and a covalent bond, resulting from the sharing of two electrons between the two bound atoms, will be appreciated more fully as discussion proceeds.

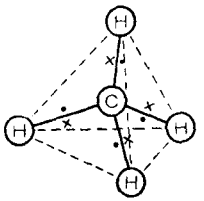


Fig. 1
Methane Molecule

C Carbon atom
H Hydrogen atom

Tetrahedral bonds to Carbon

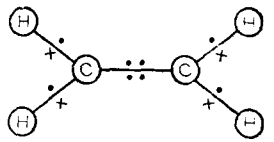


Fig. 2
Ethylene Molecule

C Carbon atom
H Hydrogen atom

All bonds to Carbon in same plane

• Electron from Carbon
× Electron from Hydrogen

} The uses of these symbols anticipates an explanation in the text

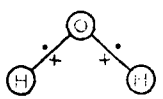


Fig. 3
Water Molecule

O Oxygen atom
H Hydrogen atom

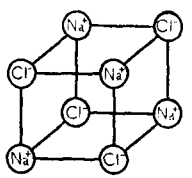


Fig. 4
Part of Sodium Chloride Lattice

Na^+ Sodium ion
Cl^- Chloride ion

The two electrons required for a covalent bond need not necessarily come initially one from each atom. Both may be

supplied by the same atom in certain cases and the bond so formed is then indistinguishable from an ordinary covalent bond. Ammonia (NH_3) is one of the molecules which can donate a pair of electrons to form such a bond, which is usually referred to as a 'co-ordinate bond' and the compounds formed in this way are called 'co-ordination compounds'. The bond is conveniently represented by an arrow pointing in the direction in which the electrons are donated, i.e., *away* from the atom from which they came; thus, $N \xrightarrow{\cdot\cdot} Cu$ represents a co-ordinate bond between nitrogen and copper, the nitrogen atom supplying both electrons. Compounds with this type of bond are shown in Figs. 11, 16 and 24. As a further variant, more than one pair of electrons may operate between the same two atoms at the same time, giving rise to a 'multiple bond' (Fig. 2). There are also occasions when the bond between two atoms may be regarded as partly ionic and partly covalent, but such cases need not concern us here.

BONDS AND THE STRUCTURE OF MATTER

Crystalline matter is built up by joining atoms together by means of ionic or covalent bonds, or sometimes both, in a way which is regularly repeated. The packing together of these small units gives rise to the crystalline form of matter with which we are familiar. The shape of these crystals is the oldest, and the most obvious, way in which crystals can be classified. However, the shape of a crystal depends both upon the shape of the molecules and the way in which they are packed together; because of the latter, the shape of a crystal gives little information concerning the shape of the molecules present. However, there is a second and much more fundamental classification of crystals which depends upon the way in which the atoms of a crystal are joined together, as was mentioned by P. R. Rowland in *Science News* 15. On this basis, crystals may be classified as molecular, adamantine, ionic and metallic – and these will now be described.

In molecular crystals covalent bonds join together atoms to form small molecules which retain their individuality in the crystal. For example, in ice (H_2O), two atoms of hydrogen and

one atom of oxygen are held together by strong covalent bonds to give a molecule of water. This uses up the main attractive forces which are available, and so the forces between one molecule and another are relatively weak. The characteristic properties of such a molecular crystal are due to the weak forces between molecules. They are: low melting point, solubility in various solvents, softness and low electrical conductivity both in solution and when molten. It is important to note that when molecular crystals dissolve in solvents the dissolved substance is present as individual molecules; the relevance of this will be apparent on page 25.

In adamantine crystals covalent bonds hold together an infinitely extended network of atoms and no definite molecule can be isolated. In fact, the whole crystal is a single giant molecule. As a result, the properties of adamantine crystals are: very high melting point, complete insolubility in solvents and great hardness. Diamond is a good example of this type of crystal.

Ionic crystals consist of a network of positive and negative ions and, once again, it is impossible to point to any definite molecule in the crystal since the crystal as a whole is a continuous framework of ions. The properties of ionic crystals are: high melting point, insolubility in organic solvents as a rule but solubility in liquids capable of decreasing the forces between the ions (e.g., water), and electrical conductivity when in solution or in the molten state.

Metallic crystals are essentially assemblages of positively charged metal ions in a cloud of free electrons. They resemble ionic crystals in some of their properties, but their most important characteristic is their high electrical conductivity. This arises from the large number of free electrons which are present.

Which of these four types of crystal is formed by any chemical compound depends upon the *types of bonds* which link the atoms together and the *orientation* of these bonds about particular atoms. We shall now see what decides this orientation in the case of inorganic, and in particular metallic, atoms which are the main subject of this article.

ORIENTATION OF BONDS IN AN IONIC COMPOUND

The orientation of bonds around a particular atom depends very much upon the type of bond. Let us consider first the shape expected when the bonds are ionic. The two important cases are those in which the central atom is surrounded by *four* or *six* other atoms. An example of the first case is the complex ion formed by the combination of a doubly positively charged cobalt ion (Co^{++}) with four negatively charged chloride ions (Cl^-). This complex ion is represented by the formula $[CoCl_4]^{--}$ The square brackets are used by chemists as a convenient way of indicating that the complex ion consists of all five atoms joined together. Now since like charges repel one another and unlike charges attract one another, it follows that the four chloride ions will try to get as *close* as they can to the positive cobalt ion but as *far away* as possible from each other. These conditions are satisfied when the four chloride ions are placed at the four corners of a tetrahedron, at the centre of which is the cobalt ion, as shown in Fig. 5. This is the shape which is obtained whenever the central atom is linked to four other atoms by essentially ionic bonds. When there are six atoms linked to the central atom by ionic bonds we get the octahedral configuration shown in Fig. 6. This is illustrated by the complex ion $[FeF_6]^{---}$, which is formed when one ferric ion (Fe^{+++}) unites with six fluoride ions ($6F^-$). In this case the two opposing

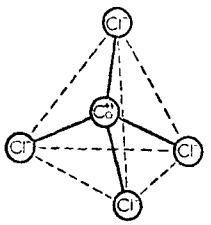


Fig. 5
Tetrahedral $[CoCl_4]^{--}$ ion
Co Cobalt ion
Cl Chloride ion

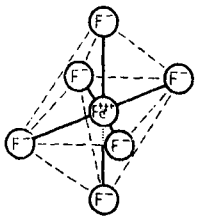


Fig. 6
Octahedral $[Fe\,F_6]^{---}$ ion
Fe Ferric ion
F Fluoride ion

tendencies – attraction between the Fe^{+++} ion and the six F^- ions, and mutual repulsion between the six F^- ions – result in the six bonds being directed towards the six corners of a regular octahedron at the centre of which is the Fe^{+++} ion. The chief factor which decides whether there will be four or six atoms surrounding the central atom is the relative sizes involved, as we might expect. Thus, small atoms like that of beryllium have only four such bonds, whereas heavy atoms like iron can have six.

ORIENTATION OF BONDS IN A COVALENT MOLECULE

In order to understand how covalent bonds determine the shape of a molecule it is necessary for us to examine more closely the structure of the atom itself and certain of the properties of the electrons used for bond formation. Fig. 7 is a purely diagrammatic representation of the electrons surrounding the nucleus of an atom. In this case a sodium atom has been chosen; this atom has a positive charge of 11 units on the central nucleus and around this are grouped 11 electrons to make the atom electrically neutral. These electrons are arranged in successive shells as we move out from the centre of the atom, there being two electrons in the first shell, eight in the second, and one in the third, in this case. These layers, or 'energy levels' as we call them, are at definite distances from the nucleus, and although under certain circumstances electrons may move from one energy

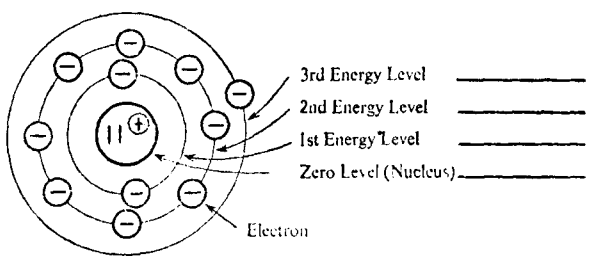


Fig. 7
Arrangement of Electrons in Sodium Atom

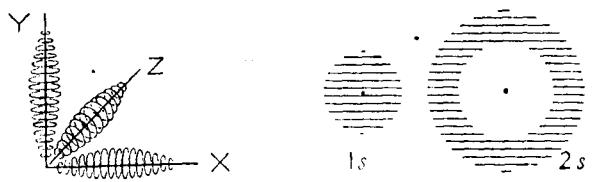
Fig. 8
Energy Levels of an Atom

level to the next, they are not permitted to stay at distances intermediate between any two successive levels. These energy levels may be compared with the floors of a building which are served by an automatic lift; we can leave the lift at the first, second, third floor, etc., but we are not permitted to get out at any position between floors. These levels are represented in Fig. 8; for an atom, the ground floor can be regarded as the nucleus itself. The energy of an electron at any higher level is equal to the work it could do if able to collapse into the nucleus. The electrons with most energy, and hence with the greatest tendency to escape from the atom, are those farthest from the nucleus, whereas the most strongly bound electrons are those closest to the nucleus.

The electrons which form covalent bonds are the ones on the outside of the atom, since they are the ones most readily accessible for combination with electrons from other atoms. They are called 'valency electrons'. We are concerned, not so much with the energy level of a valency electron, since this will be decided by the size of an atom and hence the number of electrons which it possesses, as with the *disposition* of the electron at the level which it occupies. Returning to our analogy of a lift, when we reach a given floor, the two questions which next arise are the *direction* in which we should proceed and the *room* which we shall occupy. Much the same questions arise in connection with an electron at an energy level. Atoms have available at the various energy levels a number of 'orbitals' each of which can hold a maximum of two electrons; these orbitals correspond with the rooms in our building and like them they are in definite directions from the nucleus. An orbital can be regarded as that volume of space in which the electron tends to concentrate itself or, alternatively, that volume of space in which there is the maximum probability of finding the electron. Atoms make use of four types of orbitals, but of these the directional properties of two only will concern us in this article; these are known as *s* and *p* orbitals. The other two kinds (*d* and *f*) make use of more complicated shapes.

The *p* orbitals have a definite preferred direction in space

around the nucleus. However, an *s* orbital has no preference for any one direction more than another; like a roulette board it is not directional of itself but directional from the view-point of those waiting around it – atoms in this case. An electron pair making use of an *s* orbital give rise to a covalent bond in a direction as far away as possible from any bonds using *p* orbitals which are formed to the central atom at the same time. Only one *s* orbital is available at any given energy level, but there are three *p* orbitals possible. These three *p* orbitals are arranged in three directions at right angles to one another as shown in Fig. 9. In the same figure are shown typical *s* orbitals. The shading shows the volume of space in which the electron tends to concentrate itself.



Three *p* orbitals
(one end only in each case)
Three axes at right angles
and mutually perpendicular

s orbitals
Spherically symmetrical

Fig. 9

Once we know the number of electrons which an atom contains we can decide which orbitals are used by the valency electrons. This process of atom building is for the chemist the successive filling of the orbitals available at each level, as the size of the atom is increased and successive electrons are added to it. In each *s* orbital, as already stated, there is room for two electrons. At the first energy level there is only a single *s* orbital. This is occupied by one electron in a hydrogen atom, and by two in a helium atom. The addition of any further electrons takes place at the second level. The next two go into an *s* orbital at the second level; the next six after that, into the three *p* orbitals just described. There is room, therefore, for eight

electrons at the second level, and all eight places will be filled before a start is made on the accommodation at the third level.

At higher levels, there are introduced in turn the other two kinds of orbitals, *d* and *f*, mentioned earlier, but at each level it is always the *s* and *p* orbitals, and in that order, which are filled first. By combining the rules of atom building, as just set out, with the general principles that, (i) three *p* orbitals orient themselves roughly at right angles to one another, and (ii) an *s* orbital has no preferred directional property, it is possible to understand the shape of many of the simple molecules with covalent bonds.

TWO COVALENT BONDS TO AN ATOM

A well-known molecule in which there are two covalent bonds to an atom is water (H_2O), and it is known that the oxygen atom makes use of two *p* orbitals to bind the hydrogen atoms. Since the two *p* orbitals arrange themselves at right angles, we expect the water molecule to have the angular 'V' shape shown in Fig. 3. Actually, in this molecule the angle between the two bonds is a little greater than 90° (in fact 105°), owing to the fact that the two bonds repel one another slightly. This V-shaped molecule for water has been confirmed in several ways. Other well-known molecules with this angular shape are ether, $(C_2H_5)_2O$, and hydrogen sulphide, H_2S.

A different shape to this is obtained when we make use of one *s* and one *p* orbital. Since the *s* orbital has no directional property it points *away* from the *p* orbital and we get an angle of 180° between the bonds, i.e., a straight line. When an *s* and a *p* orbital combine in this way we get a 'hybrid' of the two. This linear shape is observed in the molecule of mercuric chloride (corrosive sublimate) which has the formula $HgCl_2$. The molecule is shown in Fig. 10. Although one *s* and one *p* orbital have been used, the molecule ends up with two perfectly equivalent '*s-p* hybrid' bonds, but it is easier to arrive at the shape by considering them as separate at first. Mercuric bromide and iodide have the same shape. Copper, silver and gold make use of this

arrangement of bonds in many of their compounds. When silver chloride (AgCl) is dissolved in ammonia (NH_3) a compound with the formula $[Ag(NH_3)_2]^+Cl^-$ is obtained. As the method of writing indicates, this is a complex salt consisting of $[Ag(NH_3)_2]^+$ cations and chloride (Cl^-) anions. The bonds inside the square bracket are covalent but the positive and negative ions are held together by electrostatic (ionic) bonds. The shape of this molecule is shown in Fig. 11, the two bonds to the silver atom being in the same straight line. In this substance *both* pairs of electrons required for the two bonds to the silver atom are supplied by the ammonia molecules. However, the orbitals used by these two bonds are still one *s* and one *p* orbitals of the silver atom, and so we get a straight line molecule, apart from the hydrogen atoms of the two ammonia molecules.

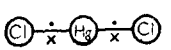


Fig. 10
Mercuric Chloride Molecule

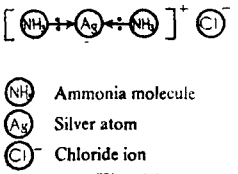


Fig. 11
Di-ammine silver chloride

THREE BONDS TO AN ATOM

Two cases are of interest. When three *p* orbitals are used by the central atom we obtain a molecule which is shaped like a pyramid with three sides. A well-known example of this type of molecule is ammonia (NH_3) in which the three hydrogen atoms form the three corners of a plane triangle above which is the nitrogen atom; this is shown in Fig. 12. As in water, partial bond repulsion results in a widening of the bond angles, in this case from 90° to 108°. Other molecules with this shape are arsenic trichloride ($AsCl_3$), phosphine (PH_3) and trimethylamine $(CH_3)_3N$. A different shape arises when the bonds to the central atom make use of two *p* orbitals and one *s* orbital. The two *p* orbitals point in two directions at right angles and when we add a third

s bond this points in a direction as far away from the other two as it can, i.e., on the other side of the atom. Actually, all of the three bonds are equivalent and they point at 120° to one another

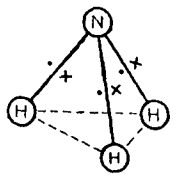


Fig. 12
Ammonia Molecule

(N) Nitrogen atom
(H) Hydrogen atom

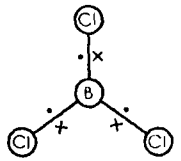


Fig. 13
Boron Trichloride Molecule

(B) Boron atom
(Cl) Chlorine atom

and all are in the same plane; this 'smoothing out' of the angles between bonds to make the molecule symmetrical is very common when we are dealing with a mixture of *s* and *p* orbitals, and the process is called 'hybridization'. An example of this kind of molecule, boron trichloride (BCl_3), is given in Fig. 13. This flat molecule with three bonds to the central atom is not very common; other examples are trimethylboron, $(CH_3)_3B$, and boron tribromide, (BBr_3).

FOUR BONDS TO AN ATOM

When there are four covalent bonds to an atom they may be arranged so as to point to the four corners of a tetrahedron or to the four corners of a square, the atom being at the centre of the figure in each case. The tetrahedral arrangement arises from the use of three *p* orbitals and one *s* orbital, there being once more a smoothing out of the angles between the bonds to make all of these equal to the so-called 'tetrahedral angle' of 109°, so that the four bonds are arranged symmetrically. A good example of this shape is the tin tetrachloride ($SnCl_4$) molecule shown in Fig. 14. Methane (CH_4), shown in Fig. 1, also owes its shape to this combination of bonds.

The square arrangement is dependent upon the use of one of the *d* orbitals to which we referred briefly earlier. It is not

easy to represent on paper the way in which *d* orbitals determine the shape of a molecule but it may be shown by calculation that the square arrangement will occur when we make use of

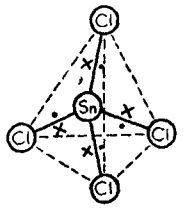


Fig. 14
Tin Tetrachloride Molecule

(Sn) Tin atom
(Cl) Chlorine atom

one *s*, two *p* and one *d* orbitals. Several of the precious metals make use of this arrangement of bonds in certain of their compounds. Thus, the compound of platinum chloride with potassium chloride with the formula K_2PtCl_4 is a square molecule of this type, as shown in Fig. 15. Another well-known substance in which this shape occurs is crystalline hydrated blue copper sulphate which has the formula $CuSO_4$, $5H_2O$. In this compound four of the water molecules are joined to the copper atom and the fifth is attached to the sulphate group. Considering only the copper atom and its four attached water

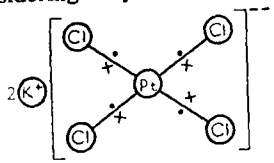


Fig. 15
Potassium platinochloride Molecule

(K+) Potassium ion
(Pt) Platinum atom
(Cl) Chlorine atom

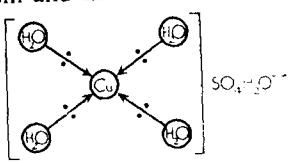


Fig. 16
Copper sulphate Molecule

(Cu) Copper atom
(H_2O) Water molecule
SO_4^{--} Sulphate group with water attached

molecules, the shape is as shown in Fig. 16. Nickel is another metal which often makes use of this square arrangement of bonds; with this metal there is a most useful general rule concerning the colour and the shape of the molecule. When the four bonds to the nickel atom point to the four corners of a square the compound is red, but when they point to the corners of a tetrahedron the compound is some shade of blue or green. Only one or two exceptions to this rule have been observed.

FIVE BONDS TO AN ATOM

Owing to the greater complexity of the orbitals involved, we shall not attempt to relate the shape of a molecule with more than four bonds to the particular orbitals employed for bond formation. When there are five bonds to an atom these are usually arranged as shown in Fig. 17, which is a molecule of phosphorus pentachloride (PCl_5) when in the form of a gas. This shape is called a 'bipyramid'; three of the bonds are in the same plane as the phosphorus atom and the other two are above and below this plane. This arrangement is the most

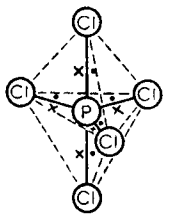


Fig. 17
Bipyramid: Phosphorus Pentachloride Molecule

(P) Phosphorus atom
(Cl) Chlorine atom

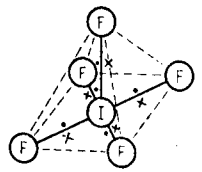


Fig. 18
Square Pyramid: Iodine Pentafluoride Molecule

(I) Iodine atom
(F) Fluorine atom

symmetrical way in which five bonds may be arranged to an atom. This usually provides us with a clue to the structure, for nature seems to have a marked tendency to make use of a

symmetrical arrangement of bonds wherever possible. Phosphorus pentachloride undergoes a very interesting change in structure when it solidifies, for it then becomes a mixture of two ions, the larger of which has the octahedral arrangement and the smaller the tetrahedral configuration. In place of (PCl_5), we have [PCl_4] [PCl_6]. Such a change in structure is not uncommon in passing from gas to the solid state. Another way in which five bonds may be arranged is shown in Fig. 18; this arrangement is known as the square pyramid. It is found in iodine pentafluoride (IF_5). There are very few cases in which there are five bonds to a metal atom, for these elements prefer four or six bonds as a rule.

SIX OR MORE BONDS TO AN ATOM

Six covalent bonds to an atom are usually arranged in a perfectly symmetrical way around the central atom and point towards the six corners of an octahedron, similar to the shape observed for a compound in which the bonds are ionic (see Fig. 6). A good example of this shape is the uranium hexafluoride (UF_6) molecule, and this is shown in Fig. 19. This compound has recently been studied rather carefully because of its importance in connexion with the preparation of the first atomic

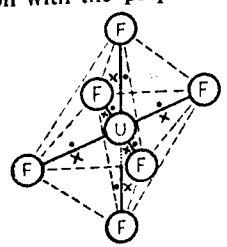


Fig. 19
Octahedral Arrangement Uranium Hexafluoride

(U) Uranium atom
(F) Fluorine atom

bomb. By allowing gaseous uranium hexafluoride to diffuse through porous tubes, uranium-235, which is explosive, was separated from uranium-238, which is relatively inactive. This octahedral arrangement of six covalent bonds is found in a very large number of metallic co-ordination compounds; although uncommon, molecules with more than six covalent bonds to the central atom are known. Examples are iodine heptafluoride (IF_7) and osmium octafluoride (OsF_8). The shape of the molecule is more complicated in each case than those described.

Table I summarizes the various shapes which we have discussed. The list is by no means exhaustive, but at least it includes all of the more common arrangements of bonds which occur in inorganic molecules.

TABLE I

Arrangement of Bonds in Inorganic Molecules

Type of Bond	Number of Bonds to Central Atom	Shape
Ionic	Four Six	Tetrahedral (5)* Octahedral (6)
Covalent	Two Three Four Five Six	Straight (10) or V-shaped (3) Planar (13) or Pyramidal (12) Square (15 and 16) or Tetrahedral (14) Square Pyramid (18) or Bipyramid (17) Octahedral (19)

* This and following numbers refer to figures in the text illustrating these shapes.

METHODS FOR STUDYING THE SHAPE OF MOLECULES

We must first distinguish between *physical* and *chemical* methods for investigating the structure of molecules. The latter involve

the definite rupture of the molecule and will not be discussed here. They are part of the chemist's stock in trade and are of special importance in organic chemistry. The use of physical methods takes the chemist into the border region between chemistry and physics; but whereas the physicist is concerned primarily with a phenomenon as such, e.g., spectroscopy, the chemist is more interested in its use as a tool for the study of chemical compounds. We might almost say that whereas the physicist uses a chemical compound in order to study spectroscopy, the chemist uses spectroscopy in order to study chemical compounds. In physical methods the molecule is unchanged in nature or composition after examination. They are those methods in which we seek to infer the structure of the molecule from its behaviour in magnetic or electric fields, its absorption spectrum, its effect upon the physical properties of a solvent when in solution, to quote but a few examples. Of the large number of physical methods now available, we shall confine our attention to some which are of special interest to the inorganic chemist.

Such physical properties as the melting point, boiling point and solubility in water or organic solvents provide the first clue to the structure of a compound. If a substance has a high melting point and will not dissolve in organic solvents it is fairly safe to assume that it is either a salt (Fig. 4) or is highly polymerized like diamond or the synthetic plastics. Polymerization is the name which we give to the process in which many small molecules unite to give a single giant molecule. If the compound will dissolve in water or in an organic solvent (without change of structure) we may then determine the size of the molecule by observing the change in freezing point or boiling point of the solvent, the change being proportional to the weight of the molecule. Ferric chloride, for example, has a molecular weight in certain solvents which corresponds with the formula Fe_2Cl_6. This indicates that each iron atom must be connected to *four* other atoms and not *three* as would be the case if the formula indicated by the molecular weight were $FeCl_3$. The shape of this molecule as it occurs in solution or in the gaseous state is shown in Fig. 20; the four bonds to the iron atoms are

approximately tetrahedral, some distortion from the tetrahedral angle taking place owing to the special requirements of the 'chlorine bridge' between the two iron atoms. Ferric chloride may be compared with boron trichloride (BCl_3), shown in Fig. 13, in which there is no such 'doubling up' or 'dimerizing'.

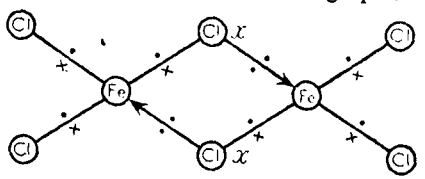


Fig. 20
Ferric Chloride Molecule in Gaseous State or in certain Organic Solvents

(Fe) Iron atom (Cl) Chlorine atom

All atoms are in the plane of the paper except the two Cl groups marked *x*; one of these is *above* the paper and the other *below*

For crystalline solids the most important method for the determination of crystal structure is unquestionably the use of *x*-ray crystallography. This method must be regarded as the final arbiter in structure determination, but it should be pointed out that the determination may be rather difficult or at the least long and tedious when the structure lacks certain symmetry. It is therefore not hard to understand why crystallographers tend to concentrate their attention upon crystals of special interest to themselves rather than to determine the shape of molecules whose structures are already known with reasonable certainty from other methods of investigation. For the use of *x* rays in the determination of the structure of matter reference should be made to *Science News* 7, 11 and 18.

When the compound is a gas, use is made of electron diffraction measurements. Owing to its dual nature, wave-system as well as particle, the electron behaves in some circumstances like *x* rays. These suffer diffraction or scattering when passed through gases, although they are less penetrating than ordinary *x* rays. From the diffraction pattern one may deduce the structure of

the molecule. This method is not so accurate as the *x*-ray method.

There are also magnetic and electrical methods which are of considerable value. The magnetic behaviour of matter has been studied in some detail since the time of Faraday in 1845, but, of course, the magnetic properties of iron and ironstone minerals have been known and made use of for centuries. When it is placed in a magnetic field, a substance is either attracted or repelled by the field. Substances which are attracted by the magnetic field are called 'paramagnetic' and those which are repelled by the field are called 'diamagnetic'. The paramagnetism of most substances is fairly small, but with certain metals, e.g., iron, it becomes very strong indeed, and the metal is then called 'ferromagnetic'; we are not concerned with ferromagnetism here. The origin of paramagnetism lies in the behaviour of the electrons surrounding the nucleus of an atom. In addition to their movement in orbits, all electrons have a spin. This spin may be clockwise or anticlockwise, and in an ordinary covalent bond one of the electrons spins in a clockwise and the other in an anticlockwise direction (*see* Fig. 21). When the number of clockwise spins equals the number of anticlockwise spins the atom is diamagnetic; but should there be an excess of one type, the atom will be paramagnetic. Furthermore, from a precise measurement of the paramagnetism we can deduce the number of these unpaired spins. Given this information it is possible in favourable cases to infer the shape of the molecule. Thus it is easy to decide whether four bonds to a nickel atom are square or tetrahedral from the magnetic behaviour. When there are four tetrahedral bonds, as in the compound of nickel sulphate with ammonia $[Ni(NH_3)_4]^{++}SO_4^{--}$, the compound is paramagnetic with two unpaired spins but when the four bonds point towards the four corners of a square, as in potassium nickel cyanide, $K_2[Ni(CN)_4]$, the two spins are 'paired off' and the molecule is diamagnetic. It must be emphasised that the magnetic behaviour only tells us whether the compound is diamagnetic or paramagnetic, and if the latter, the number of unpaired electrons; it does not give us the shape directly. However, from the number of unpaired electrons we can deduce

which orbitals are not available for bond formation and hence those which are probably used by the bonds themselves. Magnetic data must always be used in conjunction with other information concerning the shape of the molecule, and are of special value in differentiating between two alternative structures in which there are different numbers of unpaired electrons.

Fig. 21
Paired Electron spins showing Clockwise and Anti-clockwise spin

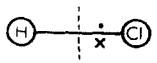


Fig. 22
Hydrogen Chloride Molecule showing electron pair closer to chlorine atom

The electrical method depends on the fact that in any covalent bond the two electrons are not usually exactly half way between the two atoms. This is because the nucleus of one of the atoms generally has a bigger positive charge than the nucleus of the other, and so attracts the electrons more strongly. This is shown diagrammatically for the molecule of hydrogen chloride (HCl) in Fig. 22. The electrons forming the covalent bond are closer to the chlorine atom. An electrical dipole also results when one of the atoms has a marked tendency to become an ion, i.e., the bond becomes partly an ionic one, but since the result is the same the origin need not concern us in detail. The molecule of hydrogen chloride is electrically unsymmetrical and it will behave as if it consisted of small positive and negative charges a short distance apart. Such an arrangement is known as an electrical dipole. When placed in an electrical field, as between the oppositely charged plates of a condenser, it tends to align itself with the field, in the same way that a small magnet aligns itself with a magnetic field. In other words it experiences an electrical twist owing to its dipole moment. The principle is illustrated in Fig. 23.

When there are more than one covalent bonds in a molecule, there will be more than one of these small electric dipoles, and when we measure the electric dipole moment for the whole

molecule, we obtain the *sum* of the individual dipoles, some of which may cancel and some reinforce one another. The most important distinction is between molecules which are perfectly symmetrical and have no dipole moment and those which are

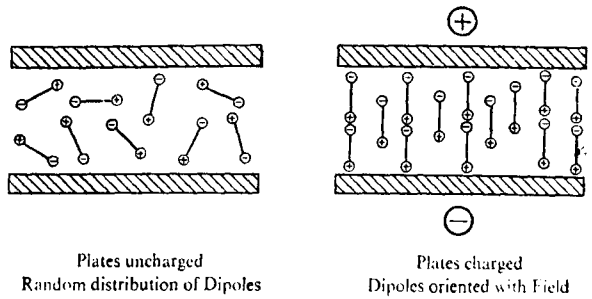


Fig. 23
Behaviour of Electrical Dipole between Plates of Condenser

unsymmetrical and therefore show a measurable dipole moment. This is often a useful way of distinguishing between possible structures, and examples of its use will be given in a later paragraph.

The size of a dipole moment is determined by measuring how the capacity of a condenser varies when the substance is placed between the plates. For the effect to be shown, the individual molecules must be free to rotate. In practice, this means that the compound must be examined either as a gas, or dissolved in a liquid which neither alters the molecules nor masks their effects. Benzene is one solvent which is often used; carbon tetrachloride is also useful, and a few others are available. The result is expressed in Debye units, so called after P. Debye, a pioneer in this field of research. The definition of this unit need not concern us for we are concerned here only with 'yes or no' answers.

When we consider the molecules of tin tetrachloride (Fig. 14) and boron trichloride (Fig. 13) we find that the total dipole moment is zero because all of the individual bond dipoles

cancel out. In the case of boron trichloride, this leaves no doubt that the molecule must be flat, with all three bonds at 120°, because no other geometric arrangement will give a dipole moment of zero. However, in the case of tin tetrachloride, it tells us only that the molecule is *either* square or tetrahedral, since the dipole moment would be zero in either case. This narrowing down of the possible structures is of great help, since other methods of elimination can also be used. For example, there is a compound of nickel which is red and diamagnetic. Formed by the union of nickel chloride and triethylphosphine, it has the formula $NiCl_2$, $2(C_2H_5)_3P$. Since there are four bonds to the nickel atom, we know that this compound may have one of three possible structures – *A*, *B* or *C* in Fig. 24. In *A* the four bonds to the nickel atom are arranged tetrahedrally, but in both *B* and *C* all four bonds are in the same plane. *B* is said to be *trans*-planar since the two chlorine atoms are directly opposite one another, but *C* is called *cis*-planar because they are adjacent to each other. It is found that the experimental dipole moment is zero, which proves that structure *B* must be correct since only in this structure do all of the bond dipole moments cancel out.

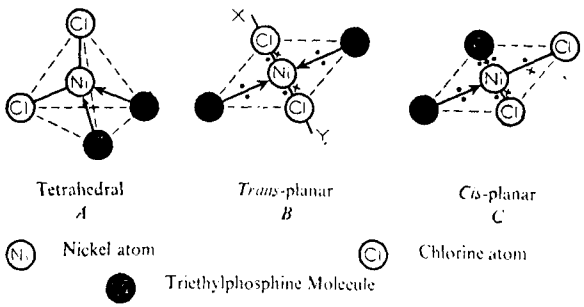


Fig. 24

This is made clear if we draw a line *XY* in *B*. The bond dipoles of the two Ni-Cl bonds cancel because they are in opposite directions. The same is true for the other two bonds in *B*. For the molecule of ammonia (NH_3), shown in Fig. 12, a value of

1.46 D. (Debye units) is obtained. This proves that the nitrogen atom cannot be in the same plane as the three hydrogen atoms, otherwise the dipole moment would be zero; hence this molecule must be pyramidal. Similarly the dipole moments of water (1.84 D.) and hydrogen sulphide (H_2S) (1.10 D.) show that these molecules must be bent (V-shaped). In the following table are given the dipole moments of a series of compounds together with their shapes. Those with a symmetrical shape are those with zero dipole moment.

TABLE II

Dipole Moments

Compound	Formula	Dipole Moment	Shape
Mercuric Chloride	$HgCl_2$	0	Straight
Mercuric Bromide	$HgBr_2$	0	Straight
Mercuric Iodide	HgI_2	0	Straight
Carbon Dioxide	CO_2	0	Straight
Water	H_2O	1.84	V shaped
Hydrogen Sulphide	H_2S	1.10	V shaped
Sulphur Dioxide	SO_2	1.76	V shaped
Ammonia	NH_3	1.48	Pyramidal
Phosphorus Trichloride	PCl_3	0.85	Pyramidal
Boron Trichloride	BCl_3	0	Planar
Tin Tetrachloride	$SnCl_4$	0	Tetrahedral
Nickel Tetracarbonyl	$Ni(CO)_4$	0	Tetrahedral

Space does not permit us to deal with spectroscopic methods for determining the structure of molecules. Studies of infra-red spectra have given much valuable information concerning the length of bonds, and the most recent development is the use of micro-wave spectroscopy. This last method gives the most accurate measurements of bond angles and bond lengths so far obtained.

In conclusion it is as well to emphasize that no single physical

method is used to solve all problems of structural chemistry and most shapes are arrived at after several different techniques have been applied. Some of these only confirm the results obtained from other methods of attack, but they are none the less valuable on that account.

In this article we have tried to cover a very wide field and at the same time have attempted to introduce some of the modern ideas of quantum mechanics in relation to molecular structure. For those who wish to go more deeply into the subjects discussed, the following short reading list will be found useful.

FURTHER READING

Pauling, Linus. *Nature of the Chemical Bond.* 2nd ed. Ithaca, N. Y., Cornell University Press (1945)

Wells, A. F. *Structural Inorganic Chemistry.* 2nd ed. Oxford, University Press (1950)

Rice, F. O., and Teller, E. *The Structure of Matter.* New York, John Wiley & Sons (1949)

Huckel, W. *Structural Chemistry of Inorganic Compounds.* Amsterdam, Elsevier (1950)

Syrkin, Y., and Dyatkina, M. *The Structure of Molecules.* London, Butterworth (1950)

[The last of these five books is more mathematical than are the four which are first listed.]

COLOURS OF THE STARS

WILLIAM H. MARSHALL

COLOUR is not one of the most marked characteristics of the stars; indeed, a casual observer would probably say that they are nearly all white, with a few red ones to vary the monotony. More careful observation, especially with a telescope, reveals a surprising variety. Table 1 shows the colours of the brightest stars in the northern sky, and it will be noticed that less than half of them are white.

Table 1: *Bright stars and their colours*

Sun	yellow	Betelgeuse	red
Capella	yellow	Rigel	bluish-white
Aldebaran	orange	Regulus	bluish-white
Arcturus	orange	Castor	white
Pollux	orange	Deneb	white
Sirius	white	Vega	white
Altair	white	Procyon	white

Now the stars are self-luminous bodies, and so, arguing by analogy from a furnace, we may expect their colours to be determined by their temperatures, so that a white star, like Sirius, should be hotter than a yellow one like the sun, which in turn should be hotter than a red one like Betelgeuse. A simple observation of colour may thus be expected to enable a star to be placed on a scale of temperatures. But the placing will be rather rough and unsatisfactory, as we might expect.

Our argument by analogy gives us no numerical values. It can only say 'hotter' or 'cooler'; it cannot say how much hotter or cooler, nor can it tell us how high on any given scale of temperature. The sun, for instance, is a yellow star, but a simple argument shows that it must be very much hotter than a yellow-hot poker. Fig. 1 shows the sun, of radius r, at the centre of a

sphere of radius R, so that the radiation leaving the surface of the sun is ultimately spread out over the whole surface of the larger sphere, where its intensity is reduced in proportion to the area over which it is spread. Thus, the radiation leaving each square centimetre of the sun is R^2/r^2 times that arriving on each square centimetre of the large sphere. For the sun and the earth, $R=$ 93,000,000 miles, and $r=432{,}000$ miles, so that $R^2/r^2=46{,}000$.

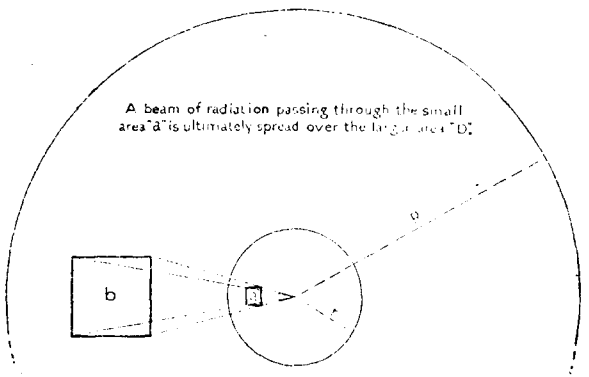


Fig. 1. The reduction in intensity of radiation, with distance from a point source, according to the area over which radiation is spread ('inverse square law').

Now the energy reaching the earth from the sun can be measured, and after correction for absorption in the earth's atmosphere it is found to be 1.94 calories per square centimetre per minute. The energy leaving the sun is therefore 46,000 times this, so that every square centimetre of the sun's surface is emitting about as much heat and light as sixty 100-watt lamps. It is obvious that the sun is much hotter than a yellow-hot poker.

The real weakness of our argument by analogy lies in the attempt to apply the idea of colour, without further investigation, to bodies so different in constitution as the poker and the sun. So the first step towards an explanation of stellar colours is to define more accurately what we mean by the colour of a star, so that it can be measured. This may be done in several different ways.

It is well known that light is a form of electromagnetic radiation, whose wavelength varies from about 4,000 to about 8,000 Ångström units (Å), 1 Å being a hundred-millionth of a centimetre. By means of the spectroscope, the different wavelengths can be separated, and the light spread out into a spectrum, and it is found that the different colours are the sensations produced by light of different wavelengths. Since the wavelength at any part of the spectrum can be measured with a high degree of precision, the spectroscope offers a means of refining our first rough estimate of colour.

The spectrum of a star consists of a continuous background, shading gradually from red through orange, yellow, green, blue, and indigo, to violet, and crossed by dark lines and bands. Some examples are shown in *Science News* 15. In passing from one star to another, we find variations in the continuous background, in the positions of the lines, and in the strength of the lines. These variations enable stellar spectra to be placed in groups of similar characteristics. The pioneer work of Sir William Huggins in Britain, Father Secchi in Italy, and Henry Draper in the United States, led finally to the elaboration of the present system of classification, in which the whole range of spectra is first divided into classes labelled O, B, A, F, G, K, M, R, N, S,* each class being further subdivided, as for example A_0, A_1, A_2, ... A_9. More than 90 per cent of stellar spectra belong to classes B, A, F, G, K, and M, those of the other classes being rarities.

This classification of stellar spectra is found to be also a classification according to colour. Stars of classes O and B are bluish-white, like Rigel and Regulus in Table 1. Those of class A are white, like Vega and Procyon. Class F stars are yellowish-white, those of class G being yellow, like the sun and Capella. Class K includes the orange-coloured stars such as Aldebaran and Arcturus; and the remainder, of classes M, R, N, and S, are red

* The peculiar order arises from changes made in the original scheme as the classification became more accurate. It may be remembered by the mnemonic 'Oh, Be A Fine Girl, Kiss Me Right Now, Sweet', which may also be taken as indicating that astronomers are not so otherworldly as some people think.

stars of various kinds. Betelgeuse, for example, is of class M. A spectrum good enough to allow of its being classified is fairly easy to photograph, and several hundred thousand stellar spectra have been classified in this way.

Another method of specifying the colour of a star depends on the fact that the intensity of its radiation is not the same at all points of its spectrum. So, when we photograph the spectrum of a star, the degree of blackening of the plate (which measures the intensity of the light) will vary from one end of the spectrum to the other. The intensity can be measured at a large number of wavelengths, care being taken to avoid the dark lines, and to measure only the continuous spectrum. In this way a curve is obtained, something like Fig. 2.

The variation in the intensity of the light, as shown by the curve, is not random, but is such that there is a quite definite maximum at a definite wavelength. In the case of the sun, for example, the maximum comes at a wavelength of about 4700 Å, which is in the yellow part of the spectrum; and we know that the sun is a yellow star.

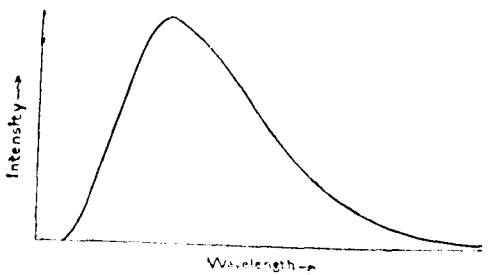


Fig. 2. Continuous spectrum of a star.

The third method of measuring stellar colours depends on the fact that different instruments for recording light differ in sensitivity at any given part of the spectrum. In the first place, we shall consider the eye and the photographic plate. The human eye is more sensitive to yellow light than to light of any other colour, as is only to be expected, since the eye evolved in an environment in which yellow light predominates. It is, however, also fairly

sensitive to both red and blue. The ordinary photographic plate, on the other hand, is highly sensitive to blue light, fairly so to green and yellow, and is practically unaffected by red. Even a panchromatic plate, which is treated so that it responds to all colours, is still much more sensitive to blue than to any other colour.

It may be well to recall here, that the brightness of a star is indicated by its position on a scale of magnitudes, such that the smaller of two numbers is attached to the brighter of two stars, magnitude 2, for example being brighter than magnitude 3. A step of one magnitude is taken to represent a light-ratio of 2.512, so that magnitude 2 is 2.512 times brighter than magnitude 3. The magnitude of a star may be measured visually, photographically, or photoelectrically, and the results are known as visual, photographic, and photoelectric magnitudes, respectively.

Now consider two stars, one red and one blue, which are such that to the eye the red one appears somewhat brighter than the blue one. If, now, we photograph them, we shall find that the blue one has so much more effect on the plate, that it appears decidedly brighter than the red one. This effect is illustrated by Fig. 3, where (*a*) shows the region of the Southern Cross as it impresses the eye, while (*b*) shows its appearance when photographed. In each diagram, the size of a dot is proportional to the brightness of the star it represents. The stars γ and δ show the effect most

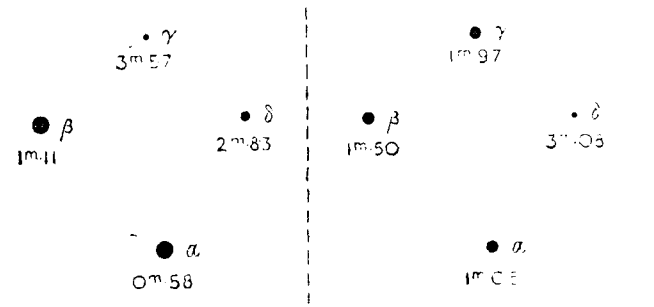


Fig. 3. Relative brightness of stars of the Southern Cross (*a*), *left*, as photographed; and (*b*), *right*, as they impress the eye.

clearly. γ is a red star of visual magnitude 1.97 and photographic magnitude 3.57, while δ is a white star of visual magnitude 3.08 and photographic magnitude 2.83. Thus γ is visually 1.11 magnitudes brighter than δ, and δ is 0.74 magnitudes brighter than γ, photographically. This gives us another way of measuring colour: we measure the visual and photographic magnitudes of a star, and the difference, photographic minus visual, is called the Colour Index. Thus, if a star is of visual magnitude (m_v) 6.2, and of photographic magnitude (m_p) 6.5, it has a colour index C.I. $=m_p-m_v=0.3$; while a star for which $m_p=8.2$ and $m_v=8.4$, has C.I. $=-0.2$.

The scale of photographic magnitudes has been chosen so that $m_p=m_v$ for stars of spectral class A_0 and between visual magnitudes 5.5. and 6.5. Thus, the colour index is small and positive for white stars, and increases for yellow, orange, and red stars; while for bluish stars it is negative and always small.

It is not now necessary to use visual estimates of magnitudes. An isochromatic plate with a suitable yellow filter in front of it has practically the same colour response as the human eye, and so it gives magnitudes which are to all intents and purposes identical with the visual ones. The magnitudes measured in this way are called photovisual magnitudes, and the colour index of a star is, of course, its photographic magnitude (measured as before from an ordinary plate) minus its photovisual magnitude.

While the photographic plate is still the fundamental instrument for measuring magnitudes and colours, it is being increasingly supplemented by the photoelectric method, which is both accurate and fast. This depends on the fact that light falling on certain substances causes the emission of electrons from their surface. These electrons constitute an electric current, which can be amplified and measured, and the strength of the current is proportional to the intensity of the light which causes it. Thus, the relative brightnesses of a set of stars can be found by focusing the light from each in turn on a photoelectric cell. Then, if one or more of the stars has already been placed on the magnitude scale, all the others can also be placed on the scale.

Until recent years the photoelectric cells available were not

sensitive enough to measure below about tenth magnitude. Even this sensitivity was achieved only by surrounding the cell with solid carbon dioxide to reduce the current due to the cell itself; and the whole apparatus was considerably less convenient than one could wish. But with the introduction of the electron-multiplier type of cell, magnitudes down to about the nineteenth can be measured quickly and accurately.

The principle of the electron multiplier may be seen from Fig. 4. A series of electrodes is arranged inside an evacuated glass envelope, and the light is admitted so that it falls on the surface of the first, causing the emission of electrons. The electrons are then focused by electric and magnetic fields on the second surface, which is specially sensitized so that it now emits several

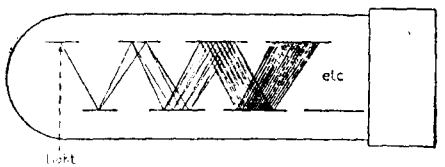


Fig. 4. Principle of the electron multiplier (diagrammatic only).

times as many electrons as fall on it. The process is repeated from one stage to another, so that after about ten stages, the original photo-current has been greatly amplified before it even leaves the tube. If the light is passed through the appropriate filters before entering the tube, we can measure both photographic and photovisual magnitudes. The tube therefore provides a rapid and accurate means of measuring colour indices.

Now, what does all this lead to? We have seen that the astronomer can measure the colour of a star with considerable accuracy, and in more than one way. But the colours are not in themselves very important; what really matters is the information we can derive from them. And the most important piece of information, as we expected from the beginning, is that we can find the temperature of a star's surface either from its colour or from its spectral class. The two are connected, as we have seen, and, since either can be used, it will be well at this point to look a little more closely at the connexion.

A star may be thought of as a sphere of hot gaseous matter at high pressure, whose surface is the photosphere, surrounded by a cooler atmosphere at lower pressure. (The transition from one to the other is, in fact, not abrupt, but the simplified picture is sufficient for present purposes.) The star's radiation is generated deep in its interior, and as it emerges from the photosphere it gives a continuous spectrum. In passing through the cooler, more tenuous atmosphere, some of the radiation at certain appropriate wavelengths is absorbed by atoms, and is reradiated *in different directions*, so that at these wavelengths the radiation is weakened, and we see dark lines against the continuous spectrum. Both the distribution of intensity in the continuous spectrum, and the nature and intensities of the absorption lines change as the temperature of the radiating surface, and hence also of the atmosphere, changes.

We shall consider first how the temperature of a star may be determined from the lines and bands on which the spectral classification is based. To understand this, we must recall some aspects of atomic structure. The typical atom consists of a massive nucleus surrounded by a 'cloud' of electrons, which are attached to it by electromagnetic forces. Now, one or more of these outer electrons can be removed from the atom, leaving the atom ionized. The more firmly an electron is attached, the more energy is needed to remove it. Thus, the outermost electron may be quite easy to remove, the next will be more difficult, the next more difficult still, and so the degree of ionization will provide an indication of how much energy has been expended on the atom. Since, in a star, the energy for ionization of atmospheric atoms comes from the radiation emerging from the photosphere, we thus have a pointer to the strength and nature of this radiation.

It is known from laboratory experiments that the particular wavelength absorbed depends on the type of atom concerned, and on its degree of ionization. From the spectrum, we can therefore tell what substances are present, and how far they are ionized. Then, since the degree of ionization is known to depend, in a way first calculated by M. N. Saha, on the temperature, a detailed

study of the lines gives an estimate of the temperature for each spectral class.

If we begin at the cool end of the spectral sequence, we find first the red stars, whose spectra show bands due to compounds such as titanium oxide. The presence of these bands shows that the temperature in these stars' atmospheres is low enough (about 3,000°) for molecules to avoid being split into their constituent atoms. As we move up the sequence, the bands disappear, and the middle classes are dominated by lines due to various metallic vapours. The lines of the neutral atoms fade as the temperature rises, those of the most easily ionized atoms disappearing first. Meantime, lines due to hydrogen become stronger and stronger, until in class A they overshadow everything else. We may note, incidentally, that the lines of hydrogen are found even in the spectra of the coolest stars, although it is difficult to ionize. This indicates the extremely great abundance of hydrogen in the stars. Finally, although temperature is the decisive factor in determining spectral class, pressure and chemical composition also exert an influence. The effect of pressure is comparatively small, but it is differences in pressure which produce the small differences between the spectra of giant and dwarf stars of the same spectra type. The effect of chemical composition is more important, and attempts to predict the spectral sequence have led to the conclusion that over 90 per cent of the stars have atmospheres of about the same chemical composition as the sun.

The temperature of the sun, estimated from its spectrum, is about 5,500° on the absolute scale.

We can get a second estimate of the temperature of a star, this time of the emitting region, from its continuous spectrum. For this method to be used, it is necessary that the light from the star shall be spread out into a fairly long spectrum. The method is therefore limited to stars of sufficient brightness for this to be done without too much loss in intensity from the spreading out.

In the study of radiation, it is found that a perfect radiator is a body which is also a perfect absorber of the radiation falling on it. Since it reflects no light at all, it is perfectly black; and so it is

customary to call a perfect radiator a 'Black Body'. Now, the radiation from a Black Body, when its intensity is plotted against wavelength, gives a curve of the same type as that in Fig. 2. Furthermore, it is found that as the temperature of the Black Body increases, the wavelength of its maximum intensity of radiation, λ_{max}, decreases according to the relation $T\lambda_{max}=0.2897$, which is known as Wien's Law. In this formula, T is the temperature on the absolute scale (reckoned from −273°C. as the zero point), and the wavelength is measured, not in Ångströms, but in centimetres. For the sun, λ_{max} is 4.70×10^{-5} cm., so that, assuming the sun to be a Black Body, we have $T=6160°$ approximately.

A third estimate of the temperature of a star is derived from its Colour Index: this gives the simplest, yet very good, basis of measurement. Using a formula derived by Planck, we can calculate the surface brightness of a Black Body at any given wavelength. Thus we can find the absolute magnitude of a star at two wavelengths, corresponding to the blue and yellow light used in measuring photographic and photovisual magnitudes respectively. These are given by the following formulae (see

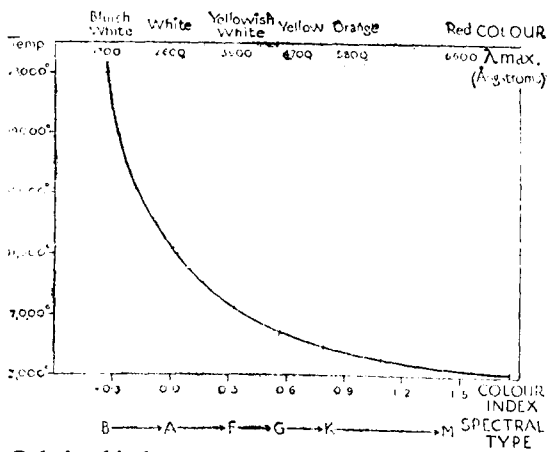


Fig. 5. Relationship between temperature (vertical scale) and colour, λ_{max}, colour index and spectral type (horizontal scale).

Astronomy by Russell, Dugan and Stewart for their derivation):

$$M_v=29{,}500/T-5\log R-0.08 \quad \ldots (1)$$

$$M_p=36{,}000/T-5\log R-0.72 \quad \ldots (2)$$

It will be noticed that these formulae involve only the temperature T and the radius R of the star; and when we subtract M_v from M_p, the terms containing R cancel out. Thus we get a formula for the colour index:

$$\text{C.I.}=M_p-M_v=7{,}200/T-0.64 \quad \ldots (3)$$

which involves only T. So, if we know the temperature of a star, we can calculate its colour index; or, what is more important, by changing the formula round, we can calculate the temperature from the colour index.

The relations between these methods of measuring stellar colours, and the corresponding temperatures are shown graphically in Fig. 5.

If we think of the stars as being bodies bearing a fairly close resemblance to one another, we may naturally expect that the hotter stars will be intrinsically brighter than the cooler ones. In other words, we may expect to find some correlation between

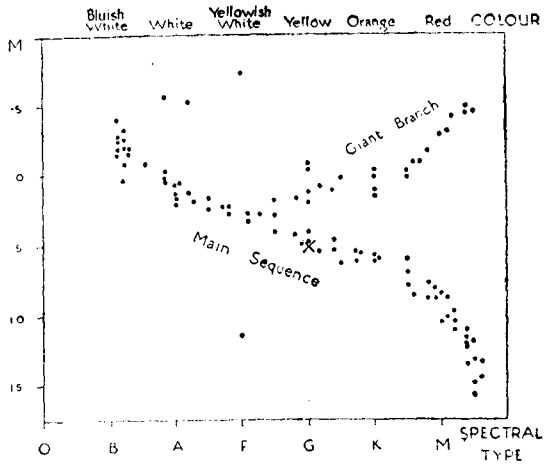


Fig. 6. Relationship between luminosity (vertical scale) and spectral type and colour (horizontal scale).

luminosity and colour or spectral class; and such a relation does indeed exist. In its usual form, as shown in Fig. 6, it is a spectrum-luminosity diagram, but since we are here chiefly concerned with colour, we shall consider it as a colour-luminosity relation, for which purpose the colours are shown for the various spectral classes.

In this diagram, known as the Hertzsprung-Russell diagram, after the two astronomers who first used this method of displaying the facts, each point represents a star. The 'x' near the middle, for example, represents the sun, of spectral type G_0 (colour, yellow) and absolute magnitude 4.8. We can see from the diagram, that a star of type A_0 may confidently be expected to have an absolute magnitude not far from 0; while one of type B_0 is likely to be of magnitude about -2.5.

This unique, though somewhat rough, correspondence of colour and luminosity holds from class B to class F; but from class G onwards, we see that the sequence is divided, and a star of class M, for instance, may be either rather faint (about magnitude 10) or very bright (about magnitude -2), with a quite distinct gap between. Thus, in the right-hand half of the diagram, a star of a given spectral type or colour may be either one of high luminosity or one of low luminosity. It was Hertzsprung who had the happy idea of calling these two kinds of star 'giants' and 'dwarfs' respectively. The 'H-R' diagram may therefore be said to consist of two branches: the main sequence runs from the blue giants diagonally down towards the red dwarfs, and the giant branch joins on to it at about class F.

It is worth noting, in passing, that the names 'giant' and 'dwarf' were not at first intended to refer to any feature of the stars except their luminosity. Later developments, however, have made it possible to measure and calculate the masses and diameters of many stars, and it has been found that the brightest stars really are much bigger and more massive than the faintest. The giants and dwarfs, therefore, are indeed giants and dwarfs in all their characteristics.

The existence of the main sequence and the giant branch suggests that two stars may be so much alike in their fundamental

structure as to give practically identical spectra, yet differ so much in some detail of their make-up that one is ten or a hundred times as big as the other. Such a difference would be expected to show in their spectra, since it must at least partly affect the outer layers, where the spectra are formed. And, in fact, slight differences of detail have been found, which provide an important method of estimating absolute magnitude. (For details, see *Science News* 15.)

Most important of all, the colour-luminosity relation implies an equally close relation between temperature and luminosity. When two such characteristics are so closely connected, we may well look for a reason in the underlying structure of the star. The relation as it is exhibited in the usual Hertzsprung-Russell diagram is not very clear-cut, but the scatter of the points in the diagram may well be a purely observational effect, masking a much more definite relation. In that case, the theoretical astrophysicist must ultimately try to construct a model of a star in which any particular temperature will entail either one definite luminosity or one of two definite alternatives, according to the pressure in the atmosphere.

This possibility is strengthened by some recent work by O. J. Eggen in the United States on the relation between colour and luminosity. Using the electron-multiplier type of cell, Eggen has measured the photographic and photovisual magnitudes, and hence the colour indices of stars in several clusters, including the Pleiades, Coma, and the Ursa Major cluster. The advantage of concentrating on a cluster is, that the dimensions of the group are small compared with its distance from the sun, and its members may therefore be taken to be all at the same distance. Once this distance has been found, the true brightness of each member of the cluster is easily calculated from its apparent brightness. This can then be related to any other characteristic, such as colour, in a diagram of the same kind as the Hertzsprung Russell one.

The significance of Eggen's work is best seen from the two diagrams reproduced here (Figs. 7 and 8). The first of these is a colour-luminosity array, showing Eggen's measures of colour

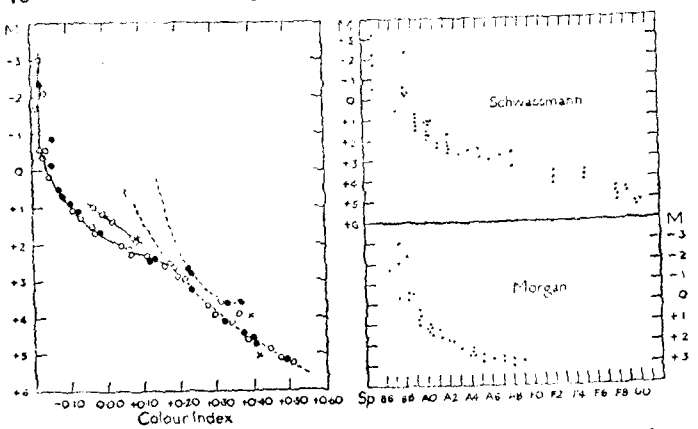


Fig. 7 (*left*). Relationship between luminosity (vertical) and colour (horizontal) as determined photoelectrically by Eggen for stars in the Pleiades cluster.

Fig. 8 (*right*). Two spectrum luminosity diagrams for the same stars. Note the 'scatter' of these diagrams, compared with Fig. 7. Both are re-drawn from Eggen's paper in the *Astrophysical Journal* (1950, **111**, 81) by permission of the author and the University of Chicago Press.

plotted against absolute magnitude for stars of the Pleiades. The second shows two spectrum-luminosity diagrams for the same stars. It is immediately clear that the relationship between colour and luminosity in Fig. 7 is much more clear and definite than the spectrum-luminosity relationship in Fig. 8. It is, in fact, so clear-cut that, in the place of one main sequence, we can distinguish several different sub-sequences; with the points for the various stars falling neatly on the appropriate lines instead of clustering round them. This is even more striking when we note that the lines themselves are those deduced from the stars of a different cluster.

We have already seen that the colour index of a star gives a good estimate of its temperature, and now it appears that it is also closely, perhaps uniquely, related to its luminosity. The extension of Eggen's work may therefore, in time, provide the astrophysicist with information which will enable him to work

out accurately the relation between temperature and luminosity. This will be a big step forward in our understanding of the stars.

Taken in conjunction with other criteria which enable the distances of the stars to be estimated, it may also give us information about the matter scattered irregularly among the stars. Starlight shining through this interstellar gas is reddened in the same way as sunlight is reddened in passing through a thick layer of the earth's atmosphere at sunrise or sunset. An estimate of its brightness, based on observation of the colour of the star and its position on a colour-luminosity array, would therefore suggest a distance actually too great; and the discovery of the discrepancy, by comparison with other estimates, would give information about the extent of the matter causing the reddening.

Now, returning to Eggen's measures on stars, it may be of interest to consider why this new work promises to be so much more accurate than any previously done. It hardly needs to be said that the difference reflects in no way on the work of all those who have contributed to the elaboration of the spectrum-luminosity diagram. The improvement is due to the difference between the classification of spectra and the measurement of colour indices. In the classification of spectra, we have six classes, each subdivided into ten sub-classes, in which to place about 99 per cent of the stars. The range of colour index corresponding to these six classes is about 2 (from −0.33 to 1.73), the colour index for any one star being measurable to 0.01. Thus the accuracy in placing by spectral class is one in sixty, while for the colour index it is one in two hundred. The colour placing is thus about three times as accurate as the spectral classification. This, together with the accuracy made possible by the new apparatus, may well account for the difference in appearance of Eggen's colour-luminosity arrays and the customary Hertzsprung-Russell diagram.

It is very likely that the measurement of star colours by the new photoelectric method will play an important part in the advancement of our knowledge of the structure of the stars, and what goes on inside them – what we may call the anatomy and physiology of the stars.

Finally, we may mention what is perhaps the most surprising

result of the measurement of star colours. From equation (1) we can clearly express R, the radius of the star, in terms of its absolute magnitude and its temperature. Then, since the temperature T can be written from equation (3) in terms of the colour index, we can ultimately express the radius of a star in terms of its colour index and its absolute magnitude. Now, when the two components of a double star are so situated that they periodically eclipse each other, study of the resulting variations in brightness enables the sizes of the stars to be calculated. Also, in the case of the largest stars, a direct measurement of the diameter may be made by means of an interferometer. In every case where these measurements are made, they agree well with the values obtained from the colour and absolute magnitude.

The measurement of colour in the stars thus enables the astronomer, not only to dip a theoretical thermometer into the stars, but also in some cases to run a theoretical tape-measure round them. And it even promises to help us to understand how the stars are made and how they work. These are indeed remarkable achievements to base on a property so apparently insignificant, and they furnish perhaps as good an illustration as any of the enormous advance made possible in any branch of science by the substitution of quantitative concepts for the purely qualitative ones with which most observation starts.

FURTHER READING

For the whole of astronomy, the reader who wants something more than a general account cannot do better than read *Astronomy* by Russell, Dugan and Stewart (Boston, Ginn and Co., 1st ed., vol. 1, 1926; vol. 2, 1927. 2nd ed., vol. 1, 1945).

For details of the theory of spectral classification, see:

Goldberg and Aller. *Atoms, Stars and Nebulae*. Philadelphia, Blakiston Co., 1943.

Johnson, M. *Astronomy of Stellar Energy and Decay*. London, Faber, 1950.

A good account of photoelectric photometry will be found in:

Hunter, A. 'Recent Advances in Astronomy'. *Sc. Progress*, 35, 75 (Jan., 1947).

Eggen's work is published in:

Astrophys. J., **111**, pp. 65, 81 and 414; **112**, 141 (1950).

ADVANCES IN ANAESTHESIA

V. KEATING

THE last twenty-five years have been characterized by remarkable advances in medicine. The science and art of anaesthesia has shared in this advance stimulated by discoveries in pharmacology and physics and by the extending field of surgery. Operations on the heart, main blood vessels and on the lungs entail serious interference with the functions of these structures, so new methods of anaesthesia have had to be devised to aid these vital organs during surgical manipulation.

Until the early twenties of this century the methods and drugs employed by the anaesthetist differed little from those described by the pioneer anaesthetists, Webb Morton and Simpson, in the mid nineteenth century, but it was not until 1923 that the subject received serious attention by pharmacologists. Until that time ether, chloroform and nitrous oxide (the 'gas' used by the dental surgeon) were the only general anaesthetics in common use. Both ether and chloroform are very useful drugs, but they have toxic effects, while nitrous oxide is too weak for use as the sole anaesthetic in serious or prolonged operations. The re-discovery by A. B. Luckhardt of the anaesthetic action of ethylene is a good example of the interdependence of apparently divergent marches of science. In 1908, carnation growers in the United States met with severe losses in sending their flowers to the Chicago markets. The blooms raised in hothouses heated by illumination gas went to sleep, i.e., the buds failed to open or the fully developed flowers wilted. Botanists of the Hull Botanical Laboratory were asked to investigate and discovered that a gas, ethylene, which formed about 4 per cent of the illumination gas was the cause. Although the anaesthetic effects of ethylene had been known since the middle of the nineteenth century and had been reported on in 1849 by Thomas Nunelly, a Leeds surgeon,

little attention was paid to them until Luckhardt's and Carter's publication in 1923 on the use of this gas in 106 surgical operations. This was an important discovery not only because it is a useful drug but because it brought the attention of experimental pharmacologists to the possibility of synthesizing a whole series of new inhalation anaesthetics. The application of this discovery was followed up by Leake who suggested the synthesis of a new chemical combining the structures of ethylene and ether. The synthesis was carried out at Princeton University by Major and Rurgh and a report published in 1931. This new anaesthetic was called di-vinyl ether and is now in common use: being a safe anaesthetic, it has largely replaced ethyl chloride for use in anaesthetic induction in children.

In the meanwhile Lucas and Henderson of Toronto were experimenting with propylene, which is in the same relationship to ethylene as is ethane to methane: it is the second member of the olefine series of hydrocarbons, ethylene being the first. An impurity of propylene, cyclopropane, was suspected to be the cause of cardiac irregularity noted by the experimenters and it was accordingly isolated for further study. Cyclopropane was found to be an excellent anaesthetic and a full report was published by Waters in 1934. This gas has been of great value; it is non-irritant to the lungs, non-toxic to body tissue, rapidly eliminated, but is nevertheless a powerful anaesthetic which can be used in low concentrations with oxygen for the seriously ill patient. It has some side actions on the heart, although widely used even in cardiac surgery because of its other good qualities. Experimental work to discover the perfect general anaesthetic still goes on. The facts detailed below will show the difficulties likely to be encountered in the search for a narcotic drug which will act on the higher centres of the brain and which will not affect the circulation, the liver and renal systems, the lungs and the lower centres in the brain which control these systems.

It is convenient to divide the drugs used in anaesthetics into two main groups; general anaesthetics which stop pain sensation from reaching the higher centres in the brain by causing unconsciousness or temporary cessation of function of these

centres, and regional anaesthetics (or analgesics) which inhibit nerve function and so prevent pain sensation being carried to the brain. The 'gas' (nitrous oxide) used by a dental surgeon will enable a tooth to be extracted painlessly by causing unconsciousness. He may also use a 'local' anaesthetic, i.e., a solution of a drug injected into the nerves supplying the decayed tooth.

While the discovery of the precise mode of action of anaesthetic drugs is held up until further advances are made in the physiology of nerve tissue, several useful working hypotheses are in existence which have been of value in estimating the probable usefulness of chemical substances as anaesthetics. The Overton-Meyer hypothesis is an example. Working independently these investigators noticed that chemicals likely to be potent general anaesthetics had a relatively higher solubility in lipoids (fat-like substances) than the less powerful anaesthetics, that is, the relative solubility of anaesthetics in oil and water roughly determines their anaesthetic potency. There are exceptions, and the parallel is not exact. Methane (marsh gas) for example has a high oil/water ratio of solubility, but is relatively ineffective as an anaesthetic agent. In general however the higher the oil/water solubility-ratio, the less the quantity required to produce anaesthesia. Meyer also showed that the concentration of all anaesthetics in the brain necessary to produce narcosis is a constant, i.e., an anaesthetic like chloroform which has a high oil/water solubility-ratio should be given in lower concentration than say ether or nitrous oxide to produce the same effect. This is in accord with clinical experience.

If an anaesthetic, say ether, is inhaled into the lungs it is absorbed by the blood stream, largely water, and distributed throughout the body. Nerve tissue is comparatively rich in lipoid content while muscle cells are comparatively low in lipoid substances, so the brain tissue will absorb more ether than will muscle. An actual estimation of the quantity of chloroform in the body shows a higher concentration in brain than in the other tissues. The Overton-Meyer theory, then, explains how certain drugs depress the action of brain tissue more than other vital tissues by

showing why a higher concentration of these drugs finds its way into the brain. It does not explain what happens when they get there. Both alcohol and motor spirit are inhalation anaesthetics; but they are not so used because the dosage required to produce general anaesthesia is too close to that which will cause respiratory arrest, i.e., the margin of safety is too low.

The Overton-Meyer theory also emphasized an important point; that while anaesthetics act mainly on the brain and nerve tissues, because they are absorbed mainly by them, they are also absorbed by and interfere with the functioning of other tissues, though perhaps to a lesser extent. For example, one hour and a half of chloroform anaesthesia causes liver damage in dogs which is detectable for eight days. While it is known that ether and chloroform cause tissue damage it is probable that all anaesthetics cause reversible injury in some degree.

Our knowledge of cell physiology has not advanced far enough to provide a useful theory of the action of anaesthetics on the individual nerve cell. Various physico-chemical changes have been noticed in all body cells subjected to the action of anaesthetics. These changes affect absorption, surface tension, hydration, particle charges and viscosity, and there are colloidal-phase effects in the living cell, but the link to correlate these phenomena is missing.

As such profound changes in cell physiology are produced by anaesthetics one can readily understand the difficulty of the clinical anaesthetist who must give sufficient narcotic to maintain the level of anaesthesia throughout a long operation while avoiding a concentration sufficient to damage the heart or the vital centres in the brain controlling the respiration and the tone of the muscles controlling the diameter of the smaller blood vessels. In a sick person and in children the margin is small.

An important quality of a safe general anaesthetic is its partially selective action inside the brain. As dosage is increased the higher centres, which determine consciousness, are first put out of action, while the 'lower' centres concerned with nervous control of the heart and respiration are put out of action later. The art of the anaesthetist, in sum, is to maintain a stable balance

between the dosage required to produce anaesthesia and good operating conditions and an overdose which would produce depression of the brain centres controlling respiration and circulation. The anaesthetist apart from this primary function must be a skilled physician able to detect the early signs of operative shock or deterioration in a patient's condition and take appropriate measures to counter them.

The giving of a general anaesthetic to-day requires the skilled administration of about half-a-dozen different drugs. The drug sequence used in a typical abdominal operation may be described as an example. The night before operation and following a clinical examination by the anaesthetist a small dose of a narcotic drug like nembutal or a similar barbiturate is given by mouth to allay anxiety and ensure a good night's sleep. About one-and-a-half hours before the patient leaves the ward for the operating theatre, an injection of morphine and atropine is given by the ward sister. The morphine allays fear, depresses the brain tissue, and reduces metabolism – so, in effect, reducing the quantity of anaesthetic required to produce a given anaesthetic level. The atropine reduces the secretion of mucus in the throat and air passages caused by irritant inhalation anaesthetics. On arrival in the anaesthetic room or occasionally in the ward the anaesthetist injects into a vein a single dose of one of the quickly acting barbiturates – usually thiopentone ('Pentothal'). This causes the patient to fall asleep but will not produce sufficient depth of anaesthesia to permit, say, an abdominal operation. A face mask is now applied and a metered quantity of nitrous oxide, cyclopropane, or ether and oxygen is administered by inhalation. The continuous administration of this mixture of gases is carried out until the end of the operation. As each person requires different quantities of anaesthetics at different stages of the operation, continuous clinical observation is required until the surgeon has finished his task. In serious and prolonged operations the physical conditions of the patient must be checked by blood pressure, pulse and respiration records and by other clinical observations. Surgical interference, haemorrhage and overdosage of anaesthetic drugs cause shock or circulatory failure;

it is part of the duty of the anaesthetist to take early action to counteract these conditions by transfusion of blood or its substitutes and by reducing the anaesthetic concentration in the blood. Rarely, in consultation with the surgeon, it may be necessary to advise modification or cessation of the operation.

It is occasionally objected that since the older anaesthetics, ether and chloroform, provide adequate operating conditions for the surgeon, the newer inhalation anaesthetics are superfluous. This is not so. Chloroform, for example, has a particularly dangerous side-action on the heart. Simpson, who introduced chloroform to medicine in 1847, had several deaths from its use, but wrongly attributed them to respiratory failure. Unlike ether, with which respiratory failure occurs long before serious heart damage, chloroform may occasionally stop the heart by direct action. It was not, however, until 1902 that a Melbourne doctor, E. H. Embly, proved this by experiment. His work was later confirmed and is now generally accepted. This toxic drug is still in restricted use.

Ether is a much safer anaesthetic – probably the safest known – because respiratory failure due to an overdose can be easily countered by an experienced anaesthetist and because sudden cardiac arrest occurring without warning is very rare. As explained above it causes marked changes in the blood chemistry and liver and kidney damage are common if the administration is prolonged. Ether is still, however, the most widely used of all general inhalation anaesthetics. Nitrous oxide, ethylene, vinyl ether, and cyclopropane are all relatively non-toxic, but for technical reasons cannot be used universally – so each patient requires a different anaesthetic 'prescription' employing one or more of these agents.

The use of these gases alone, or combined with ether and oxygen, requires special machines with accurate gas flow meters and vaporizing apparatus. Most are inhaled by the patient mixed with oxygen and then expired through non-return valves into the theatre. To reduce the danger of fire or explosion and to conserve expensive gases like cyclopropane, gas machines are constructed to allow the patient's respirations to be returned to the apparatus.

Oxygen is added, carbon dioxide chemically absorbed and the resulting mixture of volatile anaesthetic and oxygen returned for re-use. The mechanism is similar to the respiration apparatus used by firemen and deep sea divers. Accurate machinery of this type has become generally available only within the last fifteen years.

Regional anaesthesia has not in recent years changed much in method. It may be roughly classified as 'local' and 'spinal'.

The principle of the method used to obtain local analgesia is relatively simple. It depends on the fact that each organ and skin area has a fairly constant nerve supply. Many of these nerves are gathered together into main nerve trunks supplying extensive skin areas or whole organs. It is possible by injecting local anaesthetic drugs into these nerves to prevent pain sensation from the skin or deep organs from reaching the brain. Practically every known operation can be carried out under local analgesia. For various technical reasons it is not as popular as general anaesthesia in this country. It is, moreover, disliked by the majority of patients.

The spinal cord is a collection of the main nerve trunks of the body passing up the spinal column or backbone, and is bathed in fluid secreted by a structure in the brain. A spinal anaesthetic is carried out by introducing a hollow needle into this fluid and injecting a solution of novocaine or local anaesthetic into it. The local anaesthetic, of course, paralyses the nerves in the cord carrying pain sensation to the brain. The patient remains conscious. The extent of this paralysis must be carefully controlled by the dosage of the drug and the position of the patient. If the solution of the drug is of higher specific gravity than the cerebro-spinal fluid surrounding the spinal cord and the patient is sat up it will tend to anaesthetize the lower part of the body. If the patient lies in a head-down position, the solution will gravitate upwards and anaesthetize the upper part of the body. Solutions of lower specific gravity than the cerebro-spinal fluid behave in the reverse fashion.

Spinal anaesthetics (or analgesics) cause rapid loss of pain sensation – they also cause muscle paralysis in approximately the

same area as that in which sensation is lost. As operations inside the abdomen cannot be satisfactorily carried out if the muscles of the abdominal wall are fully contracted, this paralysis greatly facilitates the surgeon's work. Unfortunately a spinal block for the upper part of the abdomen will paralyse the muscles of respiration and the nerves which maintain most of the blood vessels in a state of contraction. Since the maintenance of blood pressure largely depends on the active contraction of these vessels, there is a marked fall in blood pressure which may have serious results. Artificial ventilation can be maintained during chest surgery by gas machines, as described below, but the maintenance of blood pressure presents greater difficulties. This anaesthetic method, although producing good operating conditions, has tended to be less used for upper abdominal surgery within the last decade.

A factor which has contributed to this change is the introduction of curare, a drug which is now in common use by anaesthetists, though it is not an anaesthetic. The crude drug is used by South American Indians to tip poison arrows. The purified form has a peculiar and localized action on the junction between the motor nerve and individual muscle fibres. The voluntary muscles of the body are kept in a state of 'tonus' or tension by electro-chemical impulses from the central nervous system via the motor nerves. Interference with or damage to organs covered by muscles usually causes a reflex increase in tonus. To reduce this effect during operations on the abdomen large quantities of ether, or of a similar toxic anaesthetic drug, must be given to a, sometimes, extremely ill subject. An intravenous injection of curare will greatly reduce or abolish this tonus and so enable a less powerful and less toxic anaesthetic drug to be used. Since the muscles of respiration are susceptible to the action of curare, they too can be paralysed and the anaesthetist must take over temporarily the control of respiration by mechanical means. In cardiac surgery and in some other intra-thoracic operations cessation of the movements of respiration may be required for a short period. Here curare enables the anaesthetist to vary the rate and depth of respiration at will. It is possible to do this without curare by over-ventilation of the

lungs – the effect of this being to remove from the blood the substance, carbon dioxide, which is the normal stimulus to respiration – combined with large doses of depressant narcotics. This is undesirable because of the difficulty in removing such drugs from the body when their action is no longer required and because of their side-effects on the circulation. It is interesting to note that the effect of this substance was widely known for over fifty years before it was applied to anaesthesia.

Further progress in anaesthesia is likely to be through the development of improved methods to indicate the condition of the patient's circulation. No satisfactory means has yet been devised to indicate the blood pressure continuously in a patient undergoing a long operation. An instrument has been developed by Lundy of the Mayo Clinic which gives a continuous indication of the pulse rate; but the apparatus, an electronic pulse counter, is cumbersome and expensive. A few months ago, R. G. Bickford, also working at the Mayo Clinic, succeeded in devising an apparatus by which the depth of anaesthesia is controlled by picking up the electrical wave-forms which are produced by the brain and applying these, after amplification, to work the control system. The effect is to provide a mechanical anaesthetist, so far as routine functions are concerned. If this apparatus can be simplified and developed for use in surgery, it is likely to be of value in prolonged operations in which accurate control of the depth of anaesthesia is important. Meantime the search for new inhalation anaesthetics continues, one which is now on trial in Great Britain being methyl-propyl ether.

REFERENCES

Embly, E. H. *Brit. Med. J.*, **i**, 817 (1902)
Major, F., and Rurgh, R. M. *J. Am. Chem. Soc.*, **53**, 2662 (1931)
Luckhardt, A. B., and Carter, J. B. *J. Am. Med. Ass.*, **80**, 1440 (1923)
Waters, R. M. *Anesth. and Anal.*, **13**, 56 (1934)

CONTROL OF INSECT POPULATIONS

A. G. THOMSON

DURING the past thirty years, there has been a growing realization of the importance of acquiring a better understanding of the processes which regulate the natural abundance of animal populations. Entomologists have made some outstanding contributions to knowledge of this complex subject; and the present article is concerned primarily with insect populations, and with those, in particular, which develop in stored products. It should be noted, however, that considerable attention has been devoted to the control of population in mammals and gamebirds, notably in Canada and the United States, and more lately in Great Britain.

In two papers published respectively in 1914 and 1924, K. Escherich sought to show that control of insect populations was due primarily to natural enemies, especially parasites. Escherich considered that outbreaks resulted from an unbalance between the host and its enemies brought about by the effect of weather conditions or food supply.

In 1933, Dr. A. J. Nicholson – now Chief of the Entomological Division of the Commonwealth Scientific and Industrial Research Organization, Australia – propounded a general theory of control based on the idea that populations must be in a state of balance with their environments. He held that to achieve this balance it was essential that there should be a controlling factor governed in turn by the density of the population controlled, and in his view the only factor which could be governed in this way was competition. Examination of the competition to which animals are subject led Nicholson to conclude that this generally took the form either of competition between animals when seeking the things they required for existence, or competition between their natural enemies hunting for them.

Another school of thought places the emphasis on physical control factors. Theories of this type originated from the vigorous development of experimental work showing the influence of temperature and humidity on the rates of development, reproduction, etc., of insects. Thus F. S. Bodenheimer, in papers published in 1928, 1930 and 1931, held that the population density of insects was regulated primarily by the effects of weather conditions on their development and more particularly on their survival.

Other theories are based on the conception that some populations are able, in favourable environments, to increase in density over a period of years, until a level is reached at which 'overpopulation phenomena' develop; these phenomena appear more or less suddenly and reduce the population to a low level. They then relax or disappear and the population goes through another period of increase.

In 1947, J. R. Dymond put forward a theory to explain the periodic fluctuations of certain mammals and birds in Canada. He considered that these animals were endowed with too high a reproductive capacity and that the monotony or relative uniformity of the Arctic and far northern environment accounted for the regularity with which they increased to the unstable level of numbers which brought about the periodic decline. Factors suggested as responsible for the decline included disease, emigration and starvation.

It has also been suggested that predators are the chief agents of control, but become really effective only when the carrying capacity or cover in the environment has been exceeded. For instance, N. I. Kalabukhov has pointed out that, as a population of voles or mice or ground squirrels increases, the food supply on the occupied areas will tend to diminish. When food is scarcer the animals move over larger areas. This greater movement exposes them more to their enemies, so that mortality from predators and accidents is increased. Other investigators have ascribed the periodic diminution in numbers of certain species to various diseases, or to an overflow from favourable spots into less favourable parts of the environment.

It is noteworthy that the over-population theories regard the fluctuations as characteristics of plant-eating animal populations, the fluctuations of their predators being considered as secondary. In another type of theory, fluctuations are treated as the result of interaction between the two populations. Other theories, more comprehensive again, recognize both biological and physical factors as contributory.

In a recent contribution, M. E. Solomon, of the Pest Infestation Laboratory,. the Department of Scientific and Industrial Research, has reviewed these theories and built up a picture of population control based primarily on density effects. Four phases of control are distinguished: *limitation*, which sets a variable upper limit; *conservation*, which tends to prevent extreme reduction; *suppression*, or a forced decline from a high density; and *release*, a temporary escape from normal control after a severe reduction. Each is brought about by characteristic density relationships. Suppression and release promote fluctuations.

The theoretical work from which this paper arose was undertaken as a background study to problems arising in connexion with work on the pests of stored products which is carried out at the Pest Infestation Laboratory. It is appropriate, therefore, to consider how knowledge of the natural factors limiting animal or insect populations bears on the practical aspects of pest control. I am indebted to Mr Solomon for the following points.

All economic entomologists are, of course, directly concerned with density, if only because the nuisance value of a pest increases with the population; and density effects play quite an important part in controlling the population of such pests of stored products as beetles in grain. The effects of crowding on such factors as rate of development, egg output and fertility are therefore studied. Besides the high density effects it has been noted that there is an undercrowding effect of the insects of stored products. The development of these pests is liable to be greater when there are an appreciable number of insects present than if there are only a very few. In some cases this has been ascribed to the

infrequency of encounters between the sexes, which has an adverse effect on egg-fertility. Density effects have also been studied in connexion with the relationships between hosts and parasites, where the hosts are pests of stored products, the parasite being usually a hymenopterous insect or 'wasp'.

It is also desirable to determine the physical limits of environment for the pest insects. Beyond what extremes of temperature and humidity would they be unable to complete their development, and how rapidly would they die? Within what narrower limits would they develop too slowly to be of practical importance? Such knowledge might enable control to be exercised by physical measures. And associated with this question of limits is the ability of pests to survive the British winter without the protection of heated buildings, which is essential for certain insects; and that of heat production by the insects themselves, particularly in local populations which are already large and in surroundings which are well insulated locally.

An attempt is being made to collect enough data on the life histories of insects to enable their rates of population increase, under optimum conditions, to be calculated. Theoretically some of the mites under investigation are capable of multiplying at the rate of several millionfold a year. One species has 36 generations a year and lays in effect 100 eggs per generation.

Obviously nothing remotely approaching this increase occurs in actual practice, though a very high rate of multiplication may take place during the very early stages of population growth under the most favourable conditions. Maximum performance is soon prevented, however, by various factors, chiefly density effects. It has been shown that the rate of reproduction of some grain beetles is eventually slowed down simply because, owing to overcrowding, other beetles interfere with the mating pairs by jostling them.

Possibly the most important aspect of density effects is that they are a reflection of the limited capacity of the environment. For example, in a parcel of infested grain the time comes when a shortage of food is created, particularly for insects which eat only the germ and not the floury endosperm. If the pests are

able to move outwards as they multiply, the shortage effect will be less immediate, but the insects which infest stored products may not always be able to do so quickly enough to provide sufficient food for the rapidly increasing population.

Many biologists have been struck by the remarkable disparity between the number of insects which are theoretically possible and the very tiny proportion of that number which come into existence. In examining this problem, rather an odd mistake is sometimes made. The progeny from one fertile pair of insects might theoretically be capable of reaching 100 millions within a year, but out of that vast theoretical total only perhaps a hundred insects may survive. It has therefore been concluded that the mortality must be 100 millions less 100, which is an obvious fallacy. These 100 million insects never existed except on paper, and could have existed only had there been no juvenile mortality in the earlier generations.

The mortality necessary to maintain an insect population at a steady level is not difficult to calculate. If an insect lays a hundred eggs, a 98 per cent mortality at each generation would keep the population at the existing level; whereas, with 200 eggs per generation, a 99 per cent mortality would be required. It is assumed in each case that there are equal numbers of males and females, and that the females lay the whole of their eggs within a short period. In practice, the females of some of the insects of stored products continue egg-laying even after their progeny have matured and are also laying, but this is by no means typical.

A better understanding of the complex natural factors regulating insect populations would reduce the risk of mistaken control measures. Attempts to control pests with insecticides occasionally produce unforeseen results. In many parts of the world the application of D.D.T. to fruit-trees for the control of various pests resulted in a serious outbreak of red spider mite. Investigation traced this outbreak to the destruction of the enemies which prey on the spider mites.

Cheyletus, the predator of the common grain mite, *Tyroglyphus farinae L.*, also attacks other mites including a long-haired mite

Glycyphagus, which eats minute particles of grain and does little damage. In theory, therefore, it might be advantageous to have *Glycyphagus* present in order to ensure a constant supply of food for *Cheyletus*, so that if a sudden outbreak of *Tyroglyphus* occurred there would be predators on the spot, ready to deal with it. *Cheyletus*, as it happens, is an effective predator only during the summer; but cases might arise where it was expedient to create conditions favourable for the predator.

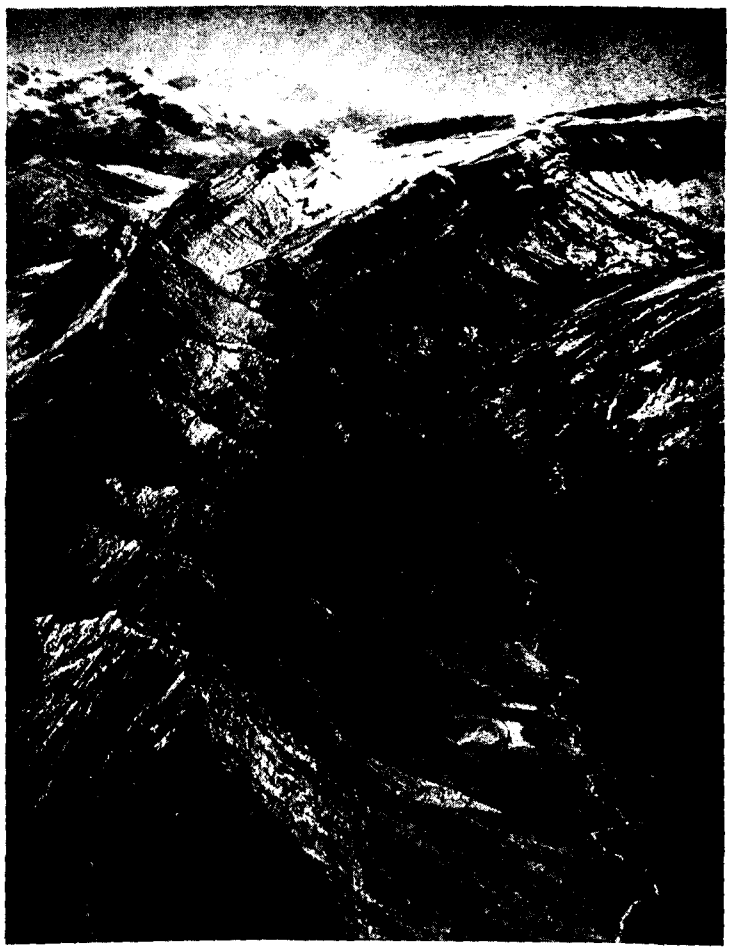
Compared with field conditions, whether in agriculture or forestry, the control of the insect pests of stored products seems to offer relatively small scope for the deliberate use of parasites and predators, since the populations to be controlled are largely temporary. As the years go by new insect poisons will doubtless be developed which can be more freely applied to foodstuffs, but the study of the biological aspects of population control may also give rise to entirely new methods of regulating the numbers of pests.

REFERENCES

Escherich, K. *Die Forstinsekten Mitteleuropas*, Vol. 1. Berlin, 1914. *Ber. Hauptversammlung Deutsch. Forstvereins*, **21**, 147 (1924).
Nicholson, A. J. *J. Anim. Ecol.*, **2**, 132 (1933).
Bodenheimer, F. S. *Biol. Zbl.*, **48**, 714 (1928). *Z. Angew. Ent.*, **16**, 433 (1930). *Arch. Zool. Ital.*, **16**, 98 (1931).
Dymond, J. R. *Trans. Roy. Soc. Can.* Secn. V, **41**, 1 (1947).
Kalabukhov, N. I. *Zool. Zh.*, **14**, 209 (1935).
Solomon, M. E. *J. Anim. Ecol.*, **18**, 1 (1949).

PHOTOGEOLOGY

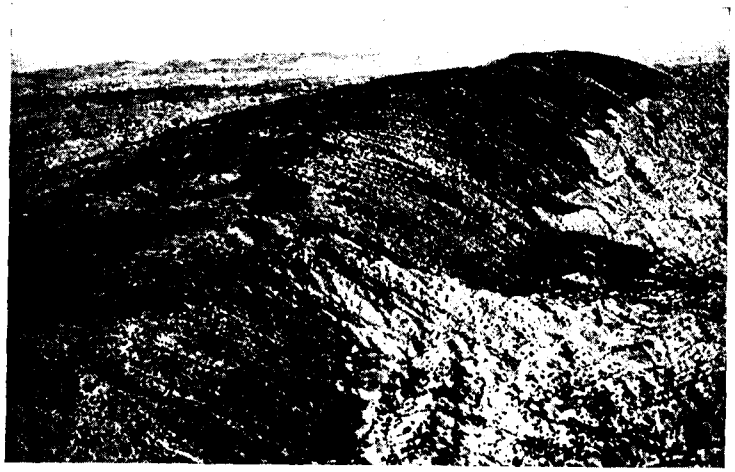
PHOTOGEOLOGY is the art and science of obtaining geological information from air photographs. Its greatest value is in country which is difficult of access and not adequately mapped. In such cases, the same air photographs can be used for mapping and rough contouring, and for geological interpretation. For geodetic and topographical surveys, the need is for completely overlapping coverage with vertical photographs taken at a uniform central scale. Overlap is required for stereoscopic viewing, which is essential in any serious use. As an indication of scale, that used in the African surveys of the Directorate of Colonial and Topographic Surveys is about 1 in 30,000, with maps from these prepared at 1 in 50,000. The obtaining of horizontal and vertical distances with sufficient accuracy for mapping is necessarily complicated: there is a distortion of scale from the centre of a photograph outwards, and superimposed on this a further distortion due to the fact that few individual photographs are strictly vertical. In mountainous country, there is the further complication that ground at a high altitude is nearer to the camera, so that the horizontal scale varies with altitude. In other words, a 'mosaic' of air photographs is not a map: rather do the air photographs provide the equivalent of survey observations from which a map can be prepared, with help from a number of control-points on the ground. On the other hand, for geological purposes the extent of detail which is needed will vary from area to area, and the quick preparation of a relatively small-scale map (possibly only 1/1,000,000) can be followed later by a more detailed examination of selected areas. Other differences are the use of oblique photographs, covering large areas, to show at a glance relationships which could only



1. Thrust block, syncline and anticline. Part of a large anticline in the Bakhtiari Mountains, S.W. Iran, with a syncline on its flank which has been partly overridden by a thrust mountain range (the S.E. end of the Zardeh Kuh). The degree of topographical dissection in the foreground is about 3,500 feet. About 7 miles of the Eocene/Upper Cretaceous valley is visible and the top of the thrust mass is about 16 miles away. (*By courtesy of the Anglo-Iranian Oil Company.*)

2. Salt plug, S.W. Iran. An oblique view of the S.W. side of Kuh-i-Namak showing the contact between the salt mass and the surrounding limestone. Note the flow lines in the salt mass. The top of the salt plug is about 4,000 feet above the foot of the mountain. The nature of the topography suggests that salt intrusion is still taking place. (*By courtesy of the Anglo-Iranian Oil Company.*)

3. An amphitheatre gorge in limestone anticline, S.W. Iran. An erosional 'amphitheatre' caused by a river cutting through an anticline of hard Lower Miocene/Oligocene (Asmari) limestone to expose underlying soft Oligocene shales. In the background a number of other structures can be identified. Width of amphitheatre, about 15,000 ft. (*By courtesy of the Anglo-Iranian Oil Company.*)

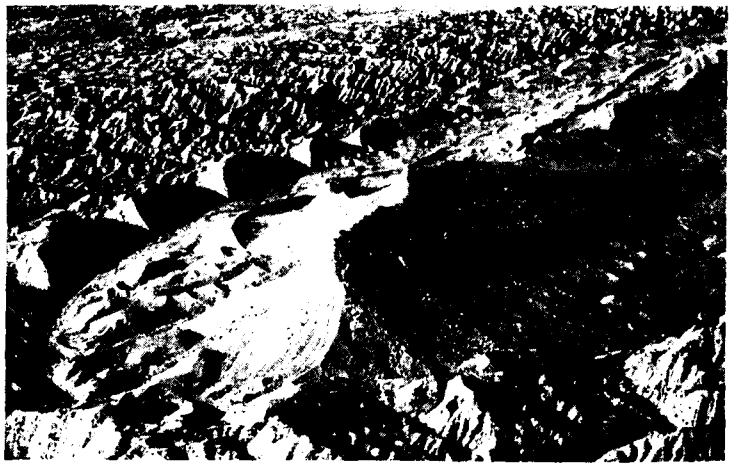


4. Anticline. An impressive view of Kuh-i-Dera, S.W. Iran, a long anticline of Asmari limestone, protruding from its cover of Lower Fars. The dark spots are oak trees. The length of the anticline which can be seen is about 12 miles. (*By courtesy of the Anglo-Iranian Oil Company.*)

5. Vertical photograph of elongated dome, Poxwell, Dorsetshire. Former quarrying has contributed to its present clearly marked dissection. It is perhaps the finest small-scale example in Great Britain of this type of structure. (*Air Ministry photograph, Crown copyright.*)

6. Vertical photograph of anticline, in Assam, seen through tropical rain forest with clearings. The lightest patches are clouds. The 'tramlines' at either side show beds dipping steeply on the flanks of the anticline. The downwards-looking 'nose' in the centre is due to a gentle dip towards the bottom of the picture along the crest of the anticline. Miocene sandstones and clays. Scale about 1/80,000. (*By courtesy of the Burmah Oil Company.*)

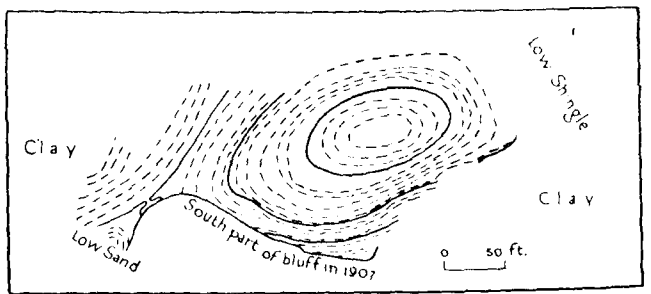
7. Vertical photograph and interpretation of folded Tertiary sediments in a desert climate (Sind, W. Pakistan). The 'nose' near the bottom-right corner is a similar feature to that in Fig. 6 above. But because of differential erosion under desert conditions, the topography is here controlled almost completely by geological structure. Note also the fault, shown as a heavy line on the diagram. Scale about 1/72,000. (*By courtesy of the Burmah Oil Company.*)



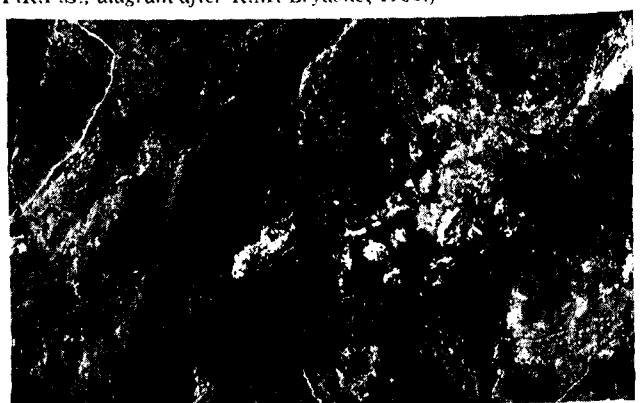
8. Oblique photograph of syncline showing convergence (S.W. Iran). Note how the syncline looks as if it was floating like a canoe in a choppy sea of Lower Fars evaporite series. Scale about 1.25 inches to one mile. For further information, see Inset 10, opposite, which is a vertical photograph of part of the same structure. (*By courtesy of the Anglo-Iranian Oil Company.*)

9. Oval limestone block, S.W. Iran. The length of the block is half-a-mile, and it is surrounded by cliffs which are some hundreds of feet high. The block, known as Qaleh Churu, is a relic of a once extensive limestone sheet, and is underlain by marls. Drainage markings show that the slope of block is downwards towards the River Rudi-zuhra, seen at bottom right. Scale about 1.8 inches to one mile. (*By courtesy of the Anglo-Iranian Oil Company.*)

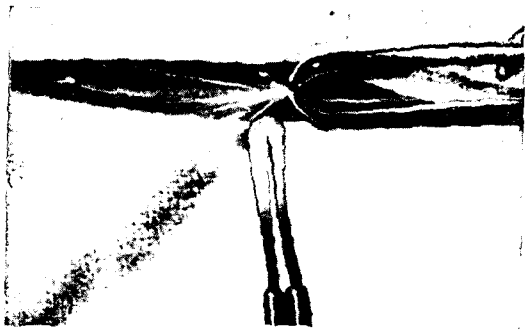
10. Vertical photograph of part of the syncline shown in Inset 8, opposite. The decrease in the thickness of the limestones in the centre of the syncline shows that gentle folding movements must have taken place during deposition. The topography shows clearly that at a geologically recent date the Lower Fars country was covered by a plain. Scale about 1.75 inches to one mile. Comparison of these two photographs provides a link between vertical and oblique photography. (*By courtesy of the Anglo-Iranian Oil Company.*)



11. Oblique photograph and diagram of eroded dome-shaped structure of *Ostrea lunata* chalk, exposed on the foreshore at Trimingham, Norfolk. The photograph was taken from the cliff-top in 1939, following a north-westerly gale which had removed the usual sand cover. It confirms in extraordinary detail the interpretation shown in the diagram, reached by more laborious methods some 31 years earlier. (*Photograph by Hallam Ashley, F.R.P.S.; diagram after R.M. Brydone*, 1908.)



12. Carboniferous strata, S. Wales. Devonian, through carboniferous limestone to millstone grit, with bands of harder limestone outcropping. (*Air Ministry photograph, Crown copyright.*)



13. The shaved and skin-sterilised animal is pressed against a cellophane barrier, above which is the sterile upper chamber. Cellophane and skin are picked up, and a small opening is burnt through.

14. The exposed uterus is laid out on sponges.

(*The photographs on this and the following pages are reproduced by courtesy of Professor J. A. Reyniers, University of Notre Dame, Indiana, and the Editor of* Crookes Digest)

15. The young are exposed and the umbilical cord is clamped off near the animal.

16. The young are removed to a clean surface and are stripped of the amnion. This stimulates as well as cleaning.

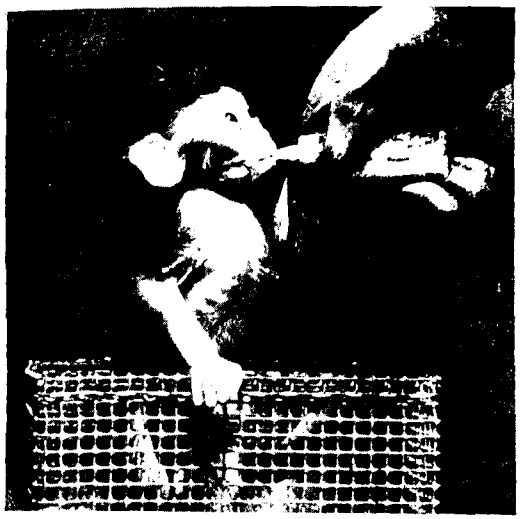
17. The young are placed in a pan after examination and passed into a rearing unit.

18. Hand-feeding of new-born rats. The milk is removed from sealed ampoules and vitamins are added with syringe and needle.

19. The feeding pipette is filled. There is a rubber nipple attached to it.

20. With the young animal held by the head, the rubber teat is pushed gently into its mouth. Its body is supported by the edge of the jar.

21. A small pledget of cotton is used to stimulate defaecation and urination.



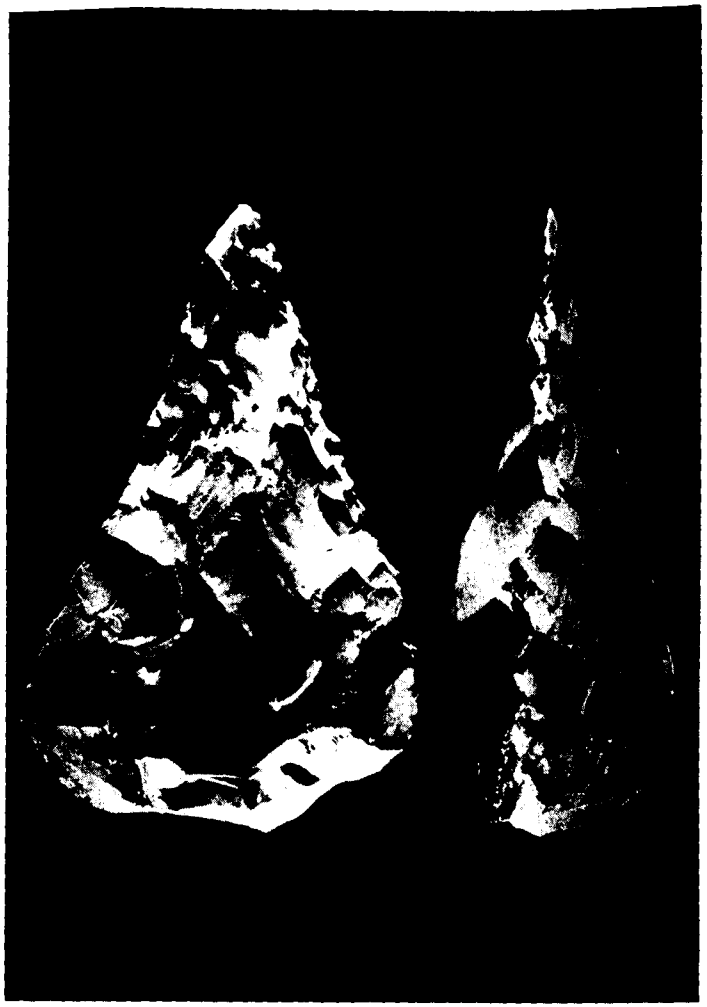
22. The stage of success. A germ-free monkey, born by Caesarian operation, is being hand-fed in a germ-free rearing unit.

23. Ape Improvising a Tool. The chimpanzee 'Sultan' is seen joining together two bamboo rods to extend his reach. After Köhler. (*By courtesy of Kegan Paul.*)

24. Skull of *Australopithecus* (restored by Professor Raymond Dart.) About one-third natural size.

25. For functional contrast: skull of South African baboon, male, on the same scale.

(*By courtesy of British Museum, Nat. Hist.*)



26. Designed for Cutting. Flint 'hand-axe', standardized tool of the Acheulian hunters. From gravels with skull of *Homo* cf. *sapiens* at Swanscombe, Kent. Two-thirds natural size. (*By courtesy of British Museum, Nat. Hist.*)

be discovered from many 'verticals', and to record rock outcrops which can be best presented in that way.

The first recorded use of air photographs for geological purposes seems to have been by Colonel A. H. Brooks, chief geologist to the United States Expeditionary Forces in France during the First World War. In a paper published in 1920, he referred briefly to the preparation of geological maps of enemy-occupied territory on a scale larger than the French official survey which provided the basis, and stated that for this purpose 'some use was made of air photographs'. According to a later account by W. T. Lee, he first familiarized himself with the appearance in air photographs of areas which he could examine on the ground, and by comparison was able to recognize similar rock formations and structures in areas held by the enemy. The point is important: for geological interpretation is essentially comparative, and the identification of the beds seen or suggested in air photographs must depend in the end on examination on the ground.

From 1924 onwards, the Canadian Geological Survey made use of air transport for observational reconnaissance; and from about 1926 onwards, air photography became of increasing importance in the Californian oilfields. A year later, a mining company surveyed an area of 10,000 square miles near the Rhodesia-Congo border in the hope – not realized in the then state of interpretation – that reasonably direct indications of ore bodies would thus be obtained. Since then, the most extended application of air survey methods to geology has been by oil companies, notably in Iran and New Guinea.

There has been a widening of interest during recent years, partly from knowledge of this work, and partly from general progress in air survey methods, particularly during the war period. As examples, a paper by A. A. Fitch and others was included in the programme of the Fourth Empire Mining and Metallurgical Congress, held in Great Britain in 1949; an exhibition of photogeology has been held lately at the Museum of Practical Geology, London; and the Directorate of Colonial Geological Surveys has established a Photogeology Section at

Bushy Park in co-operation with the Directorate of Colonial Topographic Surveys already mentioned. The object of the section is to use for geological interpretation the air photographs which have been taken, and are still being taken, for topographical surveys. This, it is stated, 'should considerably expedite' the preparation of a basic geological map of the Colonial Empire work on which is only now proceeding.

The air photographs reproduced as Inset, 1-12, have been chosen with two main objects, the inclusion of simple features which can be readily recognized, and to illustrate the variety of conditions in which the method can be applied. Two of the simplest structural features are the anticlines and synclines which result from folding. If successive beds are pictured as so many individual sheets of corrugated iron, then the effect of folding on a number of superimposed beds, initially flat, can be likened to the formation of a stack of such sheets. The comparison is not exact, since folding may be accompanied by slipping, particularly in the less coherent beds. On this analogy, an anticline is a ridge in the corrugated stack, and a syncline is a hollow. The dip towards the flanks of an anticline is the angle of downwards slope, measured crosswise to the corrugations. Inset 4 shows an anticline in south-west Iran protruding through a surface cover which is otherwise complete. Inset 5 shows an extended dome, which can be thought of as an anticline which dips at the ends as well as at the sides. It is included also as an example of the use of photogeology in the relatively familiar terrain of southern England. An anticline formed by the folding of earlier beds is likely to have suffered erosion, and the steeply dipping beds at its two flanks may then give the 'tramline' effects seen in Inset 6. This is a photograph of tropical rain forest country in Assam, and shows, incidentally, the extent to which structure may appear through a forest cover. The same photograph illustrates also the effect of erosion on the appearance of an anticline which dips ('pitches' in correct description) along the line of the fold. The effect is as would be produced by placing a stack of corrugated sheets on a surface with a downward slope in the direction of the corrugations, and then cutting a slice from the top of the

stack so as to produce a level surface. The appearance is that of a 'nose', pointed down the direction of slope (in Inset 6 towards the bottom of the picture). Inset 7 shows a similar structure under desert conditions, but crossed by a fault which runs from top to bottom of the picture, as indicated in the geological interpretation beside the photograph. The circular patterns in Inset 11 are also the result of erosion, but of erosion acting on a dome instead of a ridge. The photograph was taken from a cliff-top, not an aircraft, but is included on account of the accurate verification which it provided of a structure worked out many years earlier by more laborious methods.

The opposite case of a syncline is shown in Inset 8 and 10. These two pictures serve also to provide a link between the oblique and vertical methods of air photography, when applied to the same structure. The 'vertical' is one of a stereo-pair, and would be so used in interpretation.

A different and rather spectacular type of erosion is shown in Inset 3. Here the original structure was an anticline with an outer surface of hard limestone. This has been cut through by a river, exposing softer shales beneath, with the result that an amphitheatre has been formed of some 15,000 ft diameter. The narrowness of the gorge which is seen in front of the picture may be noted, compared with the width of excavation in relatively soft material. Inset 1, also rather spectacular, gives a vivid impression of thrust movement in a direction from top-left to bottom-right across the picture. As an indication of scale, it may be noted that the top of the thrust mass is some 16 miles distant. Each of these last two pictures is from Iran, as also is Inset 2. This shows a mammoth 'salt plug', high in the mountain mass of Kuh-i-Namak, with a 4,000 ft drop from the top of the 'plug' to the foot of the mountain.

Like other methods, photogeology has its limitations. One of them is illustrated in Inset 8, although this is a good photograph. The beds seen in the background, and described in the caption as looking like a 'choppy sea', would be taken from the photograph to be all of similar material. But on the ground, they are found to consist of beds of gypsums and marls which can be

between man and the apes of to-day indicates that they have evolved from a common ancestral stock. This conclusion is supported by the findings of palaeontologists, who have shown that some of the fossil types are, in certain features, structurally intermediate.

The common stock probably comprised a number of generalized monkey-like apes similar to those of the genus *Proconsul* which inhabited East Africa during early Miocene times, some 30 million years ago. These, it seems, were equally capable of climbing trees and of running along the ground on all fours, thus being adaptable enough to spread widely and to evolve along divergent lines of specialization. One line (or group of lines) led to the forest-bound apes of to-day, which have long powerful arms adapted for swinging from bough to bough like trapeze artists (brachiation); another line led to man, with hind limbs specialized for walking upright on open ground. This divergence of the brachiating apes and the immediate ancestors of man probably took place before the end of the Miocene period, around 15 million years ago.

Our problem of defining 'man' is resolving itself into two questions. First, what were the hall-marks of the group of 'apes' which diverged from the forest-dwellers and became the early Hominidae? Second, to what stage in the evolution of these Hominidae should the status Man be awarded – from the point of view of logical practice?

Man differs from modern apes in a number of striking physical characteristics: in the plan and form of the teeth, in habitually walking erect on two legs, and in having a brain which is considerably larger. The brain increased in size during the evolution of all the Hominoidea. Although man outstripped the forest-bound apes in this respect, it is obvious that there would have been an overlap between the range of brain-size in the earliest Hominidae and in their cousin apes. Brain-size is thus unsuitable as a criterion of relationship.

One would expect to find in 'proto-hominids' evidence that their skeleton was being adapted to allow them to leave the forests, and habitually to walk on two feet. However, adaptive

characters are usually regarded by biologists as unreliable evidence of relationship. For instance, it might well have happened that more than one line of 'apes' acquired the habit of upright bipedal gait, only one of them leading to 'man'. But it is arguable that even if that were in fact the case, all those lineages could legitimately be grouped as Hominidae. After all, the orang-utan and the gorilla represent separate, parallel lines of descent, but many zoologists do not hesitate to place them in the same family. Even so, one would naturally seek some confirmatory evidence that any fossil biped which came to light was more closely related to *Homo* than to any of the brachiating apes.

Students of fossil mammals have found that the characters of the teeth, especially when the characters are of a non-adaptive kind, are the most useful clues to relationship. Thus, arguing along general lines, one would be inclined to classify fossil ape-like forms as members of the Hominidae, irrespective of brain-size, if the evidence indicated that they walked upright and had teeth more closely resembling those of man than those of any of the existing apes. The latest discoveries in South Africa have shown that the australopithecines, although mostly with brains no larger than those of gorillas, walked upright (this is deduced from their hip bones) and moreover had a dentition of essentially human form. Their canines or eye-teeth were level with the other teeth, even from the earliest stages of wear, and even in males (Inset, 24). Although these fossil creatures are geologically perhaps too recent to be the actual ancestors of man, it is reasonable to suppose that they represent one of the side-branches of the 'proto-hominid' stock which persisted in South Africa after the main stem of Hominidae had elsewhere attained an evolutionary level counted as human (e.g., *Pithecanthropus robustus* in the earliest Pleistocene deposits of Java; see below).

But we are still in need of a working definition of man. Without it, we are in this dilemma: that, looked at logically, the South African fossils have good claims to be counted in the same family as *Homo*, yet they are obviously so much more ape-like in the general proportions of their skulls than the roughly contemporary earliest Java Man, that one hesitates to call them 'human'.

Professor Le Gros Clark, who has closely studied this problem, recently said that 'Probably the differentiation of man from ape will ultimately have to rest on a functional rather than an anatomical basis, the criterion of humanity being the ability to speak and to make tools.'

It is sometimes assumed that the difference in mental capacity between apes and men is accurately reflected in their respective brain-sizes. Unfortunately the facts do not justify the use of this simple criterion, because mental capacity depends rather more on the quality of the cortical association-areas of the brain than on its size. It is true that the comparative psychology of primates shows that the rating of mental capacity is broadly paralleled by the order of magnitude of the brain. The recorded range of adult brain-size in normal man is from *c.*830 to 2,100 c.c., whereas the brains of gorillas range from 340 to 685 c.c. But, as Dr E. I. White has lately pointed out to me, a child of our own species is usually beginning to talk at the age of two years; yet at that age a brain capacity of 650 c.c. is probably well within the normal range. Thus one cannot assume that an adult *Australopithecus* with a brain of that size was incapable of speech. (In any case Dr Broom's latest discoveries in South Africa suggest that some of the australopithecines may have had brains approaching 850 c.c.)

According to recent opinion, no reliable deductions in detail regarding mental aptitudes can be drawn from the pattern of the interior surface of the brain-case. As Weidenreich expressed it, the interpretation of the endocranial cast is no more reliable than any other form of phrenology. Thus it has been claimed that in the brain of an adult australopithecine the association-area for speech was already differentiated; but whether it was qualitatively capable of so functioning is another matter.

The difficulty about accepting the proposal to differentiate man from ape on a functional basis is to find a practical criterion. The definition of man as the tool-making primate has some advantage from this point of view. If deliberately chipped stones (artifacts) had been found in the cave lairs of *Australopithecus* (or allied genera), there would be less hesitation in accepting this fossil as

a hominid. The remains of the two Javanese species of *Pithecanthropus* were found in river or lake beds containing no recognizable artifacts; but this genus can be accepted as human because apart from the fact that its brain-size is within the known human range, stone artifacts were associated with the remains of the Chinese species, *Pithecanthropus pekinensis* (Pekin Man). Deliberately broken stones are fairly easily recognized in a cave deposit, particularly if the stones are obvious introductions, such as the pieces of quartz in the limestone caves of Pekin Man. The difficulty of distinguishing between crude stone artifacts and stones chipped by natural agencies, simulating the work of man, is a serious problem only in glacial, river or beach deposits.

The definition of man as the tool-making primate is sometimes criticized on the score that (1) lower primates occasionally make use of tools, (2) tool-making is merely one of the by-products of man's mental development, and not a primary characteristic from a zoological point of view. It is worth while to examine these criticisms in some detail.

The *use* of tools and weapons is certainly not confined to man. There are records of monkeys throwing sticks and stones; a baboon will sometimes crack open a scorpion with a pebble; while some of the apes observed by Köhler used sticks as levers for digging up objects hidden in the ground, and for extending their reach. The use of tools is not even confined to primates: when the sea-otter finds its favourite delicacy on the sea-floor – a sea-urchin – it takes it on shore together with a pebble to crack it open. The *manufacture* of tools requires mental activity of a different order. The chimpanzee is the only ape reliably reported to make tools, and then only in captivity. Sultan, one of the male chimpanzees observed by Köhler, fitted a small bamboo cane into a larger one so as to make a stick long enough to secure a bunch of bananas which could not be reached by the use of either of the rods used singly (Inset, 23). Once, he attained the same result by fitting into one of the canes a piece of wood which he pointed for the purpose with the aid of his teeth. Apes are thus evidently capable of improvising tools. But it is important to note that all the improvisations effected by Sultan were carried out with

visible reward as incentive. Köhler could obtain no indication that apes are ever capable of conceiving the usefulness of shaping an object for use in an imagined future eventuality. He expressed his conclusions as follows:

'The time in which the chimpanzee lives is limited in past and future ... it is in the extremely narrow limits in *this* direction that the chief difference is to be found between anthropoids and the most primitive human beings. The lack of an invaluable technical aid (speech) and a great limitation of those very important components of thought, so-called 'images', would thus constitute the causes that prevent the chimpanzee from attaining even the smallest beginnings of cultural development.'

In other words, apes are very limited in their capacity for visualizing and thinking about the relationships between objects when those objects are not in sight. The power of abstraction or conceptual thought is basic to the regular manufacture of tools. In apes it is no more than nascent. Mme Kohts in Moscow found that her chimpanzee could select from a collection of objects of extremely varied form those which were of the same shade of colour. This mental isolation of a single feature from a varied field of observation is the dawn of conceptual thought; but Mme Kohts found that ideas such as this rapidly fade in the mind of the ape.

To illustrate more clearly the difference between the mental capacity of an ape and of a human being, the following observations of Köhler are worth quoting. When embarrassed by lack of a stick, a chimpanzee will pull a loose board from an old box and use it. If the boards are nailed together in such a way that they make an unbroken surface, the chimpanzee although strong enough to break up the box will not see 'possible sticks' in it – even if his need of a stick is urgent. Men, on the other hand, seeking to make a tool of a form suited to a particular purpose will visualize its shape in a formless lump of stone and chip it until the imagined tool is actualized. There is the possibility of gradation between these two extremes, perceptual thought in apes, conceptual thought in man; but it seems necessary to stress the contrast because one is apt to be so impressed by the occasional manufacture of tools by apes that there is a danger of

minimizing the gap in the quality of mind needed for such efforts, compared with even the crudest tools of early man, which indicate forethought.

Language must have greatly facilitated the development of systematic manufacture of tools. Oral tradition, in effect a new kind of inheritance, is sometimes regarded as more distinctive of man than tool-making. But speech itself is really a class of tool, 'an invaluable technical aid', in Köhler's description, and depends on a capacity for conceptual thought. The brains of the earliest tool-making Hominidae were probably advanced enough functionally for speech, but nevertheless verbal language may have been a comparatively late cultural development – an invention. The earliest means of communicating ideas was perhaps by gesticulation, mainly of mouth and hands, accompanied by cries and grunts to attract attention ('gesture language').

On the evidence available it appears reasonable to conclude that systematic tool-making, in the broadest sense, would only be possible in primates whose level of cerebral development was in advance of that of modern apes.

It remains to consider at what point in the evolution of the Hominidae did tool-making arise, and whether it can be classed as a fundamental characteristic of man from a zoological point of view. It might be thought that manual dexterity was a limiting factor. Actually, however, the prehensile hands of the less specialized monkeys would be quite capable of most of the ordinary movements in making and using tools, if directed by an adequate brain.

Our unspecialized, five-fingered hands are so well adapted to grasping that they are a clear indication that our remoter ancestors were accustomed to climbing in trees. So long as they continued to lead an arboreal life their prehensile hands, fully occupied by climbing and feeding, had no opportunity or need to develop the habit of using external objects as functional extensions of the limbs. However, some of our early ancestors must have become adapted to spending part of their time walking and sitting on open ground. Their hands thus became free to handle objects, first out of idle curiosity, later perhaps, to some purpose.

full-grown poultry killed and actually eaten by baboons, mostly however by aged individuals. Eggs and chickens are taken by the dozen, by old and young baboons. I have on many occasions actually witnessed apparently organized hunts which often result in the death of the intended victim. The baboons, usually led by a veteran of the troop, surround an unsuspecting three-parts grown Mountain Reedbuck, or Duiker, as the case may be, and on one occasion a young Reedbuck doe was the victim. It would appear that on a given signal the baboons close in on their quarry, catch it and tear it asunder. As a matter of interest I have refrained from interference in these grim encounters so that I would be in a position authentically to record the results. In nine cases out of ten the game animal is devoured limb by limb and after the affair is over all that is to be found are the skull and leg bones.'

Baboons, like some other monkeys and apes too, have powerful canine teeth (Inset, 25), serving mainly for defence against carnivores and for attack in mating duels. It has been suggested that the reduction of the canine teeth in man was an outcome of the use of hand-weapons. But judging by their condition in the australopithecines, the canine teeth were reduced in the Hominidae at an evolutionary stage below that of tool-making. Certainly the 'proto-hominids' would have needed some means of defending themselves in the open, and having their hands free may well have used stones as missiles and sticks or animal long-bones as clubs.

In dry open country (such, for instance, as that inhabited by the australopithecines) the 'proto-hominids', like baboons, might readily have taken to eating flesh particularly in times of drought. Although they lacked teeth suited to carnivorous habits they could easily have killed small mammals. Life in the open set a premium on co-operation. Drawing on our knowledge of the mentality and social life of other primates, it seems not unreasonable to suppose that hunting in hordes the 'proto-hominids' could have killed medium-sized mammals, say by cornering them and using improvised hand-weapons such as they might earlier have learnt to improvise for their own defence. This is frankly speculation. Is there any real evidence that the early Hominidae were carnivorous and that they passed through a tool-using stage, before becoming tool-makers?

Even if the australopithecines are geologically too late to be actual ancestors of Man, their 'morphological dating' is plain – structurally, they may be viewed as slightly modified descendants of our Pliocene forebears. There are indications that they may have been carnivorous, and that possibly they used improvised tools and weapons. The quantities of broken animal bones and fragments of egg shells found with *Australopithecus* in the cave deposits at Taungs suggest a 'midden' of food refuse; but the possibility cannot be ruled out that the bones were introduced by a carnivore which left no other trace of its presence. Professor Dart claims that some of the long-bones of antelopes found in the australopithecine layer at Makapan show signs of having been used as weapons. But again, although there is a strong probability that these small bipedal primates were tool-users, the evidence remains *sub judice*.

By the time that the Hominidae had evolved into tool-makers they were evidently largely carnivorous – quantities of meat-bones were associated with the remains of *Pithecanthropus pekinensis*. It is easy to see how the one habit led from the other. Although the killing of game may have been accomplished easily enough in some such way as that suggested above, the early hominids must often have encountered difficulty in removing skin and fur, and in dividing the flesh. In the absence of strong canine teeth the solution would have been overcome most readily by using sharp pieces of stone. Here surely was the origin of the tradition of tool-making. Where no naturally sharp pieces of stone lay ready to hand, some of the more intelligent individuals saw that the solution was to break pebbles and produce fresh sharp edges. Once the tradition of tool-making had begun, the manifold uses of chipped stones became obvious; they were useful for sharpening sticks for use in digging out burrowing mammals, for making spears sharp enough to be effective weapons in hunting larger game, for scraping meat from bones, splitting them to get at the marrow, chopping the meat into convenient mouthfuls. All the main uses of stone tools were, I suggest, connected in the first place with adoption of semi-carnivorous habits.

From the endowment of nature we should be vegetarians. We

lack the teeth evolved by true carnivores, and we have the long gut associated with a herbivorous diet. Furthermore, we are the only members of the Hominoidea which are accustomed to eat meat on any considerable scale. It is true that anthropoid apes, like most herbivores, consume small quantities of animal protein; some of them occasionally rob birds' nests of eggs and fledglings, but by and large they are fruit and plant eaters.

One can well imagine that a changing environment, for instance during a period of desiccation, may have produced an abnormal appetite in the early hominids. Gorillas in captivity quickly develop a liking for meat, and this appears to be due to a change in the flora and fauna of their intestines. Normally their intestines are richly supplied with ciliate protozoa (Infusoria), which serve to digest cellulose. According to Reichenow, under the abnormal conditions of captivity the Infusoria are ingested, and with their disappearance from the intestine the animal develops an abnormal appetite, and readily takes to eating meat – and may even prefer meat to its normal fare.

By widening his diet and becoming a tool-maker man became the most adaptable of all primates. The change from herbivorous to semi-carnivorous habits was important from the point of view of the use of energy. To obtain a given amount of energy a carnivore subsists on a smaller quantity of food than a herbivore. Instead of eating almost continuously like their ancestors, the Hominidae spent much of their time in hunting. This led to increased interdependence. New skills and aptitudes were developed through this new way of life, and with increasing control of environment through the use of tools, man became the most adaptable of all creatures, and free to spread into every climatic zone.

An important step in the control of environment in northern climes was the making of fire. This could only have been discovered as an outcome from tool-making, but we do not know when or how. Professor Dart's claim to have found evidence of the use of fire by *Australopithecus* has not been generally accepted. Pekin Man regularly used fire, and presumably knew how to make it. The much later Acheulian hunters of Olorgesailie do

not appear to have been fire-users; probably, like some Eskimo tribes at the present time, they ate meat raw.

To sum up: I think it may fairly be claimed that tool-making is one of man's fundamental characteristics from a biological point of view. But the definition of man as the tool-making primate carries the implication that one should apply the term human only to the later members of the family Hominidae, which had attained a certain level of cerebral development, and which for convenience should perhaps be included in the single genus *Homo* (with *Pithecanthropus* one of the subgenera).

The Hominidae as a whole are bipedal primates with a distinctive dentition, which probably became differentiated during the Miocene period. The australopithecines, probably of early Pleistocene age, are believed to be slightly aberrant descendants of Pliocene Hominidae. Through following various indirect lines of evidence it appears to the author most probable that in course of adaptation to life in open country the early Hominidae relied to some extent on the use of improvised hand-weapons for defence; that they became addicted to flesh-eating during periods of aridity; and that when a certain level of cerebral development had been reached in their evolution this change of dietary habit led to regular manufacture of tools and weapons.

REFERENCES

Clark, W. E. Le Gros. *History of the Primates*, 2nd ed. London, British Museum (Natural History), 1950.

Dart, R. A. 'The Predatory Implemental Technique of Australopithecus.' *Am. J. Phys. Anthrop.*, 7 (N.S.), 1 (1950).

Köhler, W. *The Mentality of Apes*, 2nd ed. London, 1927.

Leakey, L. S. B. 'The Early History of Man.' *Sc. News* 17, p.36 (1950).

Reichenow, E. *Arch. f. Protistenk.* **41**, 1 (1920).

Yerkes, R. M. and A. W. *The Great Apes*. New Haven, 1929.

Zuckerman, S. 'Taxonomy and Human Evolution.' *Biol. Rev.*, **25**, 435 (1950).

THE ORIGIN OF LANGUAGE

R. A. S. PAGET

LANGUAGE is so enormously important to mankind that we might have expected it to take pride of place among all the sciences. It is through the invention of 'words' and 'names', and their corresponding 'units of thought', that man has acquired the power of logical thinking and the gifts of imagination and invention. But the fact is that philology – the study of language – is at present hardly at all concerned with the nature and origin of language itself. Language, it is argued, must be as old as mankind; how then can there be any evidence, now, of the ways in which early man expressed himself? Yet geologists, astronomers and anthropologists have not been afraid to study the origins of their various sciences, or to endeavour to reconstruct the distant past by the systematic study of present conditions.

And now, at long last, an orthodox philologist – Professor Alexander Jóhannesson, late Chancellor of the Icelandic University, a high authority on the Indo-European languages – has had the courage to break down the professional barriers which have separated philology from anthropology, and to risk his reputation by putting forward unorthodox views as to the origin of language itself. His results have been published, in English, as a volume of four essays, with a preface by G. R. Driver, Professor of Semitic Philology, of Magdalen College, Oxford.

Jóhannesson writes as a philologist addressing philologists. He has studied systematically 20,000 words of the Icelandic language, and the 2,200 Indo-European roots from which they were derived; but he has also examined the movements of lips, tongue, jaws, etc. – in other words, the 'gestures of articulation' by which all these words and roots are produced. As a result of this formidable examination he announced, in 1943, that a large proportion of the 2,200 Indo-European roots were due to gestures of

articulation which were themselves unconscious copies of the pantomimic hand-gestures which must originally have represented the meanings of those various roots. More lately, he has extended this work so as to include roots from six 'unrelated' families of languages, with results which are similar, so far as he has gone. He concludes that by comparing the various sound-groups of the languages of the world, it will be possible 'to reconstruct the main characteristics of human speech in its genesis.'

To most people, human speech appears to be a system of significant sounds, and the idea of relating speech-sounds to primitive hand-gestures must seem highly inconsequent. It will be well, therefore, to begin our investigation with the study of the sounds of speech, and of the mechanism by which they are produced, and then see where this leads us.

The first serious investigator, so far as I know, of the nature of human speech was M. De Kempelen, 'Conseiller aulique actuel de sa Majesté, l'Empereur Roi', who, in 1792, published his results in *Le Mechanisme de la Parole*, together with a detailed description of his talking machine. To De Kempelen, human speech in its most extended sense is the faculty of communicating man's sentiments or thoughts to his fellow-creatures by *signs*. He recognized two processes. The first is taught by nature herself: it is common to all animals, and is limited to a few ideas. [This we may call the language of the emotions.] The other has to be learnt, and is unlimited in its scope. [This we may call the language of mouth-gesture.]

The basis of all language is imitation: hence the ability of explorers to make themselves understood to the natives of all countries – viz., by making pantomimic signs. De Kempelen also realized that the language of imitation was 'capable of being perfected' to the same degree as spoken language [i.e., so that each sign became the equivalent of a spoken word] but that this development was not natural, it had to be learnt. He realized, too, the important relationship between human speech and the natural sign language of the born-deaf – a relationship which orthodox philology has almost completely ignored.

De Kempelen had studied the work of l'Abbé de l'Epée, the great French teacher of the deaf, and the surprising art of lip-reading by which the intelligent deaf (provided they were familiar with the spoken language) could understand normal speech by watching the speakers' gestures of articulation. He concluded that the various sounds of speech are simply the orderly results of the *positions of the articulating organs.* De Kempelen showed astonishing insight; but he failed to realize that the organs of articulation were themselves influenced by pantomimic hand-signs, and he thought that the sounds of speech were 'agreed' among the earliest speakers, and that this agreement insensibly acquired the force of law. De Kempelen studied the various European speech sounds, and the structure, situation and movements of each organ involved. He realized that voice is not essential to speech, but is useful to make speech audible at a distance – and that the larynx may be compared with a hunting horn, in that the vocal cords act like a trumpeter's lips, and that tightening the vocal cords raises the musical pitch of the hummed note. By trial and error he built a talking machine which reproduced nearly all the vowels and consonants, but he never realized that speech is essentially a musical effect – due to the tuning of the natural resonances of the various cavities produced, opened, or closed in the process of articulation.*

The next important contribution came from an unexpected source, Charles Dickens. In *The Pickwick Papers*, he described Sam Weller laboriously writing his valentine to Mary, the pretty housemaid, and at the same time *forming with his tongue imaginary characters to correspond.* Some thirty years later, this same sympathy of hand and mouth was noticed by Charles Darwin. He pointed out that children learning to write were seen to move their tongues in a ridiculous fashion, and that persons cutting with scissors often moved their jaws in unison with their hands.

* The natural tuning of the vocal cavities in forming the vowel sounds Ah, Eh, Ee, Aw and Oo, can be heard by shaping the mouth *as if* to articulate the sounds in succession (but without actually sounding them) and by flicking the cheek at each successive mouth-posture so as to disturb the air in the vocal cavities.

Here was direct evidence to explain how, in the beginning, primitive man's attempts to explain himself by natural hand-pantomime must have resulted in *unconscious* mouth movements accompanying the hand pantomime. And if these unconscious mouth movements were themselves accompanied by emotional sounds – expressing the pantomimist's anxiety to be noticed and understood – then speechlike sound would inevitably result.

It is interesting to note that the dawning, as it were, of hand and mouth sympathy is to be seen in the chimpanzees; but with this difference, that the sympathy appears to be limited to hand with lips and jaw, and not to hand and tongue. An example is the protruded hand and lips which, when accompanied by an emotional grunt, produces a speech-like sound which may be transcribed as ögh, or ügh – meaning 'give me' or 'come to me' or 'I want'. An extensive study of the 'language' of the anthropoid apes has been made by Dr G. Schwidetzky.

But speech-like sounds fall very far short of verbal speech as we know it. Man's natural way of observing, thinking and expressing his thoughts is seen in young children, as Professor Piaget has noted, and in the uneducated born-deaf of all ages. They observe, think, and express themselves in terms of *events as a whole*, and not in terms of words referring to one thing at a time. In fact we never see 'one thing at a time', but we have learnt to pretend that we do, and to ignore the surroundings of each selected element. We can then give each element a name or symbol, and thus acquire an entirely new power – namely, that of re-combining old symbols in our minds so as to form new patterns – the power of imagining and inventing. This was man's greatest discovery, and the basis of all verbal language; we learn it at our mother's knee, but it is not natural to man, it has to be learnt. Miss Helen Keller, deaf, dumb and blind since the age of 18 months, has described in her autobiography the wonder of the moment when, at the age of 7, she suddenly realized that 'everything had a name'. From that moment, she began to think like a normal child.

Now let us return to the pioneers of scientific philology. High honour is due to Dr J. Rae of Hana, Maui, Honolulu, who in

1862 described his theory of the Polynesian language in an essay of some 14,000 words, published in *The Polynesian*. This essay was referred to by Professor Max Müller in 1863, in a lecture at the Royal Institution. But Max Müller was not convinced, and Dr Rae's theories might have perished, had not the British Museum Library possessed a bound volume of *The Polynesian* from which the essay was reprinted as an appendix to my *Human Speech* in 1930.

Dr Rae's conclusions may be summarized as follows: (1) That the Indo-European languages have their true root and origin in the Polynesian language [which would not now be conceded]. (2) That the study of the Polynesian language gives us the key to the original formation of language itself, and to its whole mechanism. (3) That the primitive form of spoken 'word' would express simply the completed action [corresponding with De Kempelen's view]. (4) That the lips, tongue and mouth as a whole have resemblances to external objects and actions. (5) That the primitive significant sounds were all monosyllables, and expressed force, form and movement.

It will be noted that Dr Rae assumed that the human mouth pantomimed, as it were, on its own account, and not as understudy to the human hands. He was evidently unaware of the 'Sam Weller effect'. According to Dr Rae, the same word may be verb, adjective, noun or adverb, and every syllable keeps its own proper significance and force, even in the longest words. Thus, he says, *KA* denotes 'a forcible action proceeding from a definite point' [actually, *KA* means to cut, in a great variety of different languages. It is the sympathetic result of the hand gesture of a jabbing cut with the primitive hand-axe – the so-called 'coup de poing']. *KAA* means a forcible *and continued* movement commencing at a definite point – e.g., a stone rolling down hill, hence, a wheel carriage. *KAKA* denoted any quick repeated movement proceeding from a definite point – e.g., the hand – hence *KAKAWELA* 'to beat out a fire', and *KAKALA* means (*inter alia*) 'the striking of a cock with his spurs', and, hence, the spur itself. Dr Rae compares the Latin 'calcar', or *KALKAR*, meaning a 'cock's spur'. Of the word *MILO*, to spin, he says:

'the fingers take the place of lips and the thread that of the slender current' [meaning, surely, that the lips take the place of the fingers] and that *LO* is for 'long'. Dr Rae clearly saw that, in Polynesian speech, the mouth-gestures were essentially pantomimic, and that, in some cases at least, as in *MILO*, to spin, the lips imitated the human fingers.

After Dr J. Rae, the most notable contribution to the Gesture Theory came from Charles Darwin's rival, Dr Alfred Russel Wallace, who in 1895 pointed out, in the *Fortnightly Review*, that, in English speech, it is common to produce words by an appropriate gesture of the tongue, lips or jaw, so as 'to bring sense and sound into unison'. Thus, in *UP*, the jaw makes an upward movement, while in *DOWN*, the jaw moves down. Continuing consonants, such as *F, L, M, N*, etc., symbolize continuing motions, such as fly, run, swim, move. On the other hand, words for abrupt motions end with a stopped consonant – e.g., *B, D, G, K, P, T*, in stop, hop, pat, stab, kick, etc. Dr Wallace considered it in the highest degree probable that the pantomimic use of the various parts of the mouth constitute 'a fundamental principle which has always been at work, both in the origin and in the successive modifications of human speech'. Dr Wallace did not recognize Dickens' observation of hand and mouth as exemplified by Sam Weller; but he was, I believe, the first to point out that the pantomimic principle may be still active in man's unconscious development of his spoken language, and that modern languages may be just as gestural as the older ones.

Having now established a reasonable case for the gestural nature of human speech, it may be well to follow the advice of Plato and Leibnitz and study the much neglected art of 'sign language', from which oral language appears to have been derived.

The most widely distributed form of sign language is the natural pantomimic language of the uneducated born-deaf. This is a truly universal language, but its scope is very limited because it has no signs equivalent to words; it signs events. The Red Indians of North America had, until recent times, a silent sign

language, by which all the various tribes who were liable to meet on the plains hunting the buffalo could communicate freely with one another; it had a vocabulary of over 1,000 signs, many of which have been recorded. The Australian aborigines of north-west-central Queensland had a very similar sign language which they used during hunting, when silence was essential, and during various tribal ceremonies. Certain mediaeval monastic orders forbade speech to their votaries – with the result that they developed a sign language with a vocabulary covering all their normal activities. Here again the basis of the signs was pantomimic. A curious use of silent signs was discovered in Russian Armenia about the year 1934. It was used only by women: for women were, by tribal custom, forbidden to speak to any man, and even to their own sons after the age of 8. So the women used sign language while their men-folk answered in spoken words. Another sign language was discovered in the German Cameroons some 15 years ago, by Mr Ivan Sanderson; it was used for settling inter-tribal disputes.

Intermediate stages are also found, in which a primitive language depends to some extent on hand-signs to make its meaning clear. There may be no words or verbal forms to indicate present past and future, and a hand-sign has to be used to indicate the 'tense'. Or there may be no words for number, and the speaker must point to his fingers, or even to his fingers and toes, to indicate actual numbers. We ourselves, if asked to describe a concertina or a spiral staircase, would be apt to fall back on sign language.

At this stage, it may be appropriate to refer to some personal investigations which I have made of the nature and origin of human speech since the year 1921, more particularly as the books in which they were described have all long since been out of print. In contrast to De Kempelen, of whose work I did not then know, I made a systematic study by ear of the resonances of my own vocal cavities when whispering the vowel sounds. It appeared that in every case there were two principal resonances, one formed by the cavity in front of the humped tongue, and another in the cavity behind the tongue (see Fig. 1); other minor resonances of

higher pitch were heard, but these did not seem to be significant. The resonances were recorded, and I then made plasticine cavities tuned (by adjustment of their volume and size of orifice to air

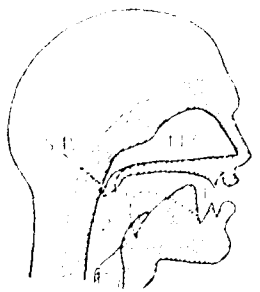


Fig. 1. Two cavities, in front of the humped tongue and behind the tongue, form the principal vocal resonances

or to the adjoining cavity) to give the pair of resonances heard in each vowel when air was blown through the cavities (Fig. 2). If the air-stream was first set in vibration by passage across a

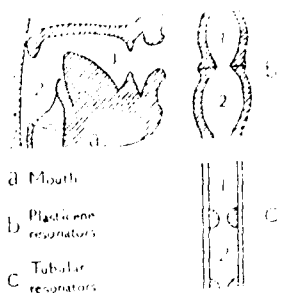


Fig. 2. Plasticine cavities can be used to give the pair of resonances heard in each vowel. Further explanation is given in the text

vibrating rubber strip, or (later) through an organ-reed – the resultant sound was a voiced vowel. The resonances of the various consonants were similarly recorded. They proved to be as musical

as those of the vowels, but differed from them in that the consonants were, generally speaking, the musical results of oral gestures, whereas the vowels were the audible effects of mouth postures. In *M*, *N* and *NG*, the nasal cavity (*N.C.* in Fig. 1) becomes a resonator by the lowering of the soft palate (*S.P.* in Fig. 1). The consonant resonances were therefore more transient effects, and could be imitated by appropriate movements of the resonating cavities. The understanding of human speech is therefore a process of musical analysis, and every child who learns to speak must have, originally, a highly developed ear for music – or he would never learn to identify and imitate these extremely subtle effects. Children with 'no ear for music' are those who are interested only in the music of human speech. As to the gestures of articulation, it soon became clear that these were pantomimic, and due to the Dickens-Darwin sympathy of hand and mouth.

My own amateur efforts at testing the validity of the 'gesture theory' in Sumerian, Chinese, Polynesian, etc., have now happily been superseded by Professor Jóhannesson's systematic technique. But a word may be said as to certain gestures in English which (so far as I know) Professor Jóhannesson has not yet investigated. We have only one tongue, and it presumably tends to work in sympathy with our leading hand. But there are many actions – e.g., clap, climb, clasp, clean (one hand wiping the other) in which *both* hands are normally involved. How does the human tongue deal with these actions? Apparently by making two *simultaneous* closures – e.g., one of the back of the tongue against the back of the throat (to produce the consonant *K* or G) and the other of the tip of the tongue against the palate (e.g., to produce the consonant *L*). We can therefore *CLAIM* (holding out *both* hands) that, in English at least, *KL* (which we inconsequently spell *CL*) commonly means 'with both hands'. Similarly, *KR* as in crush, crawl, cream (one hand skimming the contents of the other) and crowd, or *GR* in grasp, grip, grab, etc., give further examples of two-handed gestures represented by double contacts of the tongue.

Normally, lateral movements of the tongue are not used in speech, since they do not, in general, produce any recognizable

change of resonance. This can be illustrated by trying to voice the vowel sound *AH* while moving the tip of the tongue from one corner of the mouth to the other. But there is an interesting exception in the case of the consonant sounds *-DL* or *-TL*, as in saddle, waddle, dandle, toddle, and bottle. The tongue is raised and broadened so that it forms a complete transverse closure (Fig. 3, *D* or *T*). Keeping the tongue tip in contact with the roof of the mouth, the two edges of the tongue are lowered so as to leave clear passages on either side, while the tongue itself forms a saddle or bottle gesture. (Fig. 3, *L*). It seems possible that the double closure and release of *GL*, as in glance, gleam, glitter, glamour, etc., may refer to the two eyes. But all these effects need to be studied in other languages.

Fig. 3. Showing the use of the tongue in the consonant sounds *-DL* or *-TL*. Further explanation is given in the text

Anthropologically speaking, the uneducated born-deaf are living fossils – indicating the mental condition of early man, before the discovery of verbal symbolism. The long period of time – 300,000 years – during which the hand-axe was used without apparent change, becomes easy to understand if we assume that man was still in the stage of generalized pantomine and gabbling, but had not achieved word-making, and that he was consequently unable to invent. This condition may well have lasted through the succeeding and slightly more advanced Levalloisian (tortoise-core) culture of perhaps more than 150,000 years, which followed the hand-axe period. On the other hand, the Aurignacian cave drawings – of, possibly, 20,000–30,000 years ago – give clear evidence of individuals who had acquired the mental trick of isolating the element of shape and, later, those of colour, number, etc., from their general impressions of the animals over whom they wanted to gain magical power. It may

well be that these gifted individuals kept their new way of thinking to themselves, and thus maintained their authority over the gabbling tribe during the Aurignacian period.

It should be realized that all human speech is still in a very primitive stage of development, and that all our languages are – as Professor G. M. Trevelyan pointed out in the case of English – the inventions of illiterate peasants. It is not therefore surprising that, even the best of them are riddled with relics of barbarism – homophones, ambiguities and irregularities – and that they are becoming still worse owing to careless articulation and the lack of guiding principles.

And now to return to Professor Jóhannesson. In his first essay he describes how after having written half his first book – the result of 11 years' work – on the Icelandic words and their Indo-European roots – he satisfied himself that a very large proportion was due to unconscious imitation by the mouth of pantomimic hand-gestures. He then saw my book *Human Speech* (1930) and realized that very similar results had been arrived at by two very different methods. He wrote to me, we met, and since 1944 have corresponded and to some extent collaborated. In this essay, Jóhannesson compared Icelandic and Hebrew, and showed that the consonants *D*, *T*, and *TH*, and *L* and *R*, commonly had the same meaning in both languages; similarly *M* (made by closing the lips) meant 'silence' in both languages. It was also clear that the mouth gestures were the results of *unconscious* imitation of hand-gestures. Jóhannesson then examined the names of parts of the human body in Indo-European and Hebrew. He classified the various types of mouth-gestures and the hand-gestures to which they appeared to relate. Of 320 Indo-European roots, 274 (85 per cent) were explicable as pantomimic mouth-gestures. Similarly, of 78 Hebrew words designating parts of the body, 60 (80 per cent) were found to be gestural. Jóhannesson also made some investigation of Sumerian; but the material available to him was only sufficient to show 'that the Gesture Theory is evident in this language.'

In his next essay, on Hebrew and Icelandic, he discusses the development of the various Semitic languages, and the

unsuccessful efforts that philologists have made to establish direct relations between these and the Indo-European languages. He then shows that the Hebrew roots in *T–*, *D–*, *DH*, combined with *–B*, *–K*, *–Q*, *NP*, *P*, *H*, and *R* tend to show the meaning to stretch or draw – corresponding with the tongue-gestures involved. Finally, he considers the consonants *L*, *R* and *M*, and concludes that the resemblances between roots in Semitic and Indo-European languages are explicable as the results of their speaking organs imitating the sign language of the dumb preman, i.e., gestures and also natural sounds. Any common language (if such ever existed) must be at least 10,000 to 20,000 years old.

The next essay deals with the origin of the Indo-European *EU–*, of which the author examines all the 190 original 'constructed' roots containing an *–EU–* component. Most of these roots are found to mean to swell, become round, vaulted or curved. Of the 196 roots in *EU*, Jóhannesson separated those which show only abstract meanings, e.g., *GHLEU–* (be glad), *NEUDH* (desire), *LEUBH* (love, long for), and those which imitate natural sounds. Of the remaining 167 roots, 100 are found to have the curved significance. The Indo-European *EU* is replaced in Hebrew by the consonants *G*, *GH*, or *HB*, as in *GUP*, to enclose, and *GHR*, to bend or crouch. Of 90 *G* roots, at least 41 mean to turn round, be vaulted, enclose, swell, devour, finish, and 10 mean to cut.

The final essay on 'names of parts of the human body in Indo-European and in Hebrew', is the longest of the series. Jóhannesson makes an exhaustive study of the Indo-European words for the various parts of the body under the headings 'head', 'trunk', 'upper extremities', 'lower extremities', and 'common names'. Of about 320, all but 50 are shown to be gestural. He then makes a similar study of the corresponding Hebrew words, and lists 35 words – relating to the headings 'head', 'trunk', etc., in which the gestural symbolism is identical in both language groups. A final survey shows that of 78 Hebrew roots relating to parts of the human body, 60 are gestural, though the method of symbolism often differs, and the Hebrew roots are often more abstract in meaning than their Indo-European counterparts.

Under the title, 'The Gestural Origin of Language – Evidence from six "Unrelated" Languages', Jóhannesson has recently compared the words meaning 'pour', 'bend', 'cover' and 'cut', in the six 'unrelated' languages Indo-European, Hebrew, Archaic Chinese, Polynesian, Turkish and Greenlandic. Of 85 Indo-European roots of the types *GEU, GEM, GEB*, &c., 63 are gestural. Of 144 Hebrew roots, 92 are gestural. Of 107 Archaic Chinese, 90 are gestural. Of 50 examples in Polynesian, all are gestural. Of 72 Turkish examples, nearly all are gestural. In Greenlandic, 68 out of 101 are gestural. Can it be doubted that, with so massive a foundation of expert analysis spread over so wide a linguistic field, the Gesture Theory should now be taken seriously?

REFERENCES

Jóhannesson, A. *Origin of Language: Four Essays*. Preface by G. R. Driver, Reykjavik, H. F. Leiftur; Oxford, B. H. Blackwell, 1950. 'The Gestural Origin of Language'. *Nature*, **166**, 60 (1950)

Paget, R. A. S. *Human Speech*. London, Kegan Paul, 1930. *Babel*. London, Kegan Paul, 1930. *This English*. London, Kegan Paul, 1935

Piaget, J. *Language and Thought of the Child*. London, Kegan Paul, 1926

Schwidetzky, G. *Do You Speak Chimpanzee? An introduction to the study of speech of animals and of primitive men*. London, Routledge, 1932

The following references are to work on particular sign languages, as indicated:

North American Indian

Clark, W. P. *Indian Sign Language*. Philadelphia, L. R. Hamersley & Co., 1885

Seton, E. T. *Sign Talk*. New York, Doubleday Page & Co., 1918

Tomkins, W. *Universal Indian Sign Language*. San Diego, 1927–37

Australian Aboriginal

Roth, W. E. *Ethnological Studies among the North-West-Central Queensland Aborigines*. London, 1847

Medieval Monastic

Vetus Disciplina Monastica. Paris, 1726

THE SCIENCE OF LUBRICATING GREASES

THOMAS D. SMITH

THE practice of lubricating moving surfaces to reduce friction is a very ancient one. There is evidence that animal fats were used to lubricate the axles of Egyptian chariots as long ago as 1400 B.C. Lubrication must also have been used in building the pyramids. It is believed that the massive stones which were used in that enterprise were slid along wooden planks which had been smeared with olive oil. A decoration in the tomb of *Tehuti-Hetep*, 1650 B.C., shows some early methods of lubrication which employed natural fats and oils.

Simple fatty materials were still being used as lubricants in the 19th century, although the Industrial Revolution was introducing machinery of ever more complicated design. As engineering craftsmanship improved and more highly finished and precise machines were produced, untreated animal fats and oils were found to possess severe disadvantages as lubricants. Attempts to improve the quality of greases by scientific techniques, which had begun late in the 18th century, increased rapidly after a British patent was issued in 1835, describing an early cup grease. Although natural animal fats and oils have only a limited ability to act as lubricants, there was no alternative until the Drake well had been drilled in Pennsylvania in 1859, which made mineral oils generally available. Greases incorporating mineral oils were rapidly produced and solved most of the lubrication problems of the time.

In the past fifty years machinery has been designed to operate under increasingly severe conditions. Bearings run at high speeds, producing high pressures and high temperatures. Without corresponding improvements in the properties of lubricants, this progress in engineering would have been impossible and many

modern machines, which operate undamaged at the highest speeds, would not have reached the drawing-board. However, as greater demands have been made on the lubricants, scientific research on their preparation and uses has increased, until to-day there are thousands of papers and patents describing materials which will lubricate under all manner of conditions. In particular, grease-making, always a difficult and uncertain operation, is ceasing to be an art and is becoming a science.

The word grease comes from the early Latin 'crassus' and the late Latin 'grassus', meaning fat. It is used very loosely to cover a broad range of fatty and oily materials which may differ very widely in chemical constitution and physical properties. Although considered as gelatinous materials, the consistency of greases may vary from semi-fluid, liquid-like materials to very hard blocks; from very fibrous, 'stringy' gels, to materials resembling butter. Some investigations on the structure of greases will be described in later sections. It has been found that most, possibly all, greases contain extremely minute fibres. Most present-day greases consist of petroleum oils which have been thickened with soap. Other special materials have also lately been used; and, as knowledge of the structure of greases is increased and is applied, further advances may be expected.

Greases are used as lubricants in situations which require an adhesive, non-splashing material where oil would simply run off the rubbing surfaces, or in bearing housings which would not hold oil; for example in ball bearings, caterpillar tractors, exposed gears, etc. They are usually very complicated systems, in many cases little being known of their structure. However, the basic principles involved in their manufacture are quite straightforward. The soaps used in grease-making are frequently obtained from tallow, the product of rendering the suet or adipose tissue of cattle or sheep, and are made *in situ* during the grease-making process. Tallows consist mainly of triolein and tristearin, which are compounds of glycerine with oleic acid and stearic acid respectively. Their chemical formulae are shown on the next page.

$CH_3(CH_2)_{16}.COOH$ — stearic acid

$CH_3.COOH$ — acetic acid

$CH_3(CH_2)_7CH=CH(CH_2)_7COOH$ — oleic acid

CH_2OH
|
$CHOH$
|
CH_2OH
glycerine

$CH_2.O.CO.(CH_2)_{16}.CH_3$
|
$CH.O.CO.(CH_2)_{16}.CH_3$
|
$CH_2.O.CO.(CH_2)_{16}.CH_3$
tristearin

$CH_2.O.CO.(CH_2)_7.CH=CH.(CH_2)_7.COOH$
|
$CH.O.CO.(CH_2)_7.CH=CH.(CH_2)_7.COOH$
|
$CH_2.O.CO.(CH_2)_7.CH=CH.(CH_2)_7.COOH$
triolein

Stearic and oleic acid are long chain fatty acids. They are rather like acetic acid with the addition of a long chain of carbon and hydrogen atoms, but unlike it in that this long chain restricts the movement of the molecules, causing them to be solids at room temperature. Soaps are the products of reacting caustic alkali, slaked lime, or other basic material with long chain fatty acids. If tallow is treated with sodium hydroxide, the tristearin is attacked, forming glycerine and sodium stearate. Similarly, triolein yields glycerine and sodium oleate. These sodium compounds are the soaps, being similar in constitution to ordinary household soap.

Having prepared a soap, the next step is to incorporate it into an oil. If a powdered soap is simply mixed with oil, a suspension of soap particles in oil is obtained—no grease-like gel is formed. In other words, soap is insoluble, or very slightly soluble, in oil. However, if the temperature is raised to about 400°F. (*c.* 200°C.), the soap dissolves very quickly and easily. On cooling, a gelatinous grease-like material is formed. This process of the solution of a soap in a mineral oil is the fundamental process of grease-making.

There are hundreds of variations in procedure and many different soaps are employed. For example, the tallow may be replaced by other animal fats, by fish or animal oils, or by vegetable oils (such as the oil from various seeds), and the base may contain aluminium, barium, calcium, lithium, or strontium,

instead of sodium. Other bases could be used, but those mentioned are the most general. However, many of these bases form soaps which, when simply dispersed in oil, yield unstable gels from which oil separates readily. In order to prepare greases from these soaps, other materials must be added. These 'additives' combine with the soaps to produce compounds which will dissolve in oil on heating, and give stable greases when later cooled.

Calcium base greases, which constitute most cup greases, are stabilized by addition of a small quantity of water. If they are dehydrated (by heating), they break down and lose their lubricating properties. They are not heat-resistant. On the other hand, calcium soaps are insoluble in water and greases made from them are water-resistant. Sodium base greases, which do not require water as a stabilizing agent, retain their grease structure at relatively high temperatures. However, sodium soaps are soluble in water, so that these greases are not water-resistant. For many years only sodium and calcium base greases were manufactured, and the ideal grease which would be both water- and heat-resistant was energetically searched for. The problem was solved, after many attempts, by the invention of barium base greases.

This discovery initiated a period of great activity during which many new greases were produced that could be used as lubricants in a wide variety of applications – the so-called multi-purpose greases. New products were invented which could efficiently lubricate hot surfaces and others that would be useful at the freezing temperatures encountered by high-flying aircraft and in many environments. Quite recently, lithium base greases were devised which can be used at very low temperatures. Apart from the additive compound which is involved in the production of the grease structure, very small quantities of special chemicals are sometimes added to act as inhibitors of oxidation or corrosion, 'tackiness' improvers, or to improve the lubricating properties. Occasionally, finely divided solids such as graphite, metallic oxides or horsehair are used as 'fillers' to stabilize the grease structure and improve its properties for certain specialized uses.

The appearance of greases is improved by the addition of dyes, and the odour by perfumes. The finished product is put through numerous tests to ascertain its consistency and lubricating efficiency under varying conditions.

It has been stated that a grease is a system of soap, or a soap complex of some kind, in oil. Commercial greases may contain from 5 per cent to 40 per cent of soap. Certain soaps may thicken or gel an oil even though present in a concentration of only 1 per cent. Other soaps do not form stable gels with oil at any concentration. The questions now arise as to how the gel structure is formed and how such a dilute system possesses rigidity. Such phenomena have only recently been explained, and there is much that is still unknown. We are moving out of the 'art' phase of grease-making, into the era of scientific method. By the application of modern chemical and physical techniques such as electron microscopy, *x*-ray and electron diffraction, light microscopy and calorimetry, the nature of grease systems is being elucidated.

Most greases, when observed under the microscope, appear to contain minute particles or fibres of soap. Some contain long fibres which are visible to the naked eye, but even the 'buttery' greases contain invisible, microscopic fibres. Under the electron microscope it is possible to photograph particles of almost molecular dimensions, utilizing magnifications × 60,000 or more. With this instrument, all greases, with the possible exception of aluminium base greases, have been found to contain extremely minute fibres, sometimes less than 1/100,000 inch in length and less than 2 or 3 millionths of an inch in thickness. The dimensions, however, may vary widely and differences in the fibre sizes undoubtedly account for many variations in the grease properties. The grease structure may be visualized as a felt or network of soap fibres which enmesh the oil much as a sponge holds water. But why does the oil not run out?

If the end of a fine-bored capillary tube is placed in water, it is observed that the water rises in the tube and remains held there after the tube is withdrawn from the water. A similar mechanism applies in a grease on an even more exaggerated

scale. The capillaries or fine holes which wind throughout the mesh of soap fibres have diameters down to almost molecular dimensions. Oil is easily and strongly held in such spaces. Sometimes these capillaries may be a little too large to remain filled with oil and a small amount will separate and form a pool on the surface of the grease. If there is only a small quantity of 'separated' oil, the lubricating properties of the grease will be in no way affected. The oil can be forced out of the 'sponge' of soap fibres by squeezing in a filter press of some kind.

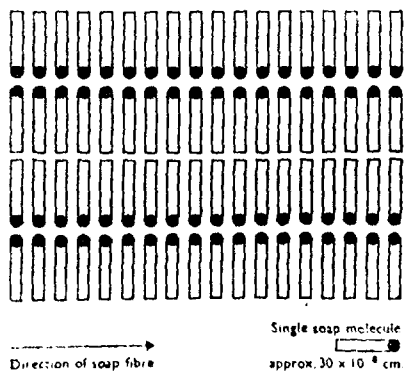


Fig. 1. Diagrammatic representation of a soap fibre.

This very fine dispersion of soap fibres can only be formed by first dissolving the soap in oil by heating, and allowing the solution to cool. When the temperature is low enough, the soap precipitates in the network form, thus holding the oil in a rigid manner. A similar structure is present in the so-called 'solid' fuels, 'canned heat', and semi-fluid liquids used in flame-throwing, which consist of petrol which has been 'thickened' or gelled by the addition of aluminium soaps. Soaps separate in this fibre-like form because of their characteristic molecular structure. This has been shown by the use of *x*-ray diffraction methods.

A diagram of a very fine soap fibre is shown in Fig. 1. It

will be noticed that the molecules lie perpendicular to the direction of the fibre, and that the ends of their long chains project outwards towards the oil. These long chains have a certain similarity to, and an affinity for, the molecules which constitute a large proportion of lubricating oils.

Another line of investigation is the use of 'phase' studies. This means, in effect, the investigation of the form and structure of these systems at different temperatures and at various concentrations of soap in oil. By heating a system of soap in oil it has been found that certain transitions occur during which one structural arrangement changes to a different arrangement, which may be observed visually, or by changes in heat or volume, or by *x*-ray diffraction methods. These changes may be thought of as a type of progressive melting. The interpretation of phase studies is a difficult and laborious process, and as extensive phase investigations on soap-oil systems were commenced only in the past decade, their precise value and meaning is still not entirely clear. However, it may be mentioned that similar studies on soap-water systems proved of tremendous value to the soap-making industry.

There are two main schools of thought on the question: how does a grease lubricate? The older theory, which is still accepted by many workers, is that the soap merely acts as a carrier for the oil. When pressure is applied, the oil is squeezed out of the grease, and lubricates the rubbing surfaces. This is probably only part of the answer. It is well known that soap itself acts as a lubricant. Carpenters, for example, frequently lubricate wood screws with soap. Recent work carried out by electron diffraction and other methods has shown that soap molecules can become absorbed (attached) to a metal surface with the long hydrocarbon chain sticking out, in a direction away from the surface. Such layers can act as very efficient lubricants in that they can separate the rubbing surfaces, and at the same time have a resilient type of structure. It is very likely that both the soap and the oil have an important part to play in lubrication by greases, the individual conditions of the machine deciding which of the constituents is to be more important. Where there

168372

are relatively large tolerances, the oil may play the larger part, but when these tolerances decrease and large pressures become operative, the layers of soap molecules will be of major importance.

Recently there have been a number of greases manufactured which do not involve soap in oil. The development of these products is due to the fact that some recent engineering developments, particularly in the aircraft industry, operate under conditions too severe for petroleum greases. Gels having a structure similar to that described for greases are very common and, although technically not including jellies, are similar to them in many respects. Few such gels have much value as lubricants. However, many materials have been used to thicken petroleum oils. A recent development is the thickening of oils with certain clays which have been given a surface hydrocarbon coating to permit them to disperse in oil. These materials can be dispersed in an extremely fine state of division, and so actually thicken the oil and form a grease. On the other hand, the mineral oil has, in some products, been replaced by synthetic organic chemicals, such as silicones, glycols and certain esters. They contain large, complicated molecules and can be used as lubricating oils. Some of their properties are considerably better than those of petroleum oils. Under suitable conditions, synthetic oils may be thickened by adding a metallic soap, and greases are then produced.

The synthetic products are, as yet, in the development stage, but it is very likely that, for some applications, they will replace greases made from petroleum oil. For example, some silicone greases retain their grease consistency at very low temperatures, at which petroleum oils would have solidified. However, the limits of petroleum lubricating greases have not yet been reached and with the application of modern physical and chemical techniques to problems of grease structure, we can expect a considerable improvement in their properties in the next several years.

RESEARCH REPORT

A. W. HASLETT

ATMOSPHERIC TURBULENCE

ATMOSPHERIC turbulence is a subject which can be approached from many angles. Formally, it is the concern of the meteorologist; but it is also of interest to astronomers and radiophysicists, and to the designers and operators of jet-propelled aircraft. In the following note, a number of recent investigations are brought together.

Dr E. C. S. Megaw, of the Royal Naval Scientific Service, has studied atmospheric turbulence in relation to the twinkling of stars and the propagation of centimetre radio waves. In outline, he has used astronomical information to confirm approximately a statistical theory of air eddies developed by a Russian mathematician, A. N. Kolmogoroff, in 1941; and applied the results to explain the reception of centimetre radio waves at distances well beyond the optical horizon and in weather conditions which apparently are normal. He has made his own observations on the radio effect.

In preliminary discussion, Dr Megaw recalls that the scientific literature of star-twinkling is considerable, though widely scattered and 'evidently not well known', or recent controversy occasioned by a theory advanced by Professor H. Hartridge, that the main or an important cause was physiological, 'would scarcely have been possible'. The weight of evidence over the past 100 years, he sums up, 'leaves no doubt of the correctness of Newton's opinion that the twinkling of stars is caused by variable atmospheric refraction.'

The astronomical information used by Dr Megaw has been of two kinds – the first consisting of fluctuations in the apparent positions of stars shown in 'star trails' from plates exposed at the Yerkes and Mount Wilson Observatories in the United States, and at the Royal Observatory, Greenwich, in Great Britain; and,

second, fluctuations in the brightness of stars recorded photo electrically by Dr J. Stebbins, formerly Director of the Washburn Observatory, and Dr G. E. Kron of the Lick Observatory. Star trails, recorded at the 'natural' rate of trailing due to the earth's rotation, show fluctuations which last from a few seconds to a few minutes, the lower limit being set by the relatively slow movement of the star-image across the plate. Fluctuations, thus observed, correspond with the larger of the supposed eddies; and from the slowest of them, it is concluded that the upper limit of eddy-size may be about 400 metres. For more rapid changes, use was made of brightness fluctuations, obvious to the naked eye, but now recorded photo-electrically. These are connected with the smaller eddies, the smallest being taken to be about 0.1 metres diameter. The distribution of both types of fluctuation – in apparent position and in brightness – is consistent with the statistical theory already mentioned; although it would be 'premature' to regard the corresponding distribution of eddy sizes as thus directly confirmed.

The radio experiments, which represent the second half of Dr Megaw's investigation, arose out of the gradual realization from wartime and later experience that the strength of radio reception on metre and centimetre wavelengths often falls off less rapidly below the horizon than would be expected from simple theory – and this even in conditions of 'normal' weather. The distinction is from special conditions in which radio waves may be canalized within a more or less sharply defined layer, usually at sea-level or shortly above it, and conveyed with relatively slight reduction in signal strength for distances of, possibly, many hundred miles. The latter effect is caused by abnormal conditions of refraction, due to an exaggeration of the usual humidity gradient, and often to an inversion of the temperature gradient as well. For lesser but still appreciable 'anomalies', Dr Megaw reached the conclusion that there was 'a common cause, not in any exceptional condition of the atmosphere', but in scattering due to 'its permanently turbulent state'. Measurements of reception strength were made on 10 cm. wavelength during six passages across the northern North Sea during the summer of 1949, and were extended to distances

of more than 300 sea miles from the transmitter. Other observations, quoted for comparison, have been made by American workers in the Caribbean Sea, and these followed a similar course up to 140 sea miles. Both sets of records agree reasonably with curves calculated from the turbulence theory, but with greater signal strength than would be expected at distances around 150 sea miles.

Dr Megaw suggests also that 'the general similarity of unexpectedly high field strengths recently reported at television wavelengths ... may not be entirely fortuitous'; and that turbulent scattering in the atmosphere probably makes an important contribution, though not the only one, to such results.

An investigation in relation to air transport has been described lately by Dr G. S. Hislop of British European Airways. Its object was to obtain information on an effect known as 'clear air turbulence' which, with the advent of jet-propelled civil aircraft designed to fly regularly at higher altitudes than in the past, has become of practical importance. The effect is described as like 'being in a fast car suddenly running over a series of deep unseen ruts'. There is no measurable height change, probably because the pressure alternation is too rapid.

Flights were carried out over western Europe during 1948 and 1949 in Mosquito aircraft. The flight-plan was to follow a 'saw-tooth' pattern, with altitude varied between 20,000 and 37,000 ft, in a search for turbulent areas. Most of the flights were within the area bounded by Rome, Stockholm, Edinburgh and Lisbon; and measurements of vertical acceleration and of temperature were made in twenty cases.

Turbulence has been found in isolated patches which, typically, are 50-100 miles long and about 3,000 ft thick. On available evidence, they are associated with a high rate of temperature variation in a horizontal direction – which would correspond with high wind gradients vertically. On the other hand, 10 out of 20 turbulent patches encountered were within 2,000 ft of the tropopause – which, as a boundary region of constant temperature, should be free from vertical wind gradients. Estimates of eddy-size were made, which are of interest in relation to

Dr Megaw's estimates based on the twinkling of stars; and of the frequency of bumping in relation to aircraft construction. On the question of eddy-size, Dr Hislop's conclusion was that those recorded were probably of 50-500 ft diameter, but with the qualification that neither larger eddies nor smaller would have been felt as bumps by the Mosquito aircraft used. This may be compared with the upper limit of about 400 metres – or something over 1,000 ft diameter – suggested by Dr Megaw's analysis. Interest in the frequency of bumping is in relation to possible resonance with the natural vibration frequency of the aircraft: for under such conditions, a dangerously large vibration could be quickly built up. Dr Hislop reports that the frequency of bumping in a turbulent patch may remain roughly constant for several seconds, and flying at 350 m.p.h. may be at as high a rate as 3 bumps per second. This, he pointed out, would be the natural frequency of vibration for an aircraft of plausible construction and 130 ft wing span.

Last, and in relation to higher altitudes, mention may be made of an investigation by J. H. McQueen of the University of Virginia, with the object of obtaining some rough indication of the extent of stirring by mixing which takes place at different levels in the upper atmosphere. His method was to compare the proportion of nitrogen-14 to nitrogen-15 in air samples collected at heights of up to 58 km. The extent of separation due to gravity can be calculated; and by partial centrifugal separation in the laboratory (giving the effect of artificial gravity), it has been shown that theory and experiment are in good agreement. The observed separation in the atmosphere was found to be negligible below 45 km., and at all heights was less than the calculated separation, if there were no stirring. The separation at 57 km. height was about 4 per cent. It is concluded that 'there is a good deal of stirring ... especially at 40 km. and below'.

Probable further evidence from the twinkling of 'radio stars' is mentioned in the following paragraph.

Megaw, E. C. S. *Nature*, **166**, 1100 (1950).
Hislop, G. S. *J. Roy. Aeronaut. Soc.* (in the press).
McQueen, J. H. *Phys. Rev.*, **80**, 100 (1950).

RADIO-ASTRONOMY

A number of radio-emitting stars having been discovered, it was to be expected that, with a large enough radio-telescope, it should be possible to detect radio waves from the extra-galactic nebulae, which are galaxies of stars. Using a 218 ft radio-telescope, this has now been accomplished for the Great Nebula in Andromeda (*M.* 31) by R. Hanbury Brown and C. Hazard, of the Jodrell Bank Experimental Station of the University of Manchester. Independently of this work, and using a different method of observation, M. Ryle and others of the Cavendish Laboratory, Cambridge, have found evidence for the emission of radio waves from directions corresponding with the Andromeda nebula and three others.

The radio-telescope used in the Manchester observations was designed by Dr J. A. Clegg, and was built by the research staff of the station during 1947 and 1948. It consists of a receiving aerial mounted on a 126 ft central tower, and around and below this a series of wire reflectors arranged in the form of a paraboloid, 218 ft across, from which radio waves are focused on to the aerial, as if by a mirror of the same diameter. The 'rim' of the mirror is supported some 24 ft above ground level by a series of posts; and, with the central tower vertical, as first erected, the whole arrangement is equivalent to a fixed-direction telescope which looks vertically upwards. Thus used, the instrument was one of enormous power, but could 'see' only such celestial objects as by the rotation of the earth were brought directly above it: those, in fact, of which the declination (elevation above the celestial equator) corresponded with the latitude of Jodrell Bank (about 53° 13′ N.). It has been in regular use for that purpose.

However, in the spring of the present year, it was pointed out by Mr Hanbury Brown that if the risk were taken of working with the central mast tilted 14° to the south, the nebula in Andromeda could be observed. Calculation confirmed that radio waves from the nebula might just be detected – provided that the distribution of radio-sources within our own galaxy was similar to that of visible stars, and that the Andromeda nebula was

comparable with our own galaxy in respect of radio-emission. The experiment was made during still weather in August and September, on a frequency of 158.5 Mc./s. (wavelength about 1.89 metres).

Observations were made on different nights with the tower tilted at varying angles 1–2° apart, so as to cover the known declination or celestial latitude of the nebula (centre, 40° 59′ N.). The beam-width of the radio-telescope is about 2°, so that this gave as good accuracy as was needed. At declinations of 38° 47′ on the one hand, and 43° 00′ on the other – to the south and north of the nebula – there was no observable rise in the level of radio intensity received. But between 39° 33′ and 41° 56′, there was a clearly marked maximum which came at the right time. From these observations, the position of the radio source has been fixed as lying between a declination of 40° 35′ and 41° 15′ (compared with 40° 59′ for the centre of the nebula), and between right ascension 00 hr. 38 min. and 00 hr. 42 min. (compared with the 00 hr. 40 min. for the nebula). Correspondence is thus perfect, within the limit of resolution of the radio-telescope; and the chance that the observed distribution could be mimicked by two or more radio stars located in the same direction as the nebula is less than 1 in 4,000.

If the Manchester radio-telescope is the equivalent of the conventional reflecting telescope used in optical astronomy, the aerial arrangement used at Cambridge is the equivalent of an interference method which, in conjunction with the 100-inch telescope at Mount Wilson Observatory, has been used to give a direct measurement of the diameters of the largest and nearest stars. As adapted for radio use at Cambridge, this consists of twin aerial systems separated by a distance of 110 wavelengths. The effect – whether in the optical form of the instrument, or in its radio adaptation – is to increase by a considerable factor the extent of angular resolution which is attainable; sources being separated and distinguished which with a simple telescope, of comparable total area, would be merged into one.

With this arrangement, some 50 radio-stars have been located in the northern hemisphere, no one of which coincides with any

outstanding visual star. The two most intense radio-stars – for which the most accurate positions are available – do not correspond in either case with any visual star brighter than 10th magnitude: they must therefore be only very feeble emitters of visible radiation, although as radio-emitters they are probably about 100 million times as 'bright' as is the sun under normal conditions. Four of the weaker radio sources, however, correspond significantly in position with four out of the five most prominent extra-galactic nebulae in the area surveyed. One, as already stated, is the Andromeda nebula: the other three are the nebulae *M.* 33, *M.* 101 and *M.* 51.

Such observations, made at Manchester and Cambridge, provide the equivalent of a view of our own nebula – the galaxy – as it would be 'seen' on radio wavelengths by an outside observer; and, in conjunction with the observed intensity of radio waves from the galaxy, as received on the earth which is inside it, give some rough confirmation that radio-sources within the galaxy are distributed in a uniform way. What they do not do is to tell us anything about the number, size or absolute brightness of these sources.

The most direct approach would be to measure their distances – or rather to make that measurement for the relatively few of them which have yet been separated. This could be done, as with visual stars, by observing their apparent change in position when observed at intervals of six months from opposite sides of the earth's orbit. But all that can be said at present is that neither of the two brightest sources can be nearer to the sun than about 100,000 million miles: for, if they were, some change in apparent position would have been detected with the Cambridge equipment. This is farther by a factor of about 1,000 than the sun's distance from the earth, but more than 100 times nearer than the nearest visible star.

Other evidence has been summarized lately by Mr Ryle. It seems likely for several reasons that the relatively few radio-stars yet identified are comparatively near neighbours of the sun, and that others more distant are responsible for the general background of radiation from the galaxy which cannot be

separated. One such argument is that there is no correspondence between the directions of known radio-stars and the shape of the galaxy, as would be expected if more distant stars were among their number. Another is that the distribution of radiation between different wavelengths is similar for the most intense local sources and for background radiation from the galaxy. Finally, from an analysis of the observed intensities of the 25 brightest radio sources in the northern hemisphere, F. G. Smith at Cambridge has obtained a figure for the average separation between radio-stars on the supposition, among others, that they are to account between them for all radio emission from the galaxy. The answer is little different from the average separation between visible stars. In sum, therefore, there is a good deal of evidence that radio-stars are extremely common bodies; that they contribute either the whole or a large part of the total radiation on radio wavelengths received from the galaxy; and that they are distributed in other spiral nebulae besides our own.

For relation to the previous paragraph, it may be noted that radio-stars 'twinkle', as do visual stars, and for the same reason. Measurements of their apparent fluctuations, made at the same time but from different sites, are therefore one more means for the study of atmospheric turbulence.

Brown, R. H., and Hazard, C. *Nature*, **166**, 901 (1950).
Ryle, M. *Monthly Notices Roy. Astronom. Soc.* (in the press).
Ryle, M. *Reports on Progress in Physics*, **13**, 184 (1950).
Smith, F. G. *Nature*, **165**, 422 (1950). Little, C. G., and Lovell, A. C. B. *Ibid.*, 423.
For Mount Wilson interferometer: Johnson, M. *Astronomy of Stellar Energy and Decay*. London, Faber, 1950.

NATURAL PLUTONIUM

The separation of plutonium in millionth-gramme quantities from concentrates of Belgian Congo pitchblende is recorded in a report issued lately by the United States Atomic Energy Commission. Plutonium, prepared on a considerable scale in atomic piles, has thus been added to the list of natural chemical

elements which have been chemically separated. The achievement, described by D. C. Peppard and others of the Atomic Energy Commission staff, is of the same pertinacious quality as the much more important isolation of radium, also from pitchblende, in 1898. But whereas the Curies used a ton of pitchblende to obtain less than a gramme of radium, the proportion of plutonium in the concentrates was little more than one-millionth part as great (7 parts in 10^{12}). The proportion of radium in pitchblende is limited by its relatively rapid rate of radioactive breakdown compared with that of uranium. That of plutonium is limited by the number of neutrons which is available – plutonium-239 being formed from uranium-238 through the absorption of a neutron as the first step. The neutrons of an atomic pile come from the fission of other uranium atoms, chiefly uranium-235, and the same source should be available also in pitchblende from spontaneous fission. But it is thought unlikely from calculation that this can provide the whole supply of neutrons.

Small though it is, the proportion of natural plutonium is apparently greater than can be thus accounted for; and one or more further reactions are invoked to provide neutrons for its formation. They may be released, it is suggested, during the disintegration of atoms of some of the lighter elements – such as oxygen, magnesium, and silicon – through the impact of the high-energy alpha particles thrown out at different stages during the radioactive breakdown of uranium.

Peppard, D. C., and others. *AECD*-2890, June 2, 1950; *Nuclear Science Abstracts*, **4**, 871 (1950).

BACTERIAL VARIANTS

It is a familiar observation that bacterial monstrosities are produced in cultures exposed to penicillin or other antibiotics in concentrations too small to kill. Among the effects described have been an 'enormous elongation', the bacteria growing without division to as much as 5 times their usual thickness and up to 20 times their usual length, and the appearance of single and multiple swellings and of branched forms – as if division had been begun, but could not be completed. There are

cases also in which bacteria, collected from different environments, have been assigned to different species or even genera but have proved to be mutually convertible by changed conditions. And, third, there are many examples of biochemical adaptation, whether through mutation or otherwise; of which the most familiar is the emergence of strains resistant to particular antibiotics, and the most amenable to experiment the ability of the bacteria to utilize this or that substance as a source of energy.

Evidence of the production of bacterial variants in an astonishing multiplicity through the combined action of chloramphenicol ('chloromycetin') and of anti-serum specific to the original strain of bacteria, has been described lately by Dr A. Voureka, working at the Wright-Fleming Institute of Microbiology, St Mary's Hospital Medical School, London. The differences observed extend to shape, size, colony form, biochemical characteristics and resistance to antibiotics. Some of the variants seem biologically stable. One has been maintained for a year, through some 200 sub-cultures, without reversion to normal. The others have reverted after longer or shorter periods. It is suggested that any agent causing the bacterial cell 'an injury or a disturbance of the same degree will probably lead ... to similar variants.'

Dr Voureka's experiments began with the observation of non-typical colonies, leading sometimes to variant strains, in cultures from patients treated with chloramphenicol for urinary infections. In one case, small and non-typical colonies, after retaining their unusual appearance through 19 sub-cultures, became in the end indistinguishable from the original micro-organisms. Another, which after 24 hours' incubation had grown only a few barely visible colonies, later developed a multiplicity of shapes, forms and sizes which persisted through many sub-cultures. Because of such variations, it was realised that the infecting organism might escape observation after treatment, though still present; that, on reversion to normal, it might be responsible for relapses; and that such changes might be significant also in the detection of symptom-free carriers. It was

therefore a natural next step to try to reproduce under controlled conditions the variations which had been observed.

The first such attempt was unsuccessful. Strains of bacteria, which had been isolated from one of the same patients before treatment with chloramphenicol, were exposed to chloramphenicol under conditions of laboratory culture, but failed to show comparable variation. It was therefore concluded that conditions in the body must have influenced the production of variant forms, and the combined effect of chloramphenicol and specific antiserum was therefore studied. The organism selected was a stock strain of *Bact. coli*, for which anti-serum was available; and in the most successful experiments, the bacteria were first inoculated into a sterile culture-broth containing the chloramphenicol, and the anti-serum added 24 hours after inoculation.

Variant strains were usually isolated within 1 to 6 days from the addition of anti-serum, and mostly within 24 hours. Regular observations were made by phase-contrast microscopy (which does not involve the use of staining); and there was available a micro-manipulator of the kind which was developed by Dr de Fonbrune at the Institut Pasteur and which was mentioned in *Science News* 19 in connection with amoebal transplants. The initial effects have been thus described by Dr Voureka: within a few hours from the addition of anti-serum, 'the culture consists of greatly swollen cells, arranged singly, in chains or in masses of chains'. But as the hours pass, 'the bacteria show increased damage. The single bacilli are greatly swollen and practically round. Their lining membrane is no longer sharply defined, but looks hairy, as if it is deteriorating and about to disintegrate. Still later, only severely damaged round bodies and masses of tiny particles can be seen.' It was from such material that variant strains were obtained.

Three types of culture are described with a corresponding 'true-forming variant' bred from each. One, maintained for a year, arose abruptly. Another was produced by steps, by survival of healthy bacteria out of an obviously 'ill' population. One showed increased resistance to chloramphenicol; but in all

others susceptibility was increased compared with the original strain. There were also variations in sensitivity to penicillin, but with differences in detail from that to chloramphenicol; in the virulence of the infections produced when the different strains were inoculated into mice; and in the nature of the sugars which the bacteria could ferment – as well as in the appearance of colonies, and in the size and form of individual bacteria. In addition, other variants, not yet described, were obtained from a strain of *Salmonella typhi*, and from a strain of penicillin-resistant staphylococcus.

Any explanation in terms of selection and favoured growth is dismissed as 'highly improbable' – among other reasons because of the increased susceptibility to chloramphenicol shown by most of the variants. But the one which arose by steps is provisionally attributed to 'a mutation induced by profound disturbance or injury'; and, in general, the most significant feature seems to be the emergence of small and, at first, slowly-dividing forms from the wreckage of disturbance. It is suggested more speculatively that these may illustrate the course of evolution from one serological type of micro-organism to another, or from one species to another. The experiments described can be no more than a beginning: it is of interest that they were undertaken by a Greek doctor, working with an American antibiotic, in a British hospital, using techniques of microscopy which originated in Holland and of micromanipulation which were developed in France, and with financial support from the United States.

Voureka, A. *Lancet*, i, pp. 27 and 28 (1951).

TORRICELLIAN CHAMBER

An unusual problem presented by the unexpectedly high water level in a Derbyshire cave roof has been solved by Dr R. E. Davies, of the University of Sheffield, as the result of observations carried out since August, 1947, with the help of members of the Cave Diving Group and the Sheffield University Mountaineering Club. The problem consists in the fact that the cave roof is normally filled with water to the top, although connected by a

continuous water passage with a section in an upstream direction which is open to the air and where the level is more than 5 ft lower. The roof chamber, which lies in the Buxton Water Passage of Peak Cavern, Castleton, was discovered three summers ago, during exploration from a diving base 300 ft. distant, using diving dresses and self-contained oxygen apparatus. It has been named the Torricellian Chamber from its resemblance to the simple mercury tube type of barometer invented by the Italian physicist Torricelli in 1643. The difficulty is to account for the filling of the cave roof in the first place; and to explain why the escape of the air, with which the cave water is saturated, does not empty it again.

In an attempt to throw light on these problems, gas was released on two occasions into the cave roof, either from the diving equipment or through a hose pipe. In this way the level of water in the chamber was lowered by amounts ranging from 6 in. to 4½ ft. In each case, the level was found to have been restored, four months or more later, by a major flood. But it was still necessary to explain how the gas which had been released was removed from the cavern roof. Fuller examination showed that, upstream from the chamber, there was a gap in a pile of boulders, which under flood conditions might direct a jet of water obliquely upwards which would hit one side of the rock and might start a whirlpool. Laboratory experiments were therefore carried out to test the ability of the supposed vortex action to remove air trapped above water in a standard glass measuring vessel. It was found that a jet of water from a normal tap could be so directed as to spiral round the walls of the vessel, and then down the centre at greatly increased speed. The descending column had a hollow, irregular core down which air was carried.

From the volume of water passing through Peak Cavern at times of flood, it was concluded that the current at the boulder dam must reach 20 m.p.h. or more; and Dr G. K. T. Conn, also of Sheffield University, calculated that this would be sufficient to produce a pumping action of the kind which had been demonstrated. Finally, the real existence of such a vortex was confirmed by making a visit to the cavern at a time when flood

conditions were developing. Handfuls of mud, released from the floor of the passage into the cavern, were seen to swirl slowly round. The sucking out of air from the chamber, and the rising of water to take its place, was thus as fully confirmed as could be expected – unless under the impossible conditions, for divers, of a major flood.

The second problem was to account for the retention of this high water level during the intervals, sometimes of many months or even a year, between successive floods. This seemed difficult to explain, if the water in the chamber, like that in the caves as a whole, was saturated with air; and a biological explanation was sought in the existence of bacteria on the chamber walls by which oxygen could be absorbed. Wall scrapings were therefore collected by divers from the chamber and brought to Sheffield for examination. But the amount of bacteria which these contained was found to be negligible. Finally, it was shown by calculation that, if only the top 1 cm. of water in the chamber were assumed to be stagnant, then the rate at which air would diffuse upwards would be so slow at cave-water temperature (about 8° C.) as to lead to a drop in water level at the rate of not more than 1 cm. per year. It is concluded that the filling of the chamber with water is due to the removal of air by a pumping action of a whirlpool formed during floods; and that the maintenance of a permanent high water level is due to the relatively slow rate of diffusion of air from the water, compared with the time intervals between successive major floods.

Davies, R. E. *Nature*, **166**, 894 (1950).

GERM-FREE MAMMALS

The rearing of germ-free mammals involves a number of specialized techniques, beginning with a sterile Caesarean operation, and continuing with sterile feeding, at first at hourly intervals and by hand. The necessary procedures have been developed almost to a fine art by Professor J. A. Reyniers and his colleagues of Loblund – the Laboratories of Bacteriology of the University of Notre Dame, Indiana – and have been applied to the rearing of germ-free guinea-pigs, mice, rats, rabbits, monkeys,

dogs and cats. They are shown in Inset, *13–22 following p. 64.* The use of the animals, when reared, is in nutritional research, which is thus freed from the complication of intestinal bacteria, and in experiments on antibody-formation and immunity. It is hoped shortly to establish a colony of germ-free rats, maintained through successive generations under sterile conditions and breeding naturally, in order that experimental animals can be made available in the numbers required for most forms of research.

Loblund Reports, Nos. 1 *and* 2. 'Germ-Free Life Studies'. University of Notre Dame, 1946 and 1949.
Reyniers, J. A., and others. *J. Nutrition*, **41**, 31 (1950).
du Vigneaud, V., and others. *Science*, **112**, 267 (1950).

CORRESPONDENCE

A HUNDRED YEARS OF THERMODYNAMICS

I SHOULD like to take up one or two points in Dr J. A. V. Butler's article in *Science News* 18. After giving an outline of Gibbs' contribution, he says on page 28:

'There is a curious footnote to this. Some years before the 1939 war, Mr J. G. Crowther, an exponent of the view that science springs entirely from economic necessities, surprisingly chose Willard Gibbs as an example to illustrate his thesis. He says: "As Isaac Newton supplied the scientific needs of the merchant traders of his day, Willard Gibbs supplied the scientific needs of the rationalizing and efficiency-hunting industrialists who have controlled Western civilisation since the middle of the nineteenth century". Could anything be more wrong-headed ...?'

A more careful reading of Crowther's chapter on Gibbs will, I am sure, convince Dr Butler that he is making a totally incorrect interpretation of Crowther's meaning. It is quite clear when this quoted passage is read in its context, that the author meant simply that the *result* of Gibbs' work was to provide the necessary scientific basis for the efficient conduct of chemical operations. Crowther was, in fact, at considerable pains to point out that this result was not achieved until many years after the publication of Gibbs' memoir. Moreover he devoted a considerable amount of space to a discussion of the motivation of Gibbs' researches, and nowhere does he suggest anything so crude as a direct inspiration from the needs of contemporary industry, nor, for that matter, anything so simple as a single-minded devotion to the 'urge to know, to penetrate the darkness, to find order and meaning in the chaotic images of the world.'

It is really surprising that a man who shows a proper degree of scientific caution when dealing with his own subject, as Dr Butler does when discussing the relation of life to the field of application of the second law, should throw this caution so completely to the winds when dealing with another branch,

namely that part of psychology which deals with human motivation. One would really think that Dr Butler had never heard of the idea of subconscious motivation, which by now rests on as solid a body of experimental evidence as many parts of physical science. Obviously this cannot be true, but it is evident that in forming his opinions about the motives for scientific research and theorizing, he has paid a great deal less attention to what the psychologists have to say on the subject than Mr Crowther has done. I find the approach of Mr Crowther to the question considerably more scientific than that of Dr Butler.

The truth is, of course, that the motivation of any action is complex. The motivation of such an extremely complicated action as that of Gibbs in developing the consequences of the second law for phase equilibria must have been especially so. The scientist himself is generally the last person to say what his motives really were, for he can only say (if he will) what was in his conscious mind, and, as we now know, that was the result of extremely involved processes in his subconcious. An example of this seems to me to be provided by Langmuir. He has said that the conscious motivation of his early researches was pure scientific interest, but that on many occasions he discussed with his director of research at the General Electric Company whether it was right that he should concentrate so exclusively on fundamental work of no immediate obvious practical value. The reply was always in the affirmative, and Langmuir continued in his fundamental researches. The subsequent enormous practical importance of nearly all this early work may be a pure accident, but it is equally arguable that Langmuir's choice of a problem was subconciously affected by his feeling that his work ought to have some practical result.

If then it is so difficult to make positive statements about the motivation of scientific developments in historic times, how much more difficult, if not impossible, to make them about events before the dawn of recorded history! Yet Dr Butler categorically states (p. 29): ' ... but the impetus to increased knowledge has not been the practical end but the urge to know, to penetrate the darkness, to find order and meaning in the

chaotic images of the world, which confronts even primitive man and which has given rise to language, to calculation, to philosophy, to art, and literature and every manifestation of civilization.' What is the scientific evidence for such a statement? There is none, and in the nature of the case there cannot be any. Positive assertions that such and such a process *must* have been the origin of the universe, unscientific as they are, are vastly more scientific than this statement of Dr Butler's. The origins of language and of every manifestation of civilization are lost in the mists of antiquity*: we can only measure the cranial capacity of the skulls of those we imagine to have been the originators, but we can never have positive knowledge of the processes that went on in the grey matter which once filled them. What evidence there is rather suggests that the origins at least of calculation are to be sought in the urgent practical needs of primitive society. One has only to think of the obviously close connection between the development of astronomy in ancient Egypt, and the urgent need for the accurate prediction of the date of the Nile flood, or of how science and indeed civilization generally has stagnated in communities whose economy was essentially static, such as was that of the Aztecs and Mayas.

No-one who has studied the matter at all closely, and Mr Crowther least of all, would deny that the desire to find out how the Universe works (or however one may wish to express the matter) has been a very important part of the conscious motivation of all really important scientific work, and has even been the dominant conscious motive in much of it. But to suggest that it is the only motive, in every case, is to turn a completely blind eye on what experimental psychology has found out about the structure of the mind, and also implies a divorce between the scientist and the community in which he lives which just does not occur in practice. Surely many scientists share with poets the virtue of being humanity's windows on the future? Is it not a plausible hypothesis, and one which, if true, would be very

* See article on 'The Origin of Language' on page 82 of this issue —ED.

creditable to Gibbs, that his work in thermodynamics was motivated both by his interest in it and by a far-sighted realization of the future practical value of the application of thermodynamics to chemistry?

R. C. MURRAY

Dr J. A. V. Butler writes:

I should be sorry if I have attributed to Mr Crowther views which he does not hold, but the whole point of his book seemed to be to illustrate the importance of practical needs and objectives as providing the impetus of scientific discovery, and the unimportance of an intellectual attempt merely to understand the Universe. He returns to this theme, in various guises, again and again. Willard Gibbs is perhaps the most unfavourable example which could be chosen to illustrate this thesis, and I attempted to show that although the development of thermodynamics up to about 1850 is an excellent example of science being led by practice, in the hands of Clausius and Thomson about 1850 it joined the older tradition of mathematical physics which began with Kepler, Galileo, and Newton, and had its origin in the attempt to understand the motions of the planets, and it is to this tradition that Gibbs's work belongs.

Mr Murray takes me up on a paragraph in which I attempt to place Gibbs's work in scientific history in terms of a philosophy of science which is at least widely held. It is of course impossible to do more than speculate on the origin of language, but a great deal of thought has been given to the nature and structure of existing languages, which is not inconsistent with the view I expressed. The idea that human beings are naturally curious and have an urge to unite their experiences into a consistent theory (which in its more advanced stages has led to science), so far from being inconsistent with psychology, is supported by what is known about how the brain deals with its sensory data, as I have attempted to show in my book *Man is a Microcosm*. Professor J. Z. Young's recent lectures on *Doubt and Certainty in Science* appear to tend in the same direction.

But as Mr Murray thinks that 'the desire to find out how the Universe works has been a very important part of the conscious motivation of all really important scientific work and has even been the dominant conscious motive in much of it', I do not think there is really any important difference of opinion between us. I did not suggest that this was the only motive; to the contrary, my article endeavoured to bring out in this limited field how the practical and the intellectual approaches had interacted to their mutual advantage.

THEORETICAL IMPLICATIONS OF TELEPATHY

REFERRING to the Soal experiments in which Mr Shackleton correctly guessed one card in advance in a long series of experiments, Margaret Knight says that '... the most disturbing feature of the experiments is that they involve precognition.' But do they? A hypothesis must fit all the known facts; if it has to ignore one of them, however inconvenient, it is not tenable.

Precognition – the knowledge of an event before it occurs – is here assumed because Shackleton correctly guessed the next card ahead. The event in this case is presumably the turning up and recognition of the card by the agent – not the *fact* that the next card in the pile at the time of the guess bore a certain symbol. In a run of, say, 50 guesses, then, Shackleton's 48th guess was correct for the 49th card: his 49th guess was correct for the 50th card. What was his 50th guess correct for? There was no future event to correspond with it. It was a guess at a future event which did not happen. Therefore, his last guess must always have been incorrect, and could not have been caused by precognition. However much his guessing in advance improves with practice, he can never score 100 per cent. And the hypothesis of precognition must, by its very nature, permit of 100 per cent success. Moreover, if the length of each run is reduced, the percentage chance of complete success is also reduced. In the extreme case of runs of only one card Shackleton would never be right.

In any case, supposing that the runs were not taken to a previously decided total but were stopped, either accidentally or

arbitrarily, just after Shackleton had made a guess. If the precognition theory is correct this implies something even more disturbing, i.e., that a future event, already known by one individual, can be cancelled. It might be said that in this case the next card, even though not turned up, could be the one which Shackleton had already guessed. But this would not be precognition, merely the correct guessing of the order of a pack of cards – an already existing event.

It would be interesting to examine the records of these experiments to see whether any of Shackleton's 'last guesses' corresponded with the first card of the same run (for which incidentally, his first guess of a series must always be incorrect). If so, it cannot be precognition: but it would be very interesting indeed if his last guess of a series corresponded with the first card of the next series.

Another fact appearing in the previous article (by E. J. Dingwall and Denys Parsons in *Science News* 9), though not taken into account by Margaret Knight, is that in the original experiments two people, one of them Shackleton, also scored significant hits on the *preceding* card. This is not precognition.

Surely the answer to all this is that the order of the cards, though unknown, was already determined before the start of the experiment, in fact from the moment that the agent finished shuffling the cards. The order in which they would then be turned up was fixed by the already prepared sheet of random numbers. Shackleton was in fact guessing the predetermined order of a pack of cards, which, though very mysterious, was not precognition. Indeed it was not even telepathy. This could easily be put to the test by running a series of experiments in which the agent picked up the card, but did not turn it over or otherwise see the symbol on it.

One more point which is brought out by this and previous articles is that it is always the receiving mind '(percipient') which is tested and not the transmitting one, which is merely referred to as the 'agent'. Hence all telepathy experiments so far have been directed at testing the receiving ability of a mind. Surely transmitting ability is equally or more important? It is

ıdeed mentioned that when the agent was changed, results fell o chance expectations, which seems to show that the success f the experiments is a measure of the success of the agent. hould not research now be directed at discovering successful ransmitters?

N. G. STRONG

Irs Margaret Knight writes:

Mr Strong's first argument is a curious one. He says that hackleton's 'advance hits' could not have been due to pre-ognition, since 'the hypothesis of precognition must, by its ery nature, permit of 100 per cent success', and 100 per cent uccess was impossible in the conditions of the experiment since one cannot score an 'advance hit' with the last guess of a series.) I am bound to say that I can see no force in this argument. Let us consider an analogous case. Suppose that we are doing an experiment on 'straight' telepathy. The agent, behind a screen, repeatedly tosses a penny, and the subject is told to guess 'heads' or 'tails' whenever he receives the signal to do so. He knows that, once in twenty-five times, the signal will be given though the penny has not, in fact, been tossed; but he must still say 'heads' or 'tails'. The conditions of the experiment make it impossible for him to score more than 24 hits out of 25, but this does not mean that the hypothesis of telepathy cannot be invoked, if the hits that he does score are greater than chance expectation.

Mr Strong's second point is that it is unnecessary to assume precognition, since the facts can equally well be explained on the hypothesis that Shackleton was clairvoyantly perceiving the order of the cards in the pack. This is certainly a possibility, and has been discussed by Soal and many other writers. Mr Strong suggests that the hypothesis 'could easily be put to the test by ... experiments in which the agent picked up the card, but did not turn it over or otherwise see the symbol on it'. Such experiments were in fact made by Soal, with Shackleton as subject; the results, which were negative, were described on pp. 12–13 of my article. Further and more extensive experiments

on clairvoyance have been carried out in this country by Thouless – again with negative results. But since Rhine in America claims positive results with some subjects (see again p. 12 of my article), the hypothesis of clairvoyance cannot be said to have been completely eliminated.

BOOKS RECEIVED

CALCULATING INSTRUMENTS AND MACHINES by Douglas R. Hartree (Cambridge University Press, 1950), pp. 138. 21s.

This is the English edition of a course of lectures given by Professor Hartree in 1948 at the University of Illinois to an audience assumed to be without specialized knowledge, though with enough of mathematics to cope with the problems and methods discussed. The distinction implied in the title is between 'instruments', in which accuracy is limited by some form of measurement or reading, and 'machines' which deal with numbers as such.

THE SCIENCE OF HUMANITY by K. G. Collier (Edinburgh, Nelson, 1950), pp. 339. 12s. 6d.

The author is a former physics master at a public school, and now a lecturer in education. He approaches the problem of human organization and social structure through physiology, psychology, anthropology and sociology, and with a political viewpoint which he describes as Radical Liberal. He has obviously read widely, and deserves to succeed in his objects, which are to encourage discussion and provide material for it.

SEAWEEDS AND THEIR USES by V. J. Chapman (London, Methuen, 1950), pp. 287. 25s.

Professor Chapman is a botanist with more enthusiasm for seaweeds than for their economic utilization under western conditions, though he thinks that successive generations should keep on trying. He has made an interesting book which, however, has been longer between first draft (1946) and publication than could fairly be expected.

IONIZATION CHAMBERS AND COUNTERS by D. H. Wilkinson (Cambridge University Press, 1950), pp. 226. 25s.

Dr Wilkinson writes for the working physicist, whose interest is in the principles and limitations of the devices discussed. As such, his book is clearly written and logically presented: it is not, for example, for the biologist in search of easy advice.

ABOUT OUR CONTRIBUTORS

VICTOR KEATING was born in 1913 and, following medical qualification, has been engaged in the speciality of anaesthetics since 1939. He is the author of papers on cyclopropane anaesthesia and the effects of the common anaesthetics on the heart and circulation; and is interested in the pharmacology of anaesthetic drugs, the design of anaesthetic instruments, and physiology applied to anaesthesia. He is now anaesthetist to a neuro-surgical unit.

R. S. NYHOLM was born in Australia and is a graduate of Sydney University. He lectured in chemistry at Sydney for several years before coming to University College, London, on a research fellowship in 1947. His chief interest is the study of complex inorganic compounds, on which he has published several papers dealing with their structures. He has made special use of magnet chemistry and other physical methods.

WILLIAM H. MARSHALL, graduate of Glasgow University, Fellow of the Royal Astronomical Society, Vice-President of the Scottish Branch of the British Astronomical Association, worked on explosives during the war and now teaches mathematics in James Hamilton School, Kilmarnock.

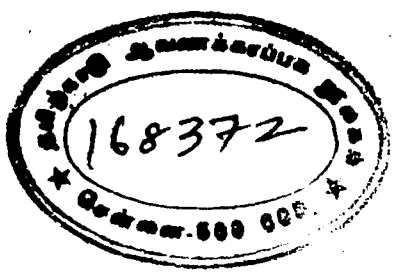
KENNETH OAKLEY graduated in geology and anthropology at University College, London, in 1933. He has 'specialized' in borderline subjects. One of his first papers (in *Proc. Roy. Soc.*) was on pathological phosphatic bodies in fossil bryozoa. He is now a principal scientific officer in the Department of Geology, British Museum (Natural History), and is in charge of 'fossil man'. He has investigated the geological dating evidence of various fossil human skulls, and in this connection has contributed to the development of the fluorine-dating method. Author of *Man the Tool-Maker*, British Museum handbook.

R. A. S. PAGET (Sir Richard Paget, 2nd Bt) was born in 1869, and combines the qualifications of barrister and Fellow of the Institute of Physics. He was educated at Eton and Magdalen College, Oxford, had early experience of arbitration, and during the first world war was assistant secretary of the Admiralty Board of Inventions and Research. He has done original research on human speech, and on this subject is the author of publications in the *Proceedings of the Royal Society* and elsewhere.

[*Continued overleaf*

T. D. SMITH is a graduate of the University of London. He carried out postgraduate research at King's College, Newcastle-upon-Tyne; in 1948 was awarded a teaching and research fellowship at the University of Southern California, Los Angeles; and later worked in the research department of a Californian oil company. He is now working in the Department of Colloid Science, the University of Cambridge. Age 27.

A. G. THOMSON was educated at Uppingham and went to South Africa in 1925, where he gained experience in scientific journalism. Since returning to England in August 1949, he has contributed articles to a number of British publications.



N71-20

2

(2)	(1)	(2)